15

Continuous random variables and their probability distributions

Objectives

- ▶ To introduce **continuous random variables**.
- ▶ To use **probability density functions** to specify the distributions for continuous random variables.
- ▶ To relate the probability for an interval to an area under the graph of a probability density function.
- ▶ To use calculus to find probabilities for intervals from a probability density function.
- ▶ To use technology to find probabilities for intervals from a probability density function.
- ▶ To calculate and interpret the **expectation** (**mean**), **variance** and **standard deviation** for a continuous random variable.

In Chapter 13 we studied discrete random variables, that is, random variables that take only a countable number of values. Most of the examples in that chapter involved the natural numbers: for example, the number of heads observed when tossing a coin several times.

In this chapter we extend our knowledge to include continuous random variables, which can take any value in an interval of the real number line. Examples include the time taken to complete a puzzle and the height of an adult. When considering the heights of adults, the range of values could be from 56 cm to 251 cm, and in principle the measurement could be any value in this interval.

We also introduce the concept of the probability density function to describe the distribution of a continuous random variable. We shall see that probabilities associated with a continuous random variable are described by areas under the probability density function, and thus integration is an important skill required to determine these probabilities.

ISBN 978-1-009-11049-5 Cambridge University Press

15A Continuous random variables

A **continuous random variable** is one that can take any value in an interval of the real number line. For example, if X is the random variable which takes its values as 'distance in metres that a parachutist lands from a marker', then X is a continuous random variable, and here the values which X may take are the non-negative real numbers.

An example of a continuous random variable

A continuous random variable has no limit as to the accuracy with which it can be measured. For example, let W be the random variable with values 'a person's weight in kilograms' and let W_i be the random variable with values 'a person's weight in kilograms measured to the ith decimal place'.

Then $W_0 = 83$ implies $82.5 \le W < 83.5$

$W_1 = 83.3$ implies $83.25 \le W < 83.35$

$W_2 = 83.28$ implies $83.275 \le W < 83.285$

$W_3 = 83.281$ implies $83.2805 \le W < 83.2815$

and so on. Thus, the random variable W cannot take an exact value, since it is always rounded to the limits imposed by the method of measurement used. Hence, the probability of W being exactly equal to a particular value is zero, and this is true for all continuous random variables.

That is,

$$\Pr(W = w) = 0 \quad \text{for all } w$$

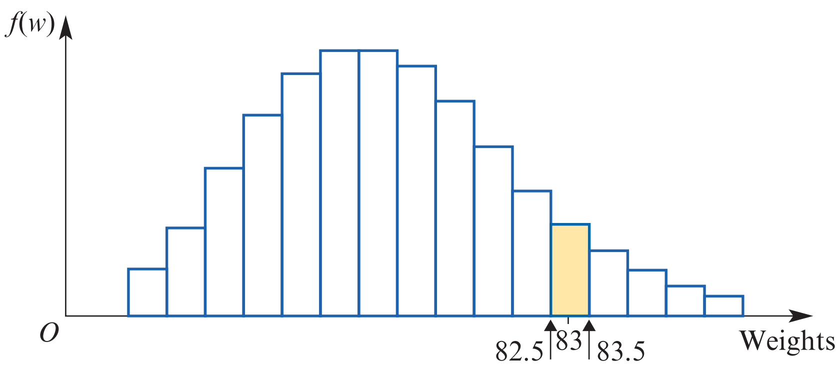
In practice, considering W_i taking a particular value is equivalent to W taking a value in an appropriate interval.

Thus, from above:

$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5)$$

To determine the value of this probability, you could begin by measuring the weight of a large number of randomly chosen people, and determine the proportion of the people in the group who have weights in this interval.

Suppose after doing this a histogram of weights was obtained as shown.



ISBN 978-1-009-11049-5 Cambridge University Press

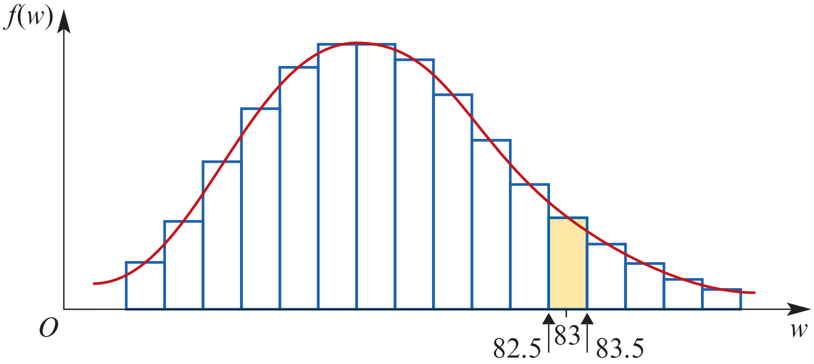
From this histogram:

$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5)$$
$$= \frac{\text{shaded area from 82.5 to 83.5}}{\text{total area}}$$

If the histogram is scaled so that the total area under the blocks is 1, then

$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5)$$
$$= \text{area under block from 82.5 to 83.5}$$

Now suppose that the sample size gets larger and that the class interval width gets smaller. If theoretically this process is continued so that the intervals are arbitrarily small, then the histogram can be modelled by a smooth curve, as shown in the following diagram.



The curve obtained here is of great importance for a continuous random variable.

The function f whose graph models the histogram as the number of intervals is increased is called the **probability density function**. The probability density function f is used to describe the probability distribution of a continuous random variable X.

Now, the probability of interest is no longer represented by the area under the histogram, but by the area under the curve. That is,

$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5)$$
$$= \text{area under the graph of the function with rule } f(w) \text{ from 82.5 to 83.5}$$
$$= \int_{82.5}^{83.5} f(w)\, dw$$

Probability density functions

In general, a **probability density function** f is a function with domain some interval (e.g. domain $[c, d]$ or $\mathbb{R}$) such that:

1 $f(x) \ge 0$ for all x in the interval, and

2 the area under the graph of the function is equal to 1.

If the domain of f is $[c, d]$, then the second condition corresponds to $\int_c^d f(x)\, dx = 1$.

ISBN 978-1-009-11049-5 Cambridge University Press

In many cases, however, the domain of f will be an 'unbounded' interval such as $[1, \infty)$ or $\mathbb{R}$. Therefore, some new notation is necessary.

- If the probability density function f has domain $[1, \infty)$, then $\int_1^{\infty} f(x)\,dx = 1$. This integral is computed as $\lim_{k\to\infty} \int_1^k f(x)\,dx$.
- If the probability density function f has domain $\mathbb{R}$, then $\int_{-\infty}^{\infty} f(x)\,dx = 1$. This integral is computed as $\lim_{k\to\infty} \int_{-k}^k f(x)\,dx$.

Note: Definite integrals which have one or both limits infinite are called **improper integrals**. There are possible complications with such integrals which we avoid in this course; you will only need the methods of evaluation illustrated in Examples 1 and 3.

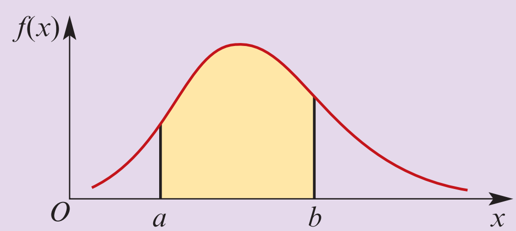
The probability density function of a random variable

Now consider a continuous random variable X with range $[c, d]$. (Alternatively, the range of X may be an unbounded interval such as $(-\infty, d]$, $[c, \infty)$ or $\mathbb{R}$.) Let f be a probability density function with domain $[c, d]$. Then:

We say that f is the **probability density function of X** if

$$\Pr(a < X < b) = \int_a^b f(x)\,dx$$

for all $a < b$ in the range of X.



Notes:

- The values of a probability density function f are not probabilities, and $f(x)$ may take values greater than 1.
- The probability of any specific value of X is 0. That is, $\Pr(X = a) = 0$. It follows that all of the following expressions have the same numerical value:
  - $\Pr(a < X < b)$
  - $\Pr(a \le X < b)$
  - $\Pr(a < X \le b)$
  - $\Pr(a \le X \le b)$
- If f has domain $[c, d]$ and $a \in [c, d]$, then $\Pr(X < a) = \Pr(X \le a) = \int_c^a f(x)\,dx$.

The natural extension of a probability density function

Any probability density function f with domain $[c, d]$ (or any other interval) may be extended to a function f^* with domain $\mathbb{R}$ by defining

$$f^*(x) = \begin{cases} f(x) & \text{if } x \in [c, d] \\ 0 & \text{if } x \notin [c, d] \end{cases}$$

This leads to the following:

A probability density function f (or its natural extension) must satisfy the following two properties:

1 $f(x) \ge 0$ for all x

2 $\int_{-\infty}^{\infty} f(x)\,dx = 1$

ISBN 978-1-009-11049-5 Cambridge University Press

Example 1

Suppose that the random variable X has the probability density function with rule:

$$f(x) = \begin{cases} cx & \text{if } 0 \le x \le 2 \\ 0 & \text{if } x > 2 \text{ or } x < 0 \end{cases}$$

a Find the value of c that makes f a probability density function.

b Find $\Pr(X > 1.5)$.

Solution

a Since f is a probability density function, we know that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

$$\text{Now } \int_{-\infty}^{\infty} f(x)\,dx = \int_0^2 cx\,dx \qquad \text{since } f(x) = 0 \text{ elsewhere}$$

$$= \left[\frac{cx^2}{2}\right]_0^2$$

$$= 2c$$

Therefore $2c = 1$ and so $c = 0.5$.

b
$$\Pr(X > 1.5) = \int_{1.5}^2 0.5x\,dx$$

$$= 0.5\left[\frac{x^2}{2}\right]_{1.5}^2$$

$$= 0.5\left(\frac{4}{2} - \frac{2.25}{2}\right)$$

$$= 0.4375$$

Example 2

Consider the function f with the rule:

$$f(x) = \begin{cases} 1.5(1 - x^2) & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \text{ or } x < 0 \end{cases}$$

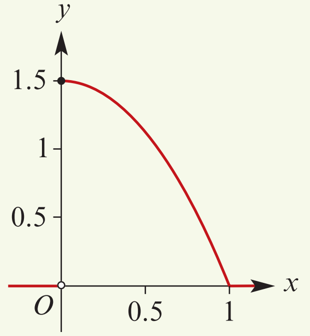
a Sketch the graph of f.

b Show that f is a probability density function.

c Find $\Pr(X > 0.5)$, where the random variable X has probability density function f.

Solution

a For $0 \le x \le 1$, the graph of $y = f(x)$ is part of a parabola with intercepts at $(0, 1.5)$ and $(1, 0)$.



ISBN 978-1-009-11049-5

b From the graph, we can see that $f(x) \geq 0$ for all x, and so the first condition holds.

The second condition to check is that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

$$\text{Now } \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 1.5(1 - x^2)\,dx \qquad \text{since } f(x) = 0 \text{ elsewhere}$$
$$= 1.5\left[x - \frac{x^3}{3}\right]_0^1$$
$$= 1.5\left(1 - \frac{1}{3}\right)$$
$$= 1$$

Thus the second condition holds, and hence f is a probability density function.

c $$\Pr(X > 0.5) = \int_{0.5}^1 1.5(1 - x^2)\,dx$$
$$= 1.5\left[x - \frac{x^3}{3}\right]_{0.5}^1$$
$$= 1.5\left[\left(1 - \frac{1}{3}\right) - \left(0.5 - \frac{0.125}{3}\right)\right]$$
$$= 0.3125$$

Probability density functions with unbounded domain

Some intervals for which definite integrals need to be evaluated are of the form $(-\infty, a]$ or $[a, \infty)$ or $(-\infty, \infty)$. For a function f with non-negative values, such integrals are defined as follows (provided the limits exist):

- To integrate over the interval $(-\infty, a]$, find $\lim_{k \to -\infty} \int_k^a f(x)\,dx$.
- To integrate over the interval $[a, \infty)$, find $\lim_{k \to \infty} \int_a^k f(x)\,dx$.
- To integrate over the interval $(-\infty, \infty)$, find $\lim_{k \to \infty} \int_{-k}^k f(x)\,dx$.

Example 3

Consider the exponential probability density function f with the rule:

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

a Sketch the graph of f.

b Show that f is a probability density function.

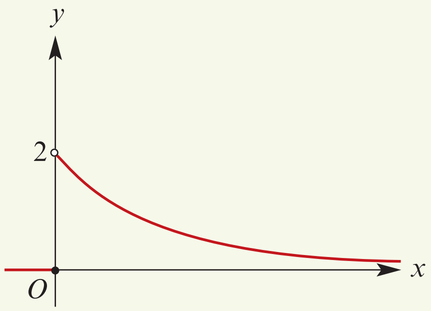
c Find $\Pr(X > 1)$, where the random variable X has probability density function f.

ISBN 978-1-009-11049-5

Cambridge University Press

Solution

a For $x > 0$, the graph of $y = f(x)$ is part of the graph of an exponential function with y-axis intercept 2. As $x \to \infty$, $y \to 0$.



b Since $f(x) \geq 0$ for all x, the first condition holds.

The second condition to check is that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Now $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} 2e^{-2x}\,dx$ since $f(x) = 0$ elsewhere

$$\begin{aligned}
&= \lim_{k\to\infty} \int_0^k 2e^{-2x}\,dx \\
&= \lim_{k\to\infty} \left[\frac{2e^{-2x}}{-2}\right]_0^k \\
&= \lim_{k\to\infty} \left[-e^{-2x}\right]_0^k \\
&= \lim_{k\to\infty} \left((-e^{-2k}) - (-e^{-0})\right) \\
&= 0 + e^0 \\
&= 1
\end{aligned}$$

Thus f satisfies the two conditions for a probability density function.

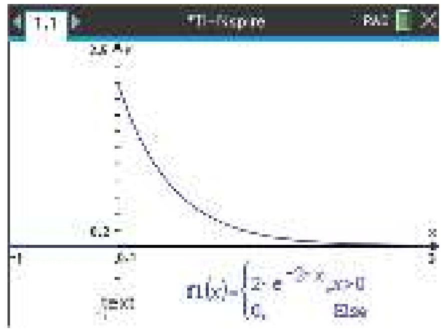
c
$$\begin{aligned}
\Pr(X > 1) &= \lim_{k\to\infty} \int_1^k 2e^{-2x}\,dx \\
&= \lim_{k\to\infty} \left[\frac{2e^{-2x}}{-2}\right]_1^k \\
&= \lim_{k\to\infty} \left[-e^{-2x}\right]_1^k \\
&= \lim_{k\to\infty} \left((-e^{-2k}) - (-e^{-2})\right) \\
&= 0 + e^{-2} \\
&= \frac{1}{e^2} \\
&= 0.1353 \quad \text{correct to four decimal places}
\end{aligned}$$

ISBN 978-1-009-11049-5 Cambridge University Press

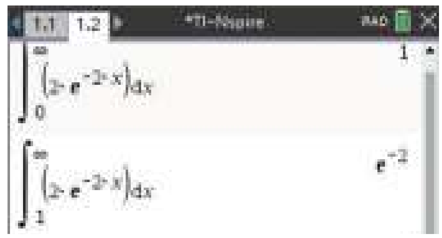
Using the TI-Nspire

This is an application of integration.

a The graph is as shown. The piecewise function template has been used in this example; access the template using the template key.



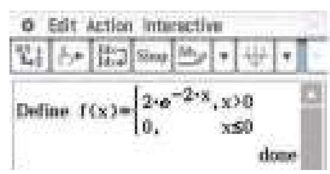
b, c The two required integrations are shown. The symbol ∞ can be found using π▸ or ctrl then the catalog key.

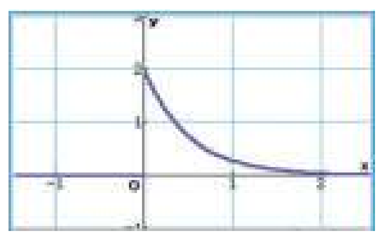


Using the Casio ClassPad

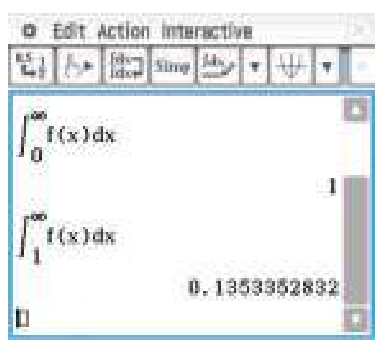
a To sketch the graph:

- Select the Math3 keyboard and tap on the piecewise template.
- Enter the function as shown, highlight and go to **Interactive** > **Define**.
- Now select the graph icon, highlight $f(x)$ and drag into the graph screen.
- Adjust the window using the window icon.





b, c Find the definite integrals as shown.



ISBN 978-1-009-11049-5

Cambridge University Press

Conditional probability

Next is an example involving conditional probability with continuous random variables.

Example 4

The time (in seconds) that it takes a student to complete a puzzle is a random variable X with a density function given by

$$f(x) = \begin{cases} \dfrac{5}{x^2} & x \geq 5 \\ 0 & x < 5 \end{cases}$$

a Find the probability that a student takes less than 12 seconds to complete the puzzle.

b Find the probability that a student takes between 8 and 10 seconds to complete the puzzle, given that he takes less than 12 seconds.

Solution

a
$$\Pr(X < 12) = \int_5^{12} f(x)\,dx = \int_5^{12} \frac{5}{x^2}\,dx = \left[-\frac{5}{x}\right]_5^{12} = -\frac{5}{12} + 1 = \frac{7}{12}$$

b
$$\Pr(8 < X < 10 \mid X < 12) = \frac{\Pr(8 < X < 10 \cap X < 12)}{\Pr(X < 12)} = \frac{\Pr(8 < X < 10)}{\Pr(X < 12)} = \frac{\int_8^{10} f(x)\,dx}{\int_5^{12} f(x)\,dx} = \frac{-\frac{1}{2} + \frac{5}{8}}{\frac{7}{12}} = \frac{3}{14}$$

Summary 15A

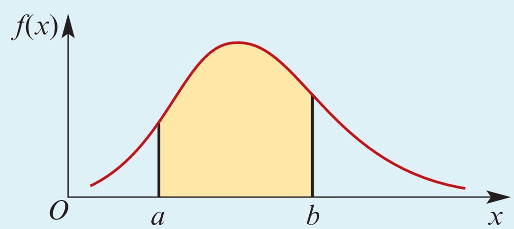
- A probability density function f (or its natural extension) must satisfy the following two properties:

  1 $f(x) \geq 0$ for all x **2** $\int_{-\infty}^{\infty} f(x)\,dx = 1$

- If X is a continuous random variable with density function f, then

$$\Pr(a < X < b) = \int_a^b f(x)\,dx$$

which is the area of the shaded region.



- Definite integrals may need to be evaluated over unbounded intervals:
  - To integrate over the interval $(-\infty, a]$, find $\lim\limits_{k\to-\infty} \int_k^a f(x)\,dx$.
  - To integrate over the interval $[a, \infty)$, find $\lim\limits_{k\to\infty} \int_a^k f(x)\,dx$.
  - To integrate over the interval $(-\infty, \infty)$, find $\lim\limits_{k\to\infty} \int_{-k}^k f(x)\,dx$.

ISBN 978-1-009-11049-5 Cambridge University Press

Skill-sheet

Exercise 15A

1 Show that the function f with the following rule is a probability density function:

$$f(x) = \begin{cases} \dfrac{24}{x^3} & 3 \le x \le 6 \\ 0 & x < 3 \text{ or } x > 6 \end{cases}$$

Example 1

2 Let X be a continuous random variable with the following probability density function:

$$f(x) = \begin{cases} x^2 + kx + 1 & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Determine the constant k such that f is a valid probability density function.

Example 2

3 Consider the random variable X having the probability density function with the rule:

$$f(x) = \begin{cases} 12x^2(1 - x) & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Sketch the graph of $y = f(x)$.

b Find $\Pr(X < 0.5)$.

c Shade the region which represents this probability on your sketch graph.

Example 3

4 Consider the random variable Y with the probability density function:

$$f(y) = \begin{cases} ke^{-y} & y \ge 0 \\ 0 & y < 0 \end{cases}$$

a Find the constant k.

b Find $\Pr(Y \le 2)$.

Example 4

5 The quarantine period for a certain disease is between 5 and 11 days after contact. The probability of showing the first symptoms at various times during the quarantine period is described by the probability density function:

$$f(t) = \frac{1}{36}(t - 5)(11 - t)$$

a Sketch the graph of the function.

b Find the probability that the symptoms appear within 7 days.

c Find the probability that the symptoms appear within 7 days, given that they appear after 5.5 days.

d Find the probability that the symptoms appear within 7 days, given that they appear within 10 days.

6 A probability model for the mass, X kg, of a 2-year-old child is given by

$$f(x) = k\sin\left(\frac{\pi(x - 7)}{10}\right), \quad 7 \le x \le 17$$

a Show that $k = \dfrac{\pi}{20}$.

b Hence find the percentage of 2-year-old children whose mass is:

i greater than 16 kg

ii between 12 kg and 13 kg.

ISBN 978-1-009-11049-5 Cambridge University Press

7 A probability density function for the lifetime, T hours, of Electra light bulbs has rule

$$f(t) = ke^{\left(\frac{-t}{200}\right)}, \quad t > 0$$

a Find the value of the constant k.

b Find the probability that an Electra light bulb will last more than 1000 hours.

8 A random variable X has a probability density function given by

$$f(x) = \begin{cases} k(1+x) & -1 \le x \le 0 \\ k(1-x) & 0 < x \le 1 \\ 0 & x < -1 \text{ or } x > 1 \end{cases}$$

where $k > 0$.

a Sketch the graph of the probability density function. **b** Evaluate k.

c Find the probability that X lies between -0.5 and 0.5.

9 Let X be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Sketch the graph of $y = f(x)$.

b Find $\Pr(0.25 < X < 0.75)$ and illustrate this on your graph.

10 A random variable X has a probability density function f with the rule:

$$f(x) = \begin{cases} \dfrac{1}{100}(10+x) & \text{if } -10 < x \le 0 \\ \dfrac{1}{100}(10-x) & \text{if } 0 < x \le 10 \\ 0 & \text{if } x \le -10 \text{ or } x > 10 \end{cases}$$

a Sketch the graph of f. **b** Find $\Pr(-1 \le X < 1)$.

11 The life, X hours, of a type of light bulb has a probability density function with the rule:

$$f(x) = \begin{cases} \dfrac{k}{x^2} & x > 1000 \\ 0 & x \le 1000 \end{cases}$$

a Evaluate k. **b** Find the probability that a bulb will last at least 2000 hours.

12 The weekly demand for petrol, X (in thousands of litres), at a particular service station is a random variable with probability density function:

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \le x \le 2 \\ 0 & x < 1 \text{ or } x > 2 \end{cases}$$

a Determine the probability that more than 1.5 thousand litres are bought in one week.

b Determine the probability that the demand for petrol in one week is less than 1.8 thousand litres, given than it is more than 1.5 thousand litres.

ISBN 978-1-009-11049-5

13 The length of time, X minutes, between the arrival of customers at an ATM is a random variable with probability density function:

$$f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

a Find the probability that more than 8 minutes elapses between successive customers.

b Find the probability that more than 12 minutes elapses between successive customers, given that more than 8 minutes has passed.

14 A random variable X has density function given by

$$f(x) = \begin{cases} 0.2 & -1 < x \le 0 \\ 0.2 + 1.2x & 0 < x \le 1 \\ 0 & x \le -1 \text{ or } x > 1 \end{cases}$$

a Find $\Pr(X \le 0.5)$.

b Hence find $\Pr(X > 0.5 \mid X > 0.1)$.

15 The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

a Sketch the graph of f.

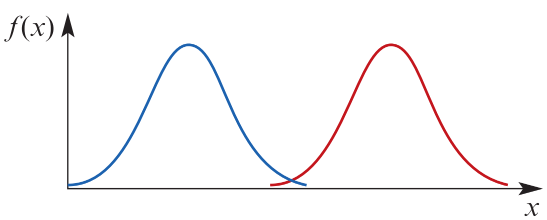
b Find:

i $\Pr(X < 0.5)$ **ii** $\Pr(X \ge 1)$ **iii** $\Pr(X \ge 1 \mid X > 0.5)$

15B Mean and percentiles for a continuous random variable

The centre is an important summary feature of a probability distribution.

The following diagram shows two probability distributions which have the same shape and the same spread, but differ in their centres.



More than one measure of centre may be determined for a continuous random variable, and each gives useful information about the random variable under consideration. The most generally useful measure of centre is the mean.

ISBN 978-1-009-11049-5

Mean

We defined the mean for a discrete random variable in Section 13D. We can also define the mean for a continuous random variable.

> For a continuous random variable X with probability density function f, the **mean** or **expected value** of X is given by
>
> $$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$
>
> provided the integral exists. The mean is denoted by the Greek letter μ (mu).
>
> If $f(x) = 0$ for all $x \notin [c, d]$, then $\mathrm{E}(X) = \int_c^d x f(x)\, dx$.

This definition is consistent with the definition of the expected value for a discrete random variable. As in the case of a discrete random variable, the expected value of a continuous random variable is the long-run average value of the variable. For example, consider the daily demand for petrol at a service station. The mean of this variable tells us the average daily demand for petrol over a very long period of time.

Example 5

Find the expected value of the random variable X which has probability density function with rule:

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Solution

By definition,

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^2 x \times 0.5x\, dx \qquad \text{since } f(x) = 0 \text{ elsewhere}$$

$$= 0.5 \int_0^2 x^2\, dx$$

$$= 0.5 \left[\frac{x^3}{3} \right]_0^2$$

$$= \frac{4}{3}$$

Using the TI-Nspire

Assign the function f as shown; access the piecewise function template using [piecewise template key].

Notes:

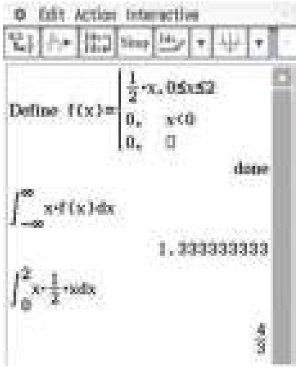
- Leave the domain for the last function piece blank; it will autofill as 'Else'.
- To obtain an exact answer, enter $\frac{1}{2}x$ instead of $0.5x$.

ISBN 978-1-009-11049-5 Cambridge University Press

Using the Casio ClassPad

- Tap the piecewise template twice.
- Define the function f as shown.
- Find $\mathrm{E}(X)$ by evaluating the definite integral as shown.

Note: Using the defined function to find $\mathrm{E}(X)$ gives the decimal answer only.



The mean of a function of X is calculated as follows. (In this case, the function of X is denoted by $g(X)$ and is the composition of the random variable X followed by the function g.)

The expected value of $g(X)$ is given by

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

provided the integral exists.

Generally, as in the case of a discrete random variable, the expected value of a function of X is not equal to that function of the expected value of X. That is,

$$\mathrm{E}[g(X)] \neq g[\mathrm{E}(X)]$$

Example 6

Let X be a random variable with probability density function f given by

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Find:

a the expected value of X^2

b the expected value of e^X.

Solution

a
$$\begin{aligned} \mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \\ &= \int_0^2 x^2 \times 0.5x\,dx \\ &= 0.5\int_0^2 x^3\,dx \\ &= 0.5\left[\frac{x^4}{4}\right]_0^2 \\ &= 2 \end{aligned}$$

b
$$\begin{aligned} \mathrm{E}(e^X) &= \int_{-\infty}^{\infty} e^x f(x)\,dx \\ &= \int_0^2 e^x f(x)\,dx \\ &= \int_0^2 e^x \times 0.5x\,dx \\ &= 4.195 \end{aligned}$$

correct to three decimal places.

ISBN 978-1-009-11049-5 Cambridge University Press

A case where the equality does hold is where g is a linear function:

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b \qquad \text{(for } a, b \text{ constant)}$$

Percentiles and the median

Another value of interest is the value of X which bounds a particular area under the probability density function. For example, a teacher may wish determine the mark, p, below which lie 75% of all students' marks. This is called the 75th percentile of the population, and is found by solving

$$\int_{-\infty}^{p} f(x)\, dx = 0.75$$

This can be stated more generally:

Percentiles

The value p of X which is the solution of an equation of the form

$$\int_{-\infty}^{p} f(x)\, dx = q$$

is called a **percentile** of the distribution.

For example, the 75th percentile is the value p found by taking $q = 75\% = 0.75$.

Example 7

The duration of telephone calls to the order department of a large company is a random variable, X minutes, with probability density function:

$$f(x) = \begin{cases} \frac{1}{3}e^{-\frac{x}{3}} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Find the value of a such that 90% of phone calls last less than a minutes.

Solution

To find the value of a, solve the equation:

$$\int_{0}^{a} \frac{1}{3}e^{-\frac{x}{3}}\, dx = 0.9$$

$$\left[-e^{-\frac{x}{3}}\right]_0^a = 0.9$$

$$1 - e^{-\frac{a}{3}} = 0.9$$

$$-\frac{a}{3} = \log_e 0.1$$

$$\therefore \quad a = 3\log_e 10$$

$$= 6.908 \qquad \text{(correct to three decimal places)}$$

So 90% of the calls to this company last less than 6.908 minutes.

ISBN 978-1-009-11049-5

Cambridge University Press

A percentile of special interest is the **median**, or 50th percentile. The median is the middle value of the distribution. That is, the probability of X taking a value below the median is 0.5, and the probability of X taking a value above the median is 0.5. Thus, if m is the median value of the distribution, then

$$\Pr(X \leq m) = \Pr(X > m) = 0.5$$

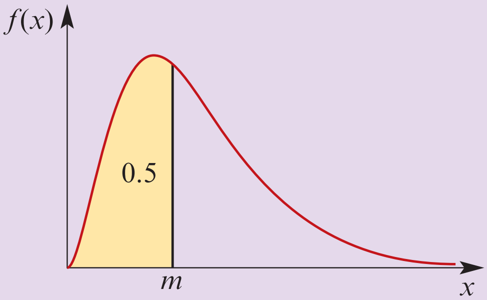
Graphically, the median is the value of the random variable which divides the area under the probability density function in half. It is a useful concept but it is not required by the Study Design for Mathematical Methods units 3&4.

The median (not required by study design)

The median is another measure of centre for a continuous probability distribution.

The median, m, of a continuous random variable X is the value of X such that

$$\int_{-\infty}^{m} f(x)\,dx = 0.5$$



Example 8

Suppose the probability density function of weekly sales of topsoil, X (in tonnes), is given by the rule:

$$f(x) = \begin{cases} 2(1-x) & 0 \leq x \leq 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

Find the value of m such that $\Pr(X \leq m) = 0.5$, and interpret.

Solution

$$\int_0^m 2(1-x)\,dx = 0.5$$

$$2\left[x - \frac{x^2}{2}\right]_0^m = 0.5$$

$$2m - m^2 = 0.5$$

$$m^2 - 2m + 0.5 = 0$$

$$\therefore \quad m = 0.293 \text{ or } m = 1.707$$

But since $0 \leq x \leq 1$, $m = 0.293$ tonnes.

This means that, in the long run, 50% of weekly sales will be less than 0.293 tonnes, and 50% will be more.

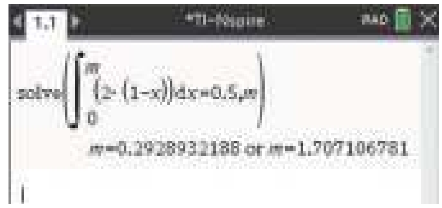
ISBN 978-1-009-11049-5

Using the TI-Nspire

This is an application of integration.

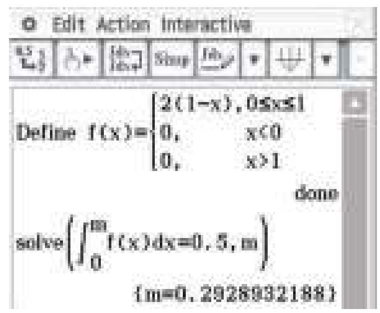
- Solve the definite integral equal to 0.5 as shown to find m (the median value).

Note: Since $0 \leq m \leq 1$, the domain constraint $|0 \leq m \leq 1$ could be added.



Using the Casio ClassPad

- Define the function f.
- Solve the definite integral equal to 0.5 as shown to find m (the median value).



Summary 15B

For a continuous random variable X with probability density function f:

- the **mean** or **expected value** of X is given by $\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$
- the expected value of $g(X)$ is given by $\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$
- The value p of X which is the solution of an equation of the form $\int_{-\infty}^{p} f(x)\, dx = q$ is called a **percentile** of the distribution.

Skill-sheet

Exercise 15B

Example 5

1 Find the mean, E(X), of the continuous random variables with the following probability density functions:

a $f(x) = 2x, \ 0 < x < 1$

b $f(x) = \dfrac{1}{2\sqrt{x}}, \ 0 < x < 1$

c $f(x) = 6x(1-x), \ 0 < x < 1$

d $f(x) = \dfrac{1}{x^2}, \ x \geq 1$

2 For each of the following, use your calculator to check that f is a probability density function and then to find the mean, E(X), of the corresponding continuous random variable:

a $f(x) = \sin x, \ 0 < x < \dfrac{\pi}{2}$

b $f(x) = \log_e x, \ 1 < x < e$

c $f(x) = \dfrac{1}{\sin^2 x}, \ \dfrac{\pi}{4} < x < \dfrac{\pi}{2}$

d $f(x) = -4x \log_e x, \ 0 < x < 1$

ISBN 978-1-009-11049-5

3 A continuous random variable X has the probability density function given by

$$f(x) = \begin{cases} 2x^3 - x + 1 & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Find μ, the mean value of X.

b Find the probability that X takes a value less than or equal to the mean.

4 Consider the probability density function given by

$$f(x) = \frac{1}{2\pi}(1 + \cos x), \quad -\pi \le x \le \pi$$

Find the expected value of X.

5 A random variable Y has the probability density function:

$$f(y) = \begin{cases} Ay & 0 \le y \le B \\ 0 & y < 0 \text{ or } y > B \end{cases}$$

Find A and B if the mean of Y is 2.

Example 6 **6** A random variable X has the probability density function given by

$$f(x) = \begin{cases} 12x^2(1 - x) & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Find $\mathrm{E}\left(\frac{1}{X}\right)$. **b** Find $\mathrm{E}(e^X)$.

Example 7 **7** The time, X seconds, between arrivals of particles at a radiation counter has been found to have a probability density function f with the rule:

$$f(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \ge 0 \end{cases}$$

a Find $\Pr(X \le 1)$. **b** Find $\Pr(1 \le X \le 2)$.

c Find the value, m, such that $\Pr(X \le m) = 0.5$. That is, the median.

8 The random variable X has a probability density function given by

$$f(x) = \begin{cases} k & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Find the value of k. **b** Find c such that $\Pr(X \le c) = 0.5$

9 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} 5(1 - x)^4 & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

Find c, such that $\Pr(X \le c) = 0.25$ correct to four decimal places.

ISBN 978-1-009-11049-5 Cambridge University Press

10 Suppose that the time (in minutes) between telephone calls received at a pizza restaurant has the probability density function:

$$f(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find c such that $\Pr(X > c) = 0.1$

Example 8 **11** A continuous random variable X has the probability density function given by

$$f(x) = \begin{cases} x & 0 \leq x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & x < 0 \text{ or } x \geq 2 \end{cases}$$

a Find μ, the expected value of X. **b** Find c such that $\Pr(X \leq c) = 0.5$

12 Let the probability density function of X be given by

$$f(x) = \begin{cases} 30x^4(1 - x) & 0 < x < 1 \\ 0 & x \leq 0 \text{ or } x \geq 1 \end{cases}$$

a Find the expected value, μ, of X.

b Find the median value, m, of X, that is, the value m such that $\Pr(X \leq m) = 0.5$, and hence show the mean is less than the median.

13 A probability model for the mass, X kg, of a 2-year-old child is given by

$$f(x) = \frac{\pi}{20}\sin\left(\frac{\pi(x - 7)}{10}\right), \quad 7 \leq x \leq 17$$

Find m such that $\Pr(X \leq m) = 0.5$

14 A random variable X has density function given by

$$f(x) = \begin{cases} 0.2 & -1 \leq x \leq 0 \\ 0.2 + 1.2x & 0 < x \leq 1 \\ 0 & x < -1 \text{ or } x > 1 \end{cases}$$

a Find μ, the expected value of X. **b** Find m, the median value of X.

15 The exponential probability distribution describes the distribution of the time between random events, such as phone calls. The general form of the exponential distribution with parameter λ is

$$f(x) = \begin{cases} \frac{1}{\lambda}e^{-\frac{x}{\lambda}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a Differentiate kxe^{-kx} and hence find an antiderivative of kxe^{-kx}.

b Show that the mean of an exponential random variable is λ.

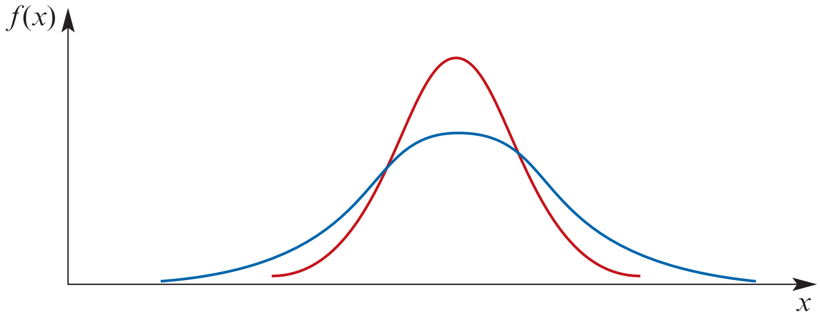
c On the same axes, sketch the graphs of the distribution for $\lambda = \frac{1}{2}$, $\lambda = 1$ and $\lambda = 2$.

d Describe the effect of varying the value of λ on the graph of the distribution.

ISBN 978-1-009-11049-5

15C Measures of spread

Another important summary feature of a distribution is variation or spread. The following diagram shows two distributions have the same shape and the same centre, but differ in their spread. their spreads.



As in the case of centre, there is more than one measure of spread. The most commonly used is the variance, together with its companion measure, the standard deviation. Others that you may be familiar with are the range and the interquartile range.

Variance and standard deviation

The **variance** of a random variable X is a measure of the spread of the probability distribution about its mean or expected value μ. It is defined as:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx\end{aligned}$$

As for discrete random variables, the variance is usually denoted by σ^2, where σ is the lowercase Greek letter *sigma*.

Variance may be considered as the long-run average value of the square of the distance from X to μ. This means that the variance is not in the same units of measurement as the original random variable X. A measure of spread in the appropriate unit is found by taking the square root of the variance.

The **standard deviation** of X is defined as:

$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

The standard deviation is usually denoted by σ.

As in the case of discrete random variables, an alternative (computational) formula for variance is generally used.

To calculate variance, use

$$\mathrm{Var(X)} = \mathrm{E}(X^2) - \mu^2$$

ISBN 978-1-009-11049-5 Cambridge University Press

Proof The computational form of the expression for variance is derived as follows:

$$\begin{aligned}\text{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \\ &= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \int_{-\infty}^{\infty} 2\mu x f(x)\,dx + \int_{-\infty}^{\infty} \mu^2 f(x)\,dx \\ &= \text{E}(X^2) - 2\mu \int_{-\infty}^{\infty} x f(x)\,dx + \mu^2 \int_{-\infty}^{\infty} f(x)\,dx\end{aligned}$$

Since $\int_{-\infty}^{\infty} x f(x)\,dx = \mu$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, we obtain

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

Example 9

Find the variance and standard deviation of the random variable X which has the probability density function f with rule:

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Solution

Use the computational formula $\text{Var}(X) = \text{E}(X^2) - \mu^2$.

First evaluate $\text{E}(X^2)$:

$$\begin{aligned}\text{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \\ &= \int_0^2 x^2 \times 0.5x\,dx \\ &= 0.5 \int_0^2 x^3\,dx \\ &= 0.5 \left[\frac{x^4}{4}\right]_0^2 \\ &= 0.5 \times 4 \\ &= 2\end{aligned}$$

Since $\text{E}(X) = \frac{4}{3}$ from Example 5, we now have

$$\text{Var}(X) = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$$

and $\quad \text{sd}(X) = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3} = 0.471$ (correct to three decimal places)

ISBN 978-1-009-11049-5 Cambridge University Press

Interquartile range

The **interquartile range** is the range of the middle 50% of the distribution; it is the difference between the 75th percentile (also known as Q3) and the 25th percentile (also known as Q1).

Example 10

Determine the interquartile range of the random variable X which has the probability density function:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

Solution

To find the 25th percentile a, solve:

$$\int_0^a 2x\,dx = 0.25$$
$$[x^2]_0^a = 0.25$$
$$a^2 = 0.25$$
$$\therefore \quad a = \sqrt{0.25} = 0.5$$

To find the 75th percentile b, solve:

$$\int_0^b 2x\,dx = 0.75$$
$$[x^2]_0^b = 0.75$$
$$b^2 = 0.75$$
$$\therefore \quad b = \sqrt{0.75} \approx 0.866$$

Thus the interquartile range is $0.866 - 0.5 = 0.366$, correct to three decimal places.

Note that the negative solutions to these equations were not appropriate, as $0 \le x \le 1$.

Exercise 15C

Example 9

1 A random variable X has probability density function:

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & x \le 0 \text{ or } x \ge 1 \end{cases}$$

Find the variance of X, and hence find the standard deviation of X.

Example 10

2 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

a Find a such that $\Pr(X \le a) = 0.25$.
b Find b such that $\Pr(X \le b) = 0.75$.
c Find the interquartile range of X.

3 A random variable X has the probability density function given by

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

a Sketch the graph of $y = f(x)$.

ISBN 978-1-009-11049-5

b Find the interquartile range of X, giving your answer correct to three decimal places.

4 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x} & 1 \le x \le 9 \\ 0 & x < 1 \text{ or } x > 9 \end{cases}$$

a Find the value of k.

b Find the mean and variance of X, giving your answer correct to three decimal places.

5 A continuous random variable X has density function f given by

$$f(x) = \begin{cases} 0 & x < 0 \\ 2 - 2x & 0 \le x \le 1 \\ 0 & x > 1 \end{cases}$$

a Find the interquartile range of X.

b Find the mean and variance of X.

6 A random variable X has probability density function f with the rule:

$$f(x) = \begin{cases} 0 & x < 0 \\ 2xe^{-x^2} & x \ge 0 \end{cases}$$

Find the interquartile range of X.

7 A random variable X has a probability density function given by

$$f(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \le x < 2 \\ 0 & x \ge 2 \end{cases}$$

a Find the interquartile range of X.

b Find the mean and variance of X.

8 The queuing time, X minutes, of a traveller at the ticket office of a large railway station has probability density function f defined by

$$f(x) = \begin{cases} kx(100 - x^2) & 0 \le x \le 10 \\ 0 & x > 10 \text{ or } x < 0 \end{cases}$$

a Find the value of k.

b Find the mean of the distribution.

c Find the standard deviation of the distribution, correct to two decimal places.

9 A probability density function is given by

$$f(x) = \begin{cases} k(a^2 - x^2) & -a \le x \le a \\ 0 & x > a \text{ or } x < -a \end{cases}$$

a Find k in terms of a.

b Find the value of a which gives a standard deviation of 2.

ISBN 978-1-009-11049-5

Cambridge University Press

10 A continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} k(3-x) & 0 \le x \le 3 \\ k(x-3) & 3 < x \le 6 \\ 0 & x > 6 \text{ or } x < 0 \end{cases}$$

where k is a constant.

a Sketch the graph of f.

b Hence, or otherwise, find the value of k.

c Verify that the mean of X is 3.

d Find $\text{Var}(X)$.

15D Properties of mean and variance*

It has already been stated that the expected value of a function of X is not necessarily equal to that function of the expected value of X. That is, in general,

$$\text{E}[g(X)] \ne g[\text{E}(X)]$$

An exception is the case where the function g is linear: the mean of a linear function of X is equal to the linear function of the mean of X.

The mean and variance of $aX + b$

> For any continuous random variable X,
>
> $$\text{E}(aX + b) = a\text{E}(X) + b$$

Proof The validity of this statement can be readily demonstrated:

$$\begin{aligned} \text{E}(aX + b) &= \int_{-\infty}^{\infty} (ax + b) f(x)\, dx \\ &= \int_{-\infty}^{\infty} ax f(x)\, dx + \int_{-\infty}^{\infty} b f(x)\, dx \\ &= a \int_{-\infty}^{\infty} x f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx \\ &= a\text{E}(X) + b \qquad \left(\text{since } \int_{-\infty}^{\infty} f(x)\, dx = 1\right) \end{aligned}$$

We can also obtain a formula for the variance of a linear function of X.

> For any continuous random variable X,
>
> $$\text{Var}(aX + b) = a^2\text{Var}(X)$$

* This section is not required for Mathematical Methods Units 3 & 4.

ISBN 978-1-009-11049-5

Proof Consider the variance of a linear function of X:

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

Now $\quad [\text{E}(aX + b)]^2 = [a\text{E}(X) + b]^2 = (a\mu + b)^2 = a^2\mu^2 + 2ab\mu + b^2$

and $\quad \text{E}[(aX + b)^2] = \text{E}(a^2X^2 + 2abX + b^2)$

$$= a^2\text{E}(X^2) + 2ab\mu + b^2$$

Thus $\quad \text{Var}(aX + b) = a^2\text{E}(X^2) + 2ab\mu + b^2 - a^2\mu^2 - 2ab\mu - b^2$

$$= a^2\text{E}(X^2) - a^2\mu^2$$

$$= a^2\text{Var}(X)$$

Although initially the absence of b in the variance may seem surprising, on reflection it makes sense that adding a constant merely translates the probability density function, and has no effect on its spread.

Example 11

Suppose that X is a continuous random variable with mean $\mu = 10$ and variance $\sigma^2 = 2$.

a Find $\text{E}(2X + 1)$.

b Find $\text{Var}(1 - 3X)$.

Solution

a $\text{E}(2X + 1) = 2\text{E}(X) + 1$

$= 2 \times 10 + 1$

$= 21$

b $\text{Var}(1 - 3X) = (-3)^2\text{Var}(X)$

$= 9 \times 2$

$= 18$

The probability density function of $aX + b$

The random variable $X + b$ If the probability density function of X has rule $f(x)$, then the probability density function of $X + b$ is obtained by the translation $(x, y) \to (x + b, y)$ and so has rule $f(x - b)$.

The random variable aX Similarly, multiplying by a is similar to a dilation of factor a from the y-axis. However, there has to be an adjustment to determine the rule for the probability density function of aX, as the transformation must be area-preserving. The rule is $\frac{1}{a}f\left(\frac{x}{a}\right)$.

The random variable $aX + b$ Thus, if the probability density function of X has rule $f(x)$, then the probability density function of $aX + b$ has rule $\frac{1}{a}f\left(\frac{x - b}{a}\right)$. The transformation is described by

$$(x, y) \to \left(ax + b, \frac{y}{a}\right)$$

In the case that a and b are positive, this is a dilation of factor a from the y-axis and factor $\frac{1}{a}$ from the x-axis, followed by a translation of b units in the positive direction of the x-axis. It

ISBN 978-1-009-11049-5 Cambridge University Press

is also important to consider the effect of the transformation on the interval within which the transformed probability density function applies, as shown in the following example.

Example 12

Suppose that X is a continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{2(x-1)}{25} & 1 \le x \le 6 \\ 0 & elsewhere \end{cases}$$

Find the probability density function g for $2X + 3$.

Solution

The transformation is described by $(x, y) \to \left(2x + 3, \dfrac{y}{2}\right)$ and therefore $(x, 0) \to (2x + 3, 0)$ for all $x \in \mathbb{R}$

Also, $(1, 0) \to (5, 0)$ and $\left(6, \dfrac{2}{5}\right)$ to $\left(15, \dfrac{1}{5}\right)$

The graph with the rule $y = \dfrac{2(x-1)}{25}$ is mapped to the graph with rule

$$y = \frac{1}{2} \times \frac{2\left(\dfrac{x-3}{2} - 1\right)}{25} = \frac{x-5}{50}$$

Thus we can say

$$g(x) = \begin{cases} \dfrac{x-5}{50} & 5 \le x \le 15 \\ 0 & elsewhere \end{cases}$$

Exercise 15D

1 The amount of flour used each day in a bakery is a continuous random variable X with a mean of 4 tonnes and a variance of 0.25 tonne. The cost of the flour used is $C = 300X + 100$.

a Find $\mathrm{E}(C)$.

b Find $\mathrm{Var}(C)$.

Example 12

2 Let X be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} x^2\left(2x + \dfrac{3}{2}\right) & \text{if } 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Find $\mathrm{E}(X)$.

b Hence find $\mathrm{E}(V)$, where $V = 2X + 3$.

ISBN 978-1-009-11049-5 Cambridge University Press

3 A random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{10}{x^2} & 5 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

Suppose that U is a random variable, and $U = 3X + 25$.

a Find $\mathrm{E}(U)$ giving your answer correct to three decimal places.

b Find $\mathrm{Var}(U)$ giving your answer correct to three decimal places.

4 For certain glass ornaments, the proportion of impurities per ornament, X, is a random variable with a density function given by

$$f(x) = \begin{cases} \dfrac{3x^2}{2} + x & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x < 0 \text{ or } x > 1 \end{cases}$$

The value of each ornament (in dollars) is $V = 100 - 1.5X$.

a Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

b Hence find the mean and standard deviation of V.

5 Let X be a random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{3x^2}{2} & \text{if } -1 \le x \le 1 \\ 0 & \text{if } x < -1 \text{ or } x > 1 \end{cases}$$

Find:

a $\mathrm{E}(3X)$ and $\mathrm{Var}(3X)$

b $\mathrm{E}(3 - X)$ and $\mathrm{Var}(3 - X)$

c $\mathrm{E}(3X + 1)$ and $\mathrm{Var}(3X + 1)$

d the rule of a probability density function for $3X$

e the rule of a probability density function for $3X + 1$.

15E Cumulative distribution functions

Another function of importance in describing a continuous random variable is the **cumulative distribution function** or **CDF**. An understanding of the cumulative density function is not required by the study design, but because if its usefulness is included here as an optional section. For a continuous random variable X, with probability density function f defined on the interval $[c, d]$, the cumulative distribution function F is given by

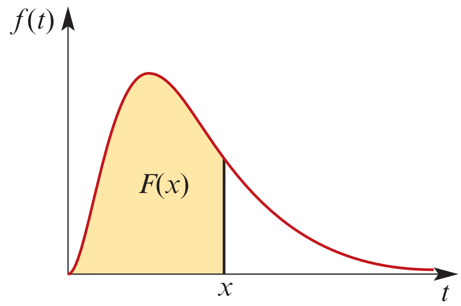
$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \int_c^x f(t)\, dt \end{aligned}$$

where t is the variable of integration. The cumulative distribution function at a particular value x gives the probability that the random variable X takes a value less than or equal to x.

ISBN 978-1-009-11049-5 Cambridge University Press

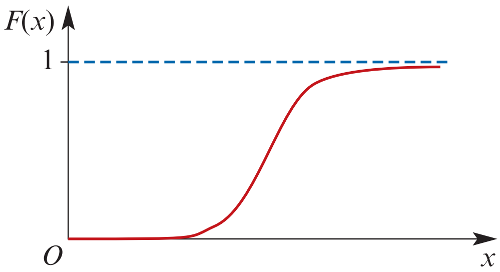
The diagram on the right shows the relationship between the probability density function f and the cumulative distribution function F.

The function F describes the area under the graph of the probability density function between the lower bound of the domain of f and x. (In the diagram, the lower bound is 0.)



For every continuous random variable X, the cumulative distribution function F is continuous.

Using the general version of the fundamental theorem of calculus, it can be shown that the derivative of the cumulative distribution function is the density function. More precisely, we have $F'(x) = f(x)$, for each value of x at which f is continuous.



There are three important properties of a cumulative distribution function. For a continuous random variable X with range $[c, d]$:

1. The probability that X takes a value less than or equal to c is 0. That is, $F(c) = 0$.
2. The probability that X takes a value less than or equal to d is 1. That is, $F(d) = 1$.
3. If x_1 and x_2 are values of X with $x_1 \le x_2$, then $\Pr(X \le x_1) \le \Pr(X \le x_2)$. That is,

   $$x_1 \le x_2 \quad \text{implies} \quad F(x_1) \le F(x_2)$$

   The function F is a **non-decreasing** function.

For a probability density function f defined on $\mathbb{R}$, the cumulative distribution is given by

$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \int_{-\infty}^{x} f(t)\, dt \end{aligned}$$

In this case, we have $F(x) \to 0$ as $x \to -\infty$, and $F(x) \to 1$ as $x \to \infty$.

Example 13

The time, X seconds, that it takes a student to complete a puzzle is a random variable with density function given by

$$f(x) = \begin{cases} \dfrac{5}{x^2} & x \ge 5 \\ 0 & x < 5 \end{cases}$$

Find $F(x)$, the cumulative distribution function of X.

ISBN 978-1-009-11049-5

Solution

$$F(x) = \int_5^x f(t)\,dt = \int_5^x \frac{5}{t^2}\,dt$$
$$= \left[\frac{-5}{t}\right]_5^x$$
$$= \frac{-5}{x} + 1$$

Thus $F(x) = 1 - \dfrac{5}{x}$ for $x \geq 5$.

The importance of the cumulative distribution function is that probabilities for various intervals can be computed directly from $F(x)$.

Exercise 15E

Example 13

1 The probability density function for a random variable X is given by

$$f(x) = \begin{cases} \frac{1}{5} & 0 < x < 5 \\ 0 & x \leq 0 \text{ or } x \geq 5 \end{cases}$$

a Find $F(x)$, the cumulative distribution function of X. **b** Hence find $\Pr(X \leq 3)$.

2 A continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{4} & \text{if } 0 \leq x < 1 \\ \frac{x^3}{5} & \text{if } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find $F(x)$, the cumulative distribution function of X.
b Solve the equation $F(x) = 0.5$ for x.

3 A random variable X has the cumulative distribution function with rule:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x^2} & x \geq 0 \end{cases}$$

a Sketch the graph of $y = F(x)$. **b** Find $\Pr(X \geq 2)$.
c Find $\Pr(X \geq 2 \mid X < 3)$.

4 The continuous random variable X has cumulative distribution function F given by

$$F(x) = \begin{cases} 0 & x < 0 \\ kx^2 & 0 \leq x \leq 6 \\ 1 & x > 6 \end{cases}$$

a Determine the value of the constant k. **b** Calculate $\Pr\left(\frac{1}{2} \leq X \leq 1\right)$.

ISBN 978-1-009-11049-5

5 The cumulative distribution function of a random variable X is given by

$$F(x) = \begin{cases} 1 - \dfrac{10}{x} & x > 10 \\ 0 & x \le 10 \end{cases}$$

Use the cumulative distribution function to determine:

a $\Pr(X < 30)$

b Find m such that $\Pr(X < m) = 0.5$.

c Find a and b such that $\Pr(a \le X \le b) = 0.95$ assuming that $\Pr(X \le a) = \Pr(X \ge b)$.

6 The cumulative distribution function of a continuous random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 4x^3 - 3x^4 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Find a probability density function for X.

7 A continuous random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - (1 - x)^5 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Find a probability density function for X.

8 Let X be a random variable with the cumulative distribution function

$$F(x) = \begin{cases} 0.5e^{x} & x \le 0 \\ 1 - 0.5e^{-x} & x > 0 \end{cases}$$

Find a probability density function for X.

ISBN 978-1-009-11049-5

Chapter summary

Assign-ment

Nrich

- A **continuous random variable** is one that can take any value in an interval of the real number line.
- A continuous random variable can be described by a **probability density function** f. There are many different probability density functions with different shapes and properties. However, they all have the following two fundamental properties:

  1 For any value of x, the value of $f(x)$ is non-negative. That is,

  $$f(x) \geq 0 \text{ for all } x$$

  2 The total area enclosed by the graph of f and the x-axis is equal to 1. That is,

  $$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

  The probability of X taking a value in the interval (a, b) is found by determining the area under the probability density curve between a and b. That is,

  $$\Pr(a < X < b) = \int_{a}^{b} f(x)\, dx$$

- The **mean** or **expected value** of a continuous random variable X with probability density function f is given by

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

  provided the integral exists.
- If $g(X)$ is a function of X, then the expected value of $g(X)$ is given by

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

  provided the integral exists. In general, $\mathrm{E}[g(X)] \neq g[\mathrm{E}(X)]$.
- The **median** of a continuous random variable X is the value m such that

  $$\int_{-\infty}^{m} f(x)\, dx = 0.5$$

- The **variance** of a continuous random variable X with probability density function f is defined by

  $$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

  provided the integral exists. To calculate the variance, use

  $$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2$$

- The **standard deviation** of X is defined by

  $$\sigma = \mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

- Linear function of a continuous random variable:

  $$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

  $$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

ISBN 978-1-009-11049-5 Cambridge University Press

Review

- The **interquartile range** of X is

$$\text{IQR} = b - a$$

where a and b are such that

$$\int_{-\infty}^{a} f(x)\,dx = 0.25 \quad \text{and} \quad \int_{-\infty}^{b} f(x)\,dx = 0.75$$

and where f is the probability density function of X.

Technology-free questions

1 The probability density function of X is given by

$$f(x) = \begin{cases} kx & \text{if } 1 \le x \le \sqrt{2} \\ 0 & \text{otherwise} \end{cases}$$

a Find k.

b Find $\Pr\left(1 < X < \frac{11}{10}\right)$.

c Find $\Pr\left(1 < X < \frac{6}{5}\right)$.

2 If the probability density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & 0 \le x \le 1 \\ 0 & x > 1 \text{ or } x < 0 \end{cases}$$

and $\text{E}(X) = \dfrac{2}{3}$, find a and b.

3 The probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{\sin x}{2} & 0 \le x \le \pi \\ 0 & x > \pi \text{ or } x < 0 \end{cases}$$

Find c such that $\Pr(X > c) = 0.8$

4 The probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{1}{4} & 1 \le x < 5 \\ 0 & x < 1 \text{ or } x \ge 5 \end{cases}$$

a Find $\Pr(1 < X < 3)$.

b Find $\Pr(X > 2 \mid 1 < X < 3)$.

c Find $\Pr(X > 4 \mid X > 2)$.

5 Consider the random variable X having the probability density function given by

$$f(x) = \begin{cases} 12x^2(1 - x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(x)$.

b Find $\Pr(X < 0.5)$ and illustrate this probability on your sketch graph.

ISBN 978-1-009-11049-5 Cambridge University Press

6 The probability density function of a random variable X is

$$f(x) = \begin{cases} kx^2(1-x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Determine k.

b Find the probability that X is less than $\frac{2}{3}$.

c Find the probability that X is less than $\frac{1}{3}$, given that X is less than $\frac{2}{3}$.

7 Let X be a continuous random variable with probability density function:

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Find $\Pr(X < 0.2)$.

b Find $\Pr(X < 0.2 \mid X < 0.3)$.

8 A continuous random variable X has probability density function:

$$f(x) = \begin{cases} \dfrac{\pi}{4}\cos\left(\dfrac{\pi x}{4}\right) & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the median value, m, of X.

9 The probability density function f of a random variable X is given by

$$f(x) = \begin{cases} \dfrac{x+2}{16} & \text{if } 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

a Find $\mathrm{E}(X)$.

b Find a such that $\Pr(X \le a) = \dfrac{5}{32}$.

10 The probability density function f of a random variable X is given by

$$f(x) = \begin{cases} c(1-x^2) & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Find c.

b Find $\mathrm{E}(X)$.

11 Show that

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

is a probability function, where the constant n is a natural number.

12 The probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{1}{x} & 1 \le x \le e \\ 0 & x > e \text{ or } x < 1 \end{cases}$$

a Find m, the value of X such that $\Pr(X < m) = 0.5$.

b Find b such that $\Pr(X > b) = \dfrac{1}{4}$.

ISBN 978-1-009-11049-5 Cambridge University Press

13 Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} x\sin x^2 & \text{if } 0 < x < \sqrt{\pi} \\ 0 & \text{elsewhere} \end{cases}$$

a Show by differentiation that $-\dfrac{1}{2}\cos x^2$ is an antiderivative of $x\sin x^2$.

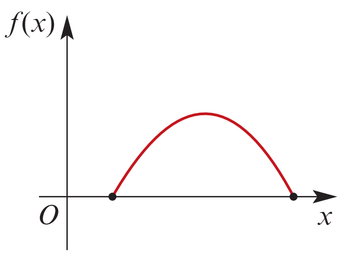
b Calculate $\Pr\left(\sqrt{\frac{\pi}{3}} < X < \sqrt{\frac{\pi}{2}}\right)$.

c Find m, the value of X such that $\Pr(X < m) = 0.5$.

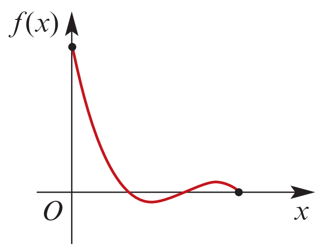
Multiple-choice questions

1 Which of the following graphs could *not* represent a probability density function f?

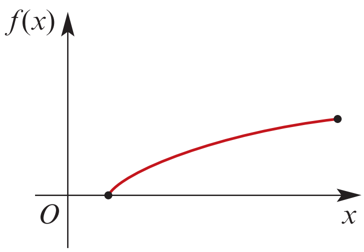
A



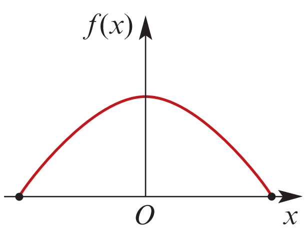
B



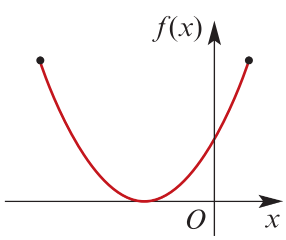
C



D



E



2 If the function $f(x) = 4x$ represents a probability density function, then which of the following could be the domain of f?

A $0 \le x \le 0.25$ **B** $0 \le x \le 0.5$ **C** $0 \le x \le 1$

D $0 \le x \le \dfrac{1}{\sqrt{2}}$ **E** $\dfrac{1}{\sqrt{2}} \le x \le \dfrac{2}{\sqrt{2}}$

3 If a random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{2}\sin x & 0 < x < k \\ 0 & x \ge k \text{ or } x \le 0 \end{cases}$$

then k is equal to

A 1 **B** $\dfrac{\pi}{2}$ **C** 2 **D** π **E** 2π

ISBN 978-1-009-11049-5

Cambridge University Press

The following information relates to Questions 4, 5 and 6.

A random variable X has probability density function:

$$f(x) = \begin{cases} \frac{3}{4}(x^2 - 1) & 1 < x < 2 \\ 0 & x \le 1 \text{ or } x \ge 2 \end{cases}$$

4 $\Pr(X \le 1.3)$ is closest to

A 0.0743 **B** 0.4258 **C** 0.3 **D** 0.25 **E** 0.9258

5 The mean, $\mathrm{E}(X)$, of X is equal to

A 1 **B** $\frac{3}{2}$ **C** $\frac{9}{4}$ **D** $\frac{27}{32}$ **E** $\frac{27}{16}$

6 The variance of X is

A $\frac{27}{16}$ **B** $\frac{67}{1280}$ **C** $\frac{81}{16}$ **D** $\frac{81}{256}$ **E** $\frac{729}{256}$

7 If a random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{x^3}{4} & 0 \le x \le 2 \\ 0 & x > 2 \text{ or } x < 0 \end{cases}$$

then the value of m such that $\Pr(X < m) = 0.5$ is closest to

A 1.5 **B** 1.4142 **C** 1.6818 **D** 1.2600 **E** 1

8 If a random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{3}{2}(x - 1)(x - 2)^2 & 1 \le x \le 3 \\ 0 & x < 1 \text{ or } x > 3 \end{cases}$$

then the mean of X is

A 1 **B** 1.333 **C** 2 **D** 2.6 **E** 3

9 If the consultation time (in minutes) at a surgery is represented by a random variable X which has probability density function

$$f(x) = \begin{cases} \frac{x}{40\,000}(400 - x^2) & 0 \le x \le 20 \\ 0 & x < 0 \text{ or } x > 20 \end{cases}$$

then the expected consultation time (in minutes) for three patients is

A $10\frac{2}{3}$ **B** 30 **C** 32 **D** 42 **E** $43\frac{2}{3}$

10 The top 10% of students in an examination will be awarded an ‘A’. If the distribution of scores on the examination is a random variable X with probability density function

$$f(x) = \begin{cases} \frac{\pi}{100}\sin\left(\frac{\pi x}{50}\right) & 0 \le x \le 50 \\ 0 & x < 0 \text{ or } x > 50 \end{cases}$$

then the minimum score required to be awarded an ‘A’ is closest to

A 40 **B** 41 **C** 42 **D** 43 **E** 44

ISBN 978-1-009-11049-5 Cambridge University Press

Review

11 The cumulative distribution function gives the probability

A that a random variable takes a particular value

B that a random variable takes a value less than or equal to a particular value

C that a random variable takes a value more than a particular value

D of two or more events occurring at once

E that a random variable takes a particular value given that another event has occurred

12 Suppose that X is a continuous random variable with cumulative distribution function given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{8}x & \text{if } 0 \le x < 1 \\ \frac{1}{8}(3x-2) & \text{if } 1 \le x < 2 \\ \frac{1}{2}(x-1) & \text{if } 2 \le x < 3 \\ 1 & \text{if } x \ge 3 \end{cases}$$

Then $\Pr(1 < X < 2.5)$ is equal to

A $\frac{1}{8}$ **B** $\frac{5}{16}$ **C** $\frac{5}{8}$ **D** $\frac{3}{4}$ **E** $\frac{7}{8}$

Extended-response questions

1 The continuous random variable X has probability density function f, where

$$f(x) = \begin{cases} \dfrac{k}{12(x+1)^3} & 0 \le x \le 4 \\ 0 & x < 0 \text{ or } x > 4 \end{cases}$$

a Find k.

b Evaluate $E(X+1)$. Hence, find the mean of X.

c Use your calculator to verify your answer to part **b**.

d Find the value of $c > 0$ for which $\Pr(X \le c) = c$.

2 The distribution of X, the life of a certain electronic component in hours, is described by the following probability density function:

$$f(x) = \begin{cases} \dfrac{a}{100}\left(1 - \dfrac{x}{100}\right) & 100 < x < 1000 \\ 0 & x \le 100 \text{ and } x \ge 1000 \end{cases}$$

a What is the value of a?

b What is the probability that a component will last longer than 950 hours? Give your answer correct to four decimal places.

c Find the expected value of the life of the components.

ISBN 978-1-009-11049-5 Cambridge University Press

d Find the the minimum life for the longest lasting 50% of components.

e Suppose that the components are sold in boxes of 50, and that the probability of any component in the box lasting less than 950 hours is independent of the probability of any other component in the box lasting less than 950 hours. What is the probability that no more that one component in the box lasts less than 950 hours? Give your answer correct to three decimal places.

3 A factory has two machines that manufacture pipes of a certain length, Machine A and Machine B. The error (in mm) in the length of the pipes produced by Machine A is a random variable wth probability density function given by

$$a(x) = \begin{cases} \frac{\pi}{20}\cos\left(\frac{\pi(x-6)}{10}\right) & 1 \le x \le 11 \\ 0 & elsewhere \end{cases}$$

a Find, correct to four decimal places, the probability that the error in length of a pipe randomly chose from Machine A is more than 10mm.

b The probability that the error in length is more than c mm is 0.01. Find the value of c correct to two decimal places.

The error (in mm) in the length of the pipes produced by Machine B is also a random variable wth probability density function given by

$$b(x) = \begin{cases} \frac{4}{2187}(x-1)^2(10-x) & 1 \le x \le 10 \\ 0 & elsewhere \end{cases}$$

c Determine the mean error in pipes produced by Machine B.

d The probability that the error in length is more than k mm is 0.05. Find the value of k, correct to two decimal places.

In this factory 2000 pipes are produced each day, 1200 from Machine A and 800 from Machine B. A pipe is deemed to be unacceptable if the error in length is more than 9.5mm.

e Find, correct to three decimal places. the probability that a pipe selected from the combined daily production of rods is unacceptable.

f Find the probability that a pipe selected from the combined daily production of rods and found to be unacceptable was produced by Machine A. Give your answer correct to three decimal places.

4 The seasonal yield of a variety of strawberries (in kgs per plant) has probability density function:

$$f(x) = \begin{cases} \frac{1}{5}x & 0 \le x < 2 \\ \frac{1}{15}(10-2x) & 2 \le x \le 5 \\ 0 & x < 0 \text{ or } x > 5 \end{cases}$$

a Find correct to four decimal places $\Pr(1 < X < 3)$.

b Find correct to four decimal places $\Pr(X \ge 1.5 \mid X < 3)$.

ISBN 978-1-009-11049-5 Cambridge University Press

c Determine the mean yield for this variety of strawberry plant.

d The lowest yielding strawberry plants are pulled out at the end of the season, and replaced with new plants.

i If 10% of the plants are to be removed, what is minimum yield produced by a strawberry plant which is not removed?

ii Again, assume 10% of the plants are to be removed. Suppose that the strawberry plants are grown in beds of 20 plants, and that the yield of each plant is independent of the yield of any other plan. What is the probability that more than one plant will be removed from a randomly selected bed? Give your answer correct to four decimal places.

5 The continuous random variable X has the probability density function f, where

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

By first expanding $(X - c)^2$, or otherwise, find two values of c such that

$$\mathrm{E}[(X - c)^2] = \frac{2}{3}$$

6 The yield of a variety of corn has probability density function:

$$f(x) = \begin{cases} kx & 0 \le x < 2 \\ k(4 - x) & 2 \le x \le 4 \\ 0 & x < 0 \text{ or } x > 4 \end{cases}$$

a Find k.

b Find the expected value, μ, and the variance of the yield of corn.

c Find the probability $\Pr(\mu - 1 < X < \mu + 1)$.

d Find the value of a such that $\Pr(X > a) = 0.6$, giving your answer correct to one decimal place.

ISBN 978-1-009-11049-5 Cambridge University Press