Supporting Australian Mathematics Project

A guide for teachers – Years 11 and 12

Probability and statistics: Module 21

Continuous probability distributions

Education Services Australia

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Continuous probability distributions – A guide for teachers (Years 11–12)

Professor Ian Gordon, University of Melbourne

Editor: Dr Jane Pitkethly, La Trobe University

Illustrations and web design: Catherine Tan, Michael Shaw

Full bibliographic details are available from Education Services Australia.

Published by Education Services Australia
PO Box 177
Carlton South Vic 3053
Australia

Tel: (03) 9207 9600
Fax: (03) 9910 9800
Email: info@esa.edu.au
Website: www.esa.edu.au

This publication is funded by the Australian Government Department of Education, Employment and Workplace Relations.

Supporting Australian Mathematics Project

Australian Mathematical Sciences Institute
Building 161
The University of Melbourne
VIC 3010
Email: enquiries@amsi.org.au
Website: www.amsi.org.au

Continuous probability distributions

Assumed knowledge

- The content of the modules:
  - *Probability*
  - *Discrete probability distributions*
  - *Binomial distribution*.
- A basic understanding of integration, as discussed in the module *Integration*.

Motivation

The module *Discrete probability distributions* introduces the fundamentals of random variables, noting that they are the numerical outcome of a random procedure.

Most of the examples considered in that module involve counts of some sort: the number of things, or people, or occurrences, and so on. When we count, the outcome is definitely discrete: it can only take integer values, and not other numerical values.

However, many measured variables are not like this. Rather, they take a value in a specified range (for example, a variable might be positive), but within that range they can take any numerical value. The paradigm phenomenon in this category is 'time'. If we ask the question 'How long does it take?' (to complete a crossword, to brush your teeth, for a person to die after a diagnosis of an illness, to run 10 kilometres), then the answer can be given to varying levels of accuracy, and we are really only limited by the precision of our instruments. We feel that, in principle, the time could be any numerical value. For example, the time take to run 10 kilometres could be 50 minutes and 23.1 seconds. But it could also be 50 minutes and 23.08 seconds, or 50 minutes and 23.082 seconds, and so on.

Many other variables are measured on a continuum like this. These variables include height, weight, blood pressure, temperature, distance, speed and many others. We need a way to represent the probability distribution of such continuous variables, and the purpose of this module is to describe this.

There are different ways to describe the probability distribution of a continuous random variable. In this module, we introduce the cumulative distribution function and the probability density function. We shall see that probabilities associated with a continuous random variable are given by integrals. This module also covers the mean and variance of a continuous random variable.

Content

Continuous random variables: basic ideas

Random variables have been introduced in the module *Discrete probability distributions*. Recall that a random variable is a variable whose value is determined by the outcome of a random procedure.

There are two main types of random variables: discrete and continuous. The modules *Discrete probability distributions* and *Binomial distribution* deal with discrete random variables; we now turn our attention to the second type, continuous random variables.

A **continuous random variable** is one that can take any real value within a specified range.

A discrete random variable takes some values and not others; we cannot obtain a value of 4.73 when rolling a fair die. By contrast, a continuous random variable can take any value, in principle, within a specified range.

We have already seen examples of continuous random variables, when the idea of a random variable was first introduced.

Example: Five people born in 1995

Five babies born in 1995 are followed up over their lives, and major health and milestone events are recorded. Here are two continuous random variables that could be defined:

- Let W be the average height of the five people at age 18. Then the value of W must be positive, but there is no obvious upper bound. The common practice in such cases is to say that the possible values are $W > 0$; we will assign extremely low probabilities to large values.
- Let T_i be the total time spent on Facebook by individual i up to age 18. Then T_i in this case is limited by the total time span being considered. If we measure T_i in years, then $0 \le T_i \le 18$; again, values anywhere near the logical maximum of 18 years will be assigned essentially zero probability.

Cumulative distribution functions

The **cumulative distribution function** (**cdf**) of any random variable X is the function $F_X : \mathbb{R} \to [0,1]$ defined by

$$F_X(x) = \Pr(X \le x).$$

Both discrete and continuous random variables have cdfs, although we did not focus on them in the modules on discrete random variables and they are more straightforward to use for continuous random variables.

As noted in the module *Discrete probability distributions*, the use of lower case x as the argument is arbitrary here: if we wrote $F_X(t)$, it would be the same function, determined by the random variable X. But it helps to associate the corresponding lower-case letter with the random variable we are considering.

The cdf is defined for all real values x, sometimes implicitly rather than explicitly. In the example in the previous section, we considered the random variable T_i, the total time spent on Facebook by individual i up to age 18. This is a measurement of time, in years, which must be between 0 and 18. So we know that the cdf of T_i must be zero for any value of $t < 0$. That is, if $t < 0$, then $F_{T_i}(t) = \Pr(T_i \le t) = 0$. At the other extreme, we know that T_i must be less than or equal to 18. So, if $t \ge 18$, then $F_{T_i}(t) = \Pr(T_i \le t) = 1$.

Remember that outcomes for random variables define events in the event space, which is why we are able to assign probabilities to such outcomes. Let X be a random variable with cdf $F_X(x)$. For $a < b$, we can consider the following events:

- $C =$ "$X \le a$"
- $D =$ "$a < X \le b$"
- $E =$ "$X \le b$".

Then C and D are mutually exclusive, and their union is the event E. By the third axiom of probability, this tells us that

$$\begin{aligned}
& \Pr(E) = \Pr(C) + \Pr(D) \\
\implies\quad & \Pr(X \le b) = \Pr(X \le a) + \Pr(a < X \le b) \\
\implies\quad & \Pr(a < X \le b) = \Pr(X \le b) - \Pr(X \le a) \\
\implies\quad & \Pr(a < X \le b) = F_X(b) - F_X(a).
\end{aligned}$$

The cumulative distribution function $F_X(x)$ of a random variable X has three important properties:

1 The cumulative distribution function $F_X(x)$ is a non-decreasing function. This follows directly from the result we have just derived: For $a < b$, we have

$$\Pr(a < X \leq b) \geq 0 \implies F_X(b) - F_X(a) \geq 0 \implies F_X(a) \leq F_X(b).$$

2 As $x \to -\infty$, the value of $F_X(x)$ approaches 0 (or equals 0). That is, $\lim_{x\to-\infty} F_X(x) = 0$. This follows in part from the fact that $\Pr(\varnothing) = 0$.

3 As $x \to \infty$, the value of $F_X(x)$ approaches 1 (or equals 1). That is, $\lim_{x\to\infty} F_X(x) = 1$. This follows in part from the fact that $\Pr(\mathscr{E}) = 1$.

All of the above discussion applies equally to discrete and continuous random variables.

We now turn specifically to the cdf of a continuous random variable. Its form is something like that shown in the following figure. We require a continuous random variable to have a cdf that is a continuous function.

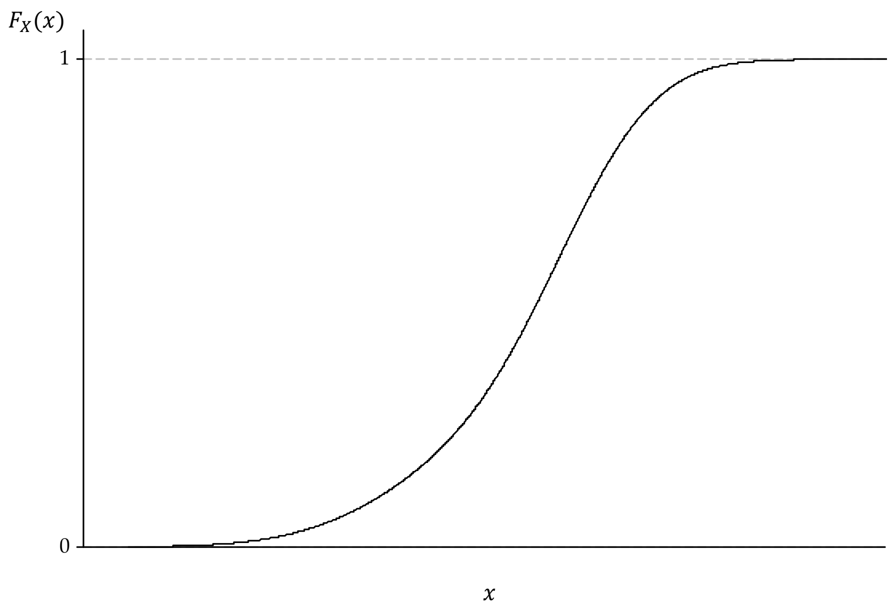

Figure 1: The general appearance of the cumulative distribution function of a continuous random variable.

We now use the cdf a continuous random variable to start to think about the question of probabilities for continuous random variables. For discrete random variables, probabilities come directly from the probability function $p_X(x)$: we identify the possible discrete values that the random variable X can take, and then specify somehow the probability for each of these values.

For a continuous random variable X, once we know its cdf $F_X(x)$, we can find the probability that X lies in any given interval:

$$\Pr(a < X \leq b) = F_X(b) - F_X(a).$$

But what if we are interested in the probability that a continuous random variable takes a specific single value? What is $\Pr(X = x)$ for a continuous random variable?

We can address this question by considering the probability that X lies in an interval, and then shrinking the interval to a single point. Formally, for a continuous random variable X with cdf $F_X(x)$:

$$\begin{aligned}
\Pr(X = x) &\leq \lim_{h\to 0^+} \Pr(x - h < X \leq x) \\
&= \lim_{h\to 0^+} \big(F_X(x) - F_X(x-h)\big) \\
&= F_X(x) - F_X(x) && \text{(since } F_X \text{ is continuous)} \\
&= 0.
\end{aligned}$$

Hence, $\Pr(X = x) = 0$. This is a somewhat disconcerting result: It seems to be saying that we can never really observe a continuous random variable taking a specific value, since the probability of observing any value is zero.

In fact, this is true. Continuous random variables are really abstractions: what we observe is always rounded in some way. It helps to think about some specific cases.

Example: Random numbers

Suppose we consider real numbers randomly chosen between 0 and 1, for which we record X_1, the random number truncated to one decimal place. For example, the number 0.07491234008 is recorded as 0.0 (as the first decimal place is zero). Note that this means we are not rounding, but truncating. If the mechanism generating these numbers has no preference for any position in the interval $(0, 1)$, then the distribution of the numbers we obtain will be such that

$$\Pr(X_1 = 0.0) = \Pr(X_1 = 0.1) = \cdots = \Pr(X_1 = 0.9) = \tfrac{1}{10}.$$

This is a discrete random variable, with the same probability for each of the ten possible outcomes.

Now suppose that instead we record X_2, the random number truncated to two decimal places (again, not rounding). For example, if the real number is 0.9790295134, we record 0.97. This random variable is also discrete, with

$$\Pr(X_2 = 0.00) = \Pr(X_2 = 0.01) = \cdots = \Pr(X_2 = 0.99) = \tfrac{1}{100}.$$

The distributions of X_1 and X_2 are shown in figure 2.

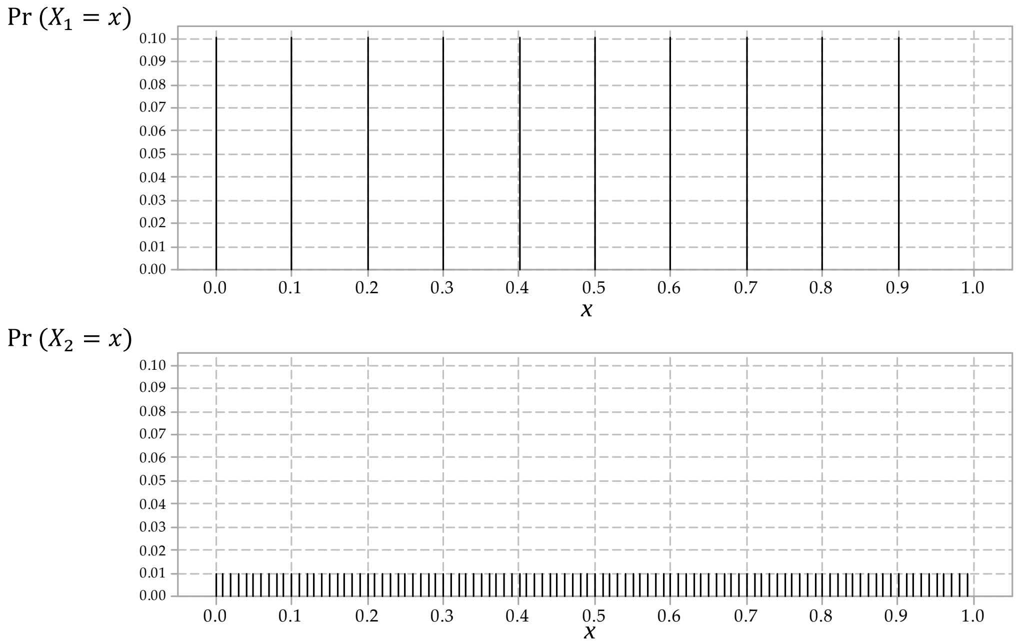

Figure 2: The probability functions $p_{X_1}(x)$ for X_1 and $p_{X_2}(x)$ for X_2.

You can see where this is going: If we record the first k decimal places, the random variable X_k has 10^k possible outcomes, each with the same probability 10^{-k}.

Excel has a function that produces real numbers between 0 and 1, chosen so that there is no preference for any position in the interval $(0, 1)$. If you enter =RAND() in a cell and hit return, you will obtain such a number. Increase the number of decimal places shown in the cell. Keep going until you get a lot of zeroes on the end of the number; you might need to increase the size of the cell. Your spreadsheet should look like figure 3, although of course the specific number will be different ... it is random, after all!

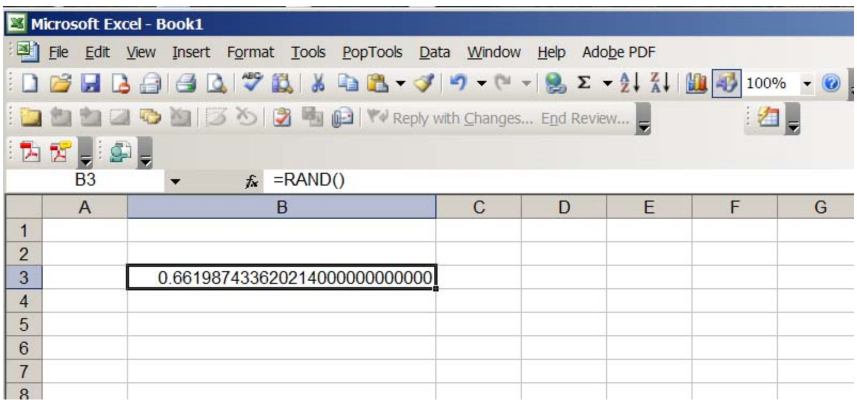

Figure 3: An Excel spreadsheet with a random number from between 0 and 1.

If you hit the key 'F9' repeatedly at this point, you will see a sequence of random numbers, all between 0 and 1. From examining these, it appears that Excel actually produces observations on the random variable X_{15}, the first 15 decimal places of the number. So the chance of any specific one of these numbers occurring is 10^{-15}.

Now consider the chance that each of these discrete random variables $X_1, X_2, X_3, \dots$ takes a value in the interval $[0.3, 0.4)$, for example. We have

$$\Pr(0.3 \le X_1 < 0.4) = \Pr(X_1 = 0.3) = 0.1,$$

$$\begin{aligned}\Pr(0.3 \le X_2 < 0.4) &= \Pr(X_2 = 0.30) + \Pr(X_2 = 0.31) + \cdots + \Pr(X_2 = 0.39)\\ &= 10 \times 0.01 = 0.1,\end{aligned}$$

$$\begin{aligned}\Pr(0.3 \le X_3 < 0.4) &= 10^2 \times 10^{-3} = 0.1,\\ \Pr(0.3 \le X_4 < 0.4) &= 10^3 \times 10^{-4} = 0.1,\\ &\vdots\\ \Pr(0.3 \le X_k < 0.4) &= 10^{k-1} \times 10^{-k} = 0.1,\\ &\vdots\end{aligned}$$

As we make the distribution finer and finer, with more and more possible discrete values, the probability that any of these discrete random variables lies in the interval $[0.3, 0.4)$ is always 0.1.

As k increases, the discrete random variable X_k can be thought of as a closer and closer approximation to a continuous random variable. This continuous random variable, which we label U, has the property that

$$\Pr(a \le U \le b) = b - a, \qquad \text{for } 0 \le a \le b \le 1.$$

For example, $\Pr(0.3 \le U \le 0.4) = 0.4 - 0.3 = 0.1$.

The cdf of this random variable, $F_U(u)$, is therefore very simple. It is

$$F_U(u) = \Pr(U \le u) = \begin{cases} 0 & \text{if } u \le 0,\\ u & \text{if } 0 < u \le 1,\\ 1 & \text{if } u > 1.\end{cases}$$

A random variable with this cdf is said to have a **uniform distribution** on the interval $(0, 1)$; we denote this by $U \stackrel{d}{=} \mathrm{U}(0, 1)$.

The following figure shows the graph of the cumulative distribution function of U.

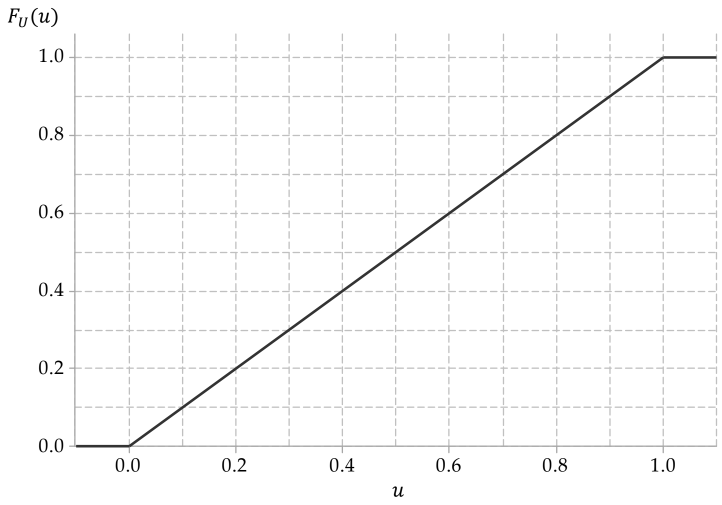

Figure 4: The cumulative distribution function of $U \stackrel{d}{=} \mathrm{U}(0,1)$.

You might have noticed a difference in how the interval between 0.3 and 0.4 was treated in the continuous case, compared to the discrete cases. In all of the discrete cases, the upper limits $0.4, 0.40, 0.400, \dots$ were excluded, while for the continuous random variable it is included. Whenever we are dealing with discrete random variables, whether the inequality is strict or not often matters, and care is needed. For example,

$$\Pr(0.30 \le X_2 \le 0.40) = 11 \times 0.01 = 0.11 \ne 0.1.$$

On the other hand, as we noted earlier in this section, for any continuous random variable X, we have $\Pr(X = x) = 0$. Consequently, for a continuous random variable X and for $a \le b$, we have $\Pr(X = a) = \Pr(X = b) = 0$ and therefore

$$\left.\begin{array}{l} \Pr(a < X < b) \\ \Pr(a < X \le b) \\ \Pr(a \le X < b) \\ \Pr(a \le X \le b) \end{array}\right\} = F_X(b) - F_X(a).$$

The cdf is one way to describe the distribution of a continuous random variable. What about the probability function, as used for discrete random variables? As we have just seen, for a continuous random variable, we have $p_X(x) = \Pr(X = x) = 0$, for all x, so there is no point in using this. In the next section, we look at the appropriate analogue to the probability function for continuous random variables.

Probability density functions

The **probability density function** (**pdf**) $f(x)$ of a continuous random variable X is defined as the derivative of the cdf $F(x)$:

$$f(x) = \frac{d}{dx} F(x).$$

It is sometimes useful to consider the cdf $F(x)$ in terms of the pdf $f(x)$:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt. \qquad (*)$$

The pdf $f(x)$ has two important properties:

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

The first property follows from the fact that the cdf $F(x)$ is non-decreasing and $f(x)$ is its derivative. The second property follows from equation $(*)$ above, since $F(x) \to 1$ as $x \to \infty$, and so the total area under the graph of $f(x)$ is equal to 1.

An infinite variety of shapes are possible for a pdf, since the only requirements are the two properties above. The pdf may have one or several peaks, or no peaks at all; it may have discontinuities, be made up of combinations of functions, and so on. Figure 5 shows a pdf with a single peak and some mild skewness. As is the case for a typical pdf, the value of the function approaches zero as $x \to \infty$ and $x \to -\infty$.

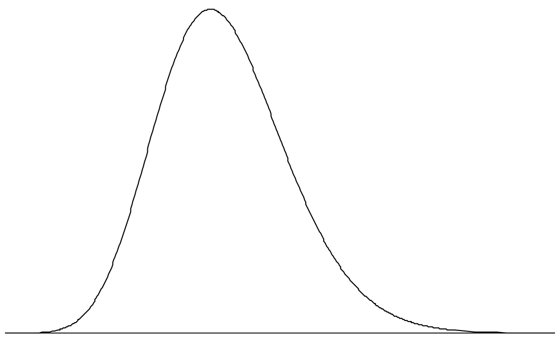

Figure 5: A pdf may look something like this.

We now explore how probabilities concerning the continuous random variable X relate to its pdf. The important result here is that

$$\Pr(a < X \leq b) = \int_{a}^{b} f(x)\, dx = \big[F(x)\big]_{a}^{b}.$$

This result follows from the fact that both sides are equal to $F(b) - F(a)$.

Notes.

- For a continuous random variable, we must consider the probability that it lies in an interval. The importance of this result is that it tells us that, to find the probability, we need to find the area under the pdf on the given interval.
- The total area under the pdf equals 1. So this result tells us that, to approximate the probability that the random variable lies in a given interval, we just have to guess the fraction of the area under the pdf between the ends of the interval.
- This result provides another perspective on why pdfs cannot be negative, since if they were, a negative probability could be obtained, which is impossible.
- The pdf is analogous to, but different from, the probability function (pf) for a discrete random variable. A pf gives a probability, so it cannot be greater than one. A pdf $f(x)$, however, may give a value greater than one for some values of x, since it is not the value of $f(x)$ but the area under the curve that represents probability. On the other hand, the height of the curve reflects the relative probability. If $f(b) = 2f(a)$, then an observation near b is approximately twice as likely as an observation near a.

Exercise 1

Consider the function

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

a Check that $f(x)$ has the two required properties for a pdf, and sketch its graph.

b Suppose that the continuous random variable X has the pdf $f(x)$. Obtain the following probabilities *without calculation*:

- **i** $\Pr(X \le -3)$
- **ii** $\Pr(0 \le X \le 1)$
- **iii** $\Pr(0.5 \le X \le 1)$.

c By looking at the graph of the pdf, guess the value of $\theta = \Pr(0.4 \le X \le 0.7)$. Then check the accuracy of your guess by calculating θ.

d
- **i** Find $f(0.2)$ and $f(0.4)$, and hence obtain $\lambda = \dfrac{f(0.4)}{f(0.2)}$.
- **ii** Find the probability that X is within 0.05 of 0.2. That is, find the probability $p_{0.2} = \Pr(0.15 \le X \le 0.25)$.
- **iii** Find the probability that X is within 0.05 of 0.4. That is, find the probability $p_{0.4} = \Pr(0.35 \le X \le 0.45)$.
- **iv** Confirm that the ratio of these two probabilities is approximately equal to λ. That is, check that $\dfrac{p_{0.4}}{p_{0.2}} \approx \dfrac{f(0.4)}{f(0.2)}$.

Example: Random numbers, continued

Consider the continuous random variable U from the first random-number example. Then $U \stackrel{d}{=} \mathrm{U}(0,1)$. The pdf of U is given by

$$f_U(u) = \begin{cases} 1 & \text{if } 0 < u < 1, \\ 0 & \text{otherwise.} \end{cases}$$

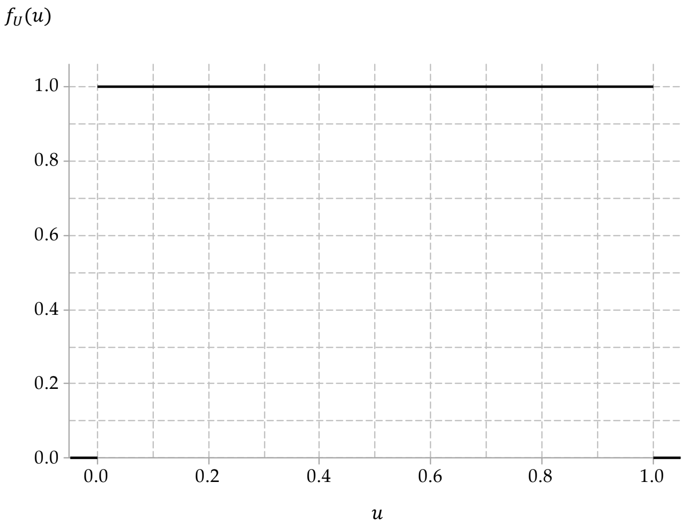

Figure 6: The probability density function of $U \stackrel{d}{=} \mathrm{U}(0,1)$.

Exercise 2

Consider the function f_V shown in figure 7; assume that $f_V(v) = 0$ for $v < 0$ and $v > 1$.

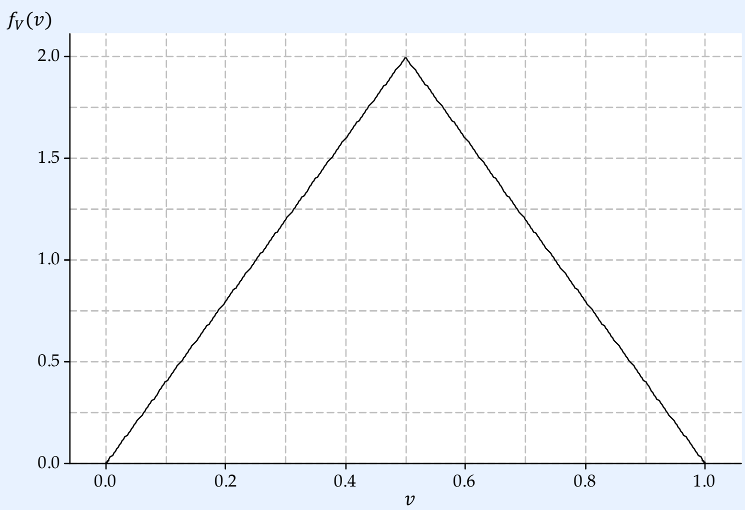

Figure 7: The probability density function of a random variable V with the triangular distribution.

- **a** Verify that f_V is a pdf.
- **b** Give a formula (involving cases) for the function $f_V(v)$.
- **c** Suppose a continuous random variable V has this pdf. Find the cdf $F_V(v)$ of V.
- **d** Find $\Pr(0.2 \le V \le 0.3)$.
- **e** Which is more likely: $V \approx 0.3$ or $V \approx 0.8$? Explain.

The triangular pdf shown in figure 7 is the pdf of the average of two $\mathrm{U}(0,1)$ random variables. That is, if $U_1 \stackrel{\mathrm{d}}{=} \mathrm{U}(0,1)$ and $U_2 \stackrel{\mathrm{d}}{=} \mathrm{U}(0,1)$ are independent, then $V = \frac{1}{2}(U_1 + U_2)$ has the pdf in figure 7.

This raises the question: What does the average of three independent $\mathrm{U}(0,1)$ random variables look like? The answer is shown in figure 8. If $U_i \stackrel{\mathrm{d}}{=} \mathrm{U}(0,1)$, for $i = 1,2,3$, and the three random variables are independent, then $W = \frac{1}{3}(U_1 + U_2 + U_3)$ has the following pdf.

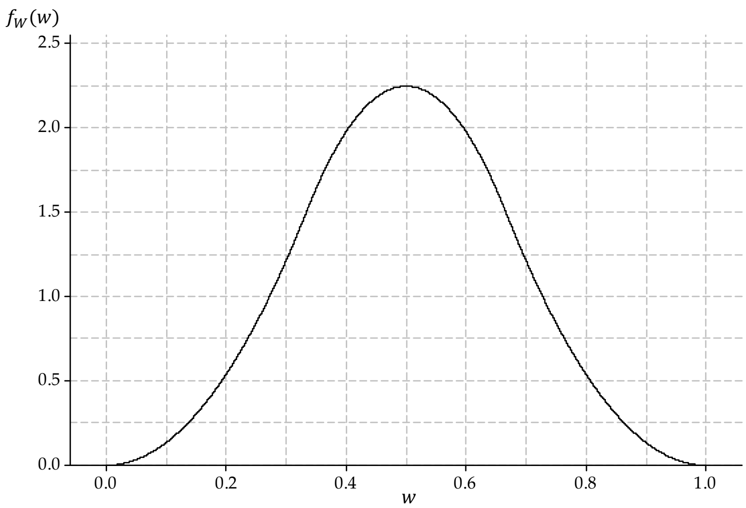

Figure 8: The probability density function of W, the average of three independent $\mathrm{U}(0,1)$ random variables.

Mean and variance of a continuous random variable

Mean of a continuous random variable

When introducing the topic of random variables, we noted that the two types — discrete and continuous — require different approaches.

In the module *Discrete probability distributions*, the definition of the mean for a discrete random variable is given as follows: The mean μ_X of a discrete random variable X with probability function $p_X(x)$ is

$$\mu_X = \mathrm{E}(X) = \sum x\, p_X(x),$$

where the sum is taken over all values x for which $p_X(x) > 0$.

The equivalent quantity for a continuous random variable, not surprisingly, involves an integral rather than a sum. The **mean** μ_X of a continuous random variable X with probability density function $f_X(x)$ is

$$\mu_X = \mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

By analogy with the discrete case, we may, and often do, restrict the integral to points where $f_X(x) > 0$.

Several of the points made when the mean was introduced for discrete random variables apply to the case of continuous random variables, with appropriate modification.

Recall that mean is a measure of 'central location' of a random variable. It is the weighted average of the values that X can take, with weights provided by the probability density function. The mean is also sometimes called the 'expected value' or 'expectation' of X and denoted by $\mathrm{E}(X)$. In visual terms, looking at a pdf, to locate the mean you need to work out where the pivot should be placed to make the pdf balance on the x-axis, imagining that the pdf is a thin plate of uniform material, with height $f_X(x)$ at x.

An important consequence of this is that the mean of any symmetric random variable (continuous or discrete) is always on the axis of symmetry of the distribution; for a continuous random variable, this means the axis of symmetry of the pdf.

Exercise 3

Two triangular pdfs are shown in figure 9.

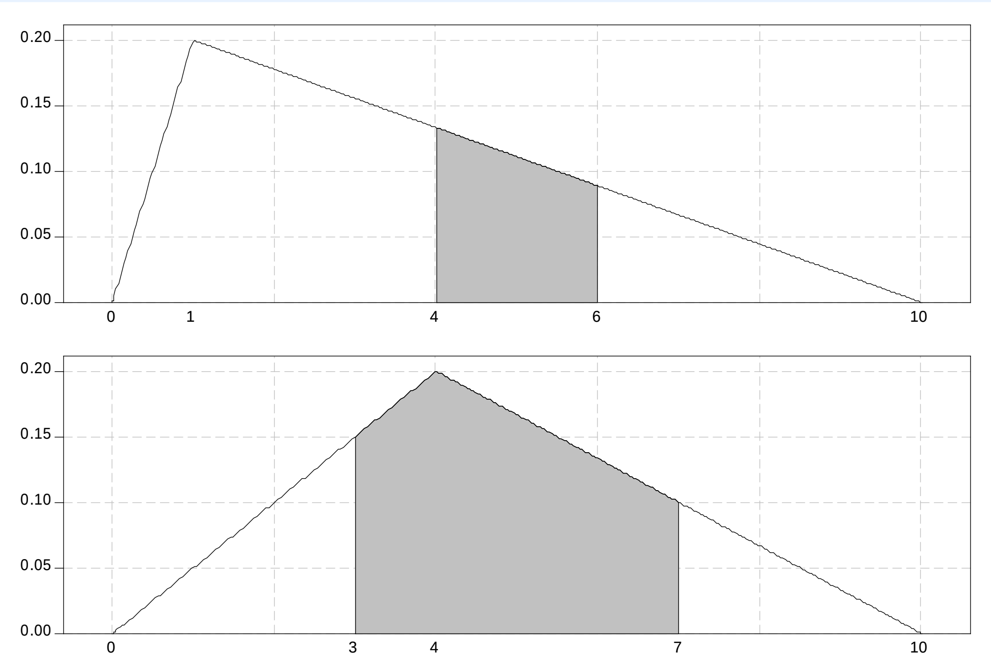

Figure 9: The probability density functions of two continuous random variables.

Each of the pdfs is equal to zero for $x < 0$ and $x > 10$, and the x-values of the apex and the boundaries of the shaded region are labelled on the x-axis in figure 9.

For each of these pdfs separately:

- **a** Write down a formula (involving cases) for the pdf.
- **b** Guess the value of the mean. Then calculate it to assess the accuracy of your guess.
- **c** Guess the probability that the corresponding random variable lies between the limits of the shaded region. Then calculate the probability to check your guess.

The module *Discrete probability distributions* gives formulas for the mean and variance of a linear transformation of a discrete random variable. In this module, we will prove that the same formulas apply for continuous random variables.

Theorem

Let X be a continuous random variable with mean μ_X. Then

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

for any real numbers a, b.

Proof

For a continuous random variable X, the mean of a function of X, say $g(X)$, is given by

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

So, for $g(X) = aX + b$, we find that

$$\begin{aligned}
\mathrm{E}(aX + b) &= \int_{-\infty}^{\infty} (ax + b)\, f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} ax\, f_X(x)\, dx + \int_{-\infty}^{\infty} b\, f_X(x)\, dx \\
&= a\int_{-\infty}^{\infty} x\, f_X(x)\, dx + b\int_{-\infty}^{\infty} f_X(x)\, dx \\
&= a\mu_X + b.
\end{aligned}$$

□

Variance of a continuous random variable

Recall that the variance of a random variable X is defined as follows:

$$\mathrm{var}(X) = \sigma_X^2 = \mathrm{E}[(X-\mu)^2], \qquad \text{where } \mu = \mathrm{E}(X).$$

The variance of a continuous random variable X is the weighted average of the squared deviations from the mean μ, where the weights are given by the probability density function $f_X(x)$ of X. Hence, for a continuous random variable X with mean μ_X, the **variance** of X is given by

$$\mathrm{var}(X) = \sigma_X^2 = \mathrm{E}[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\, dx.$$

Recall that the **standard deviation** σ_X is the square root of the variance; the standard deviation (and not the variance) is in the same units as the random variable.

Exercise 4

Find the standard deviation of

a the random variable $U \stackrel{\mathrm{d}}{=} \mathrm{U}(0,1)$; see figure 6

b the random variable V from exercise 2.

As observed in the module *Discrete probability distributions*, there is no simple, direct interpretation of the variance or the standard deviation. (The variance is equivalent to the 'moment of inertia' in physics.) However, there is a useful guide for the standard deviation that works most of the time in practice. This guide or 'rule of thumb' says that, for many distributions, the probability that an observation is within two standard deviations of the mean is approximately 0.95. That is,

$$\Pr(\mu_X - 2\sigma_X \le X \le \mu_X + 2\sigma_X) \approx 0.95.$$

This result is correct (to two decimal places) for an important distribution that we meet in another module, the Normal distribution, but it is found to be a useful indication for many other distributions too, including ones that are not symmetric.

Due to Chebyshev's theorem, not covered in detail here, we know that the probability $\Pr(\mu_X - 2\sigma_X \le X \le \mu_X + 2\sigma_X)$ can be as small as 0.75 (but no smaller) and it can be as large as 1. So clearly, the rule does not apply in some situations. But these extreme distributions arise rather infrequently across a broad range of practical applications.

Exercise 5

For the random variable V from exercises 2 and 4, find $\Pr(\mu_V - 2\sigma_V \le V \le \mu_V + 2\sigma_V)$.

We now consider the variance and the standard deviation of a linear transformation of a random variable.

Theorem

Let X be a random variable with variance σ_X^2. Then

$$\operatorname{var}(aX+b) = a^2 \operatorname{var}(X) = a^2 \sigma_X^2,$$

for any real numbers a, b.

Proof

Define $Y = aX + b$. Then $\operatorname{var}(Y) = \mathrm{E}[(Y-\mu_Y)^2]$. We know that $\mu_Y = a\mu_X + b$. Hence,

$$\begin{aligned}\operatorname{var}(aX+b) &= \mathrm{E}\big[\big(aX+b-(a\mu_X+b)\big)^2\big] \\ &= \mathrm{E}\big[a^2(X-\mu_X)^2\big] \\ &= a^2\,\mathrm{E}[(X-\mu_X)^2] \\ &= a^2 \operatorname{var}(X) \\ &= a^2\sigma_X^2. \qquad \square\end{aligned}$$

It follows from this result that

$$\operatorname{sd}(aX+b) = |a| \operatorname{sd}(X) = |a|\sigma_X.$$

Exercise 6

Suppose that X, the distance travelled by a taxi in a single trip in a major Australian city, has a mean of 15 kilometres and a standard deviation of 50 kilometres. Define Y to be the charge for a single trip (or, to be pedantic, the component of the charge that depends on flagfall and distance travelled). If the flagfall is \$3.20 and the rate per kilometre is \$2.20, what are the mean and the standard deviation of Y?

Relative frequencies and continuous distributions

Continuous random variables are technically an abstraction ($\Pr(X = x) = 0$), and all variables that we measure are *in practice, as measured,* discrete. Even quantities that seem to be intrinsically continuous, like time, are always measured to the nearest unit of time, depending on the context. The following table illustrates this.

Units used for time measurements

Context	Smallest unit of time usually used
Cosmology	10 million years (0.01 billion years)
Geological scale	MYA (million years ago)
Recorded history	year
Recent history	day
'What's the time?'	minute
Jogger	second
Many sports (athletics, swimming, ...)	hundredth of a second
Formula 1 racing	thousandth of a second

This discreteness follows from the rounding that occurs, either by necessity or by convention. There are many such examples in everyday life and also in research. We usually measure the height of individuals to the nearest centimetre, but it is easy to envisage a greater accuracy of measurement: to the nearest millimetre, for example.

Given all this intrinsic or necessary discreteness in measured variables, why not stick to the use of discrete distributions for modelling? Why use continuous distributions at all, especially if a continuous random variable is really an abstraction rather than a reality?

The answer is that it is often convenient and useful to use a continuous distribution, rather than attempting to use a discrete model.

Example: Train trips

Consider a study of a particular train trip in a timetable, from an outer suburban train station to a central city station. The purpose of the study is to check whether the actual times are adhering to the train timetable. The train's departure time and arrival time are recorded for 250 days. The times are measured to one-second accuracy, so that times such as 42 minutes and 7 seconds (42:07) are recorded.

The average time recorded, to the nearest second, was 2598 seconds (which corresponds to 43:18 minutes, or to 43.3 minutes as a decimal). The minimum time was 2466 seconds (41:06 minutes), and the maximum time was 3747 seconds (62:27 minutes).

If we regarded this random variable as being discrete, with possible values at every second $(\dots, 2800, 2801, 2802, \dots)$, and sought to model it using a discrete distribution, we would need to come up with probabilities for each second separately. In the absence of a theoretical basis for doing so, we might consider using the gathered data to estimate these probabilities.

Think of the consequences of doing this. There are 1282 discrete times (to the nearest second) in the range of the data, from 2466 to 3747 seconds. With $n = 250$ observations, most of these discrete values will not appear in the data, that is, they will have a frequency of zero. Of the rest, most will have a count of one, some will have two, and a handful of the discrete values might have occurred three or more times in the data set.

It would therefore be quite cumbersome to model the times as discrete. It is convenient to think of the times as coming from an underlying theoretical distribution which is continuous. We can then look at the continuous distribution and its properties to understand more about the pattern of the train-trip times.

A related point is that it is unlikely that we would ever need to consider the lengths of the train trips at the detailed level of the individual discrete times. It is much more likely that we would be interested in, say, the percentage of trips between 42 and 43 minutes, or the fraction of trips that are five minutes late or worse, and so on, rather than the chance that a train trip is 43 minutes 30 seconds.

Figure 10 shows a histogram of the 250 train trips. Alongside it is the pdf of a continuous random variable. This pdf is given by the following formula:

$$f(x) = \frac{\exp\left(-\frac{1}{2}[\ln(x-41)-0.4]^2\right)}{\sqrt{2\pi}(x-41)}, \qquad \text{for } x > 41.$$

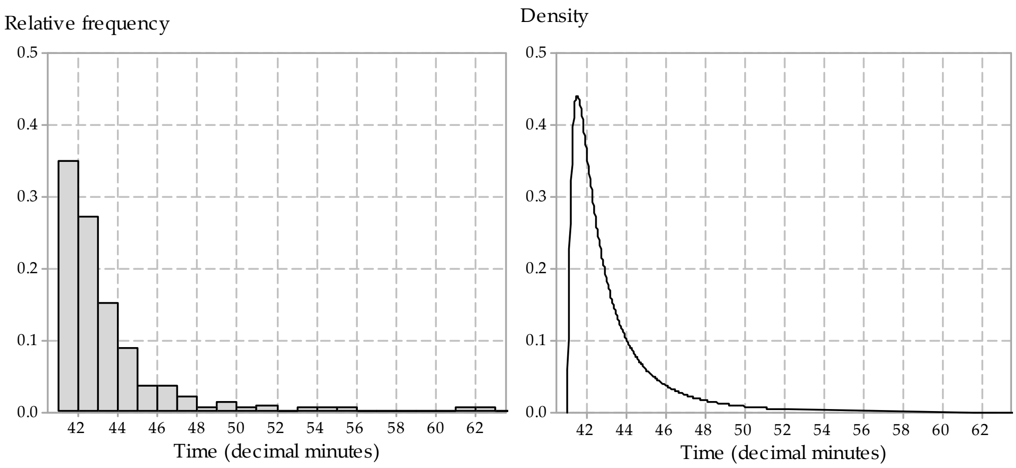

Figure 10: Comparison of relative-frequency data and a postulated model.

The relative frequencies in the histogram correspond to the areas under the graph of the probability density function. We can assess whether the model is a good fit to the data by looking at the probabilities from the pdf, and asking how close they are to the relative frequencies in the histogram.

For example, there were 87 trips between 41:00 minutes and 41:59 minutes; this is a relative frequency of $\frac{87}{250} = 0.348$. How close is this to the probability implied by the pdf? Calculating the area under the graph of this pdf is beyond the curriculum, but it is found to be 0.345, which is very close to the relative frequency. The following table shows the relative frequencies and the corresponding probabilities from the pdf, for the first five one-minute intervals.

Comparison of data and model for train-trip times

Time interval (minutes)	Frequency	Relative frequency	Probability from model
41:00 – 41:59	87	0.348	0.345
42:00 – 42:59	68	0.272	0.271
43:00 – 43:59	38	0.152	0.142
44:00 – 44:59	22	0.088	0.081
45:00 – 45:59	9	0.036	0.049

This example illustrates the way that continuous distributions are often used: as a useful approximation to a discrete random variable, when the discrete random variable is on a very fine scale. This occurs in a wide variety of contexts, such as measurements of human heights (centimetres), IQ (integers), exam marks and so on.

In circumstances where we do not have a clear basis for choosing a particular pdf to model data of this sort, the relative frequencies from the histogram serve as a guide: we obtain estimates of probabilities for given intervals directly. We then look for a continuous distribution that can closely reflect these estimates.

In the module *Exponential and normal distributions*, we will see this in practice; the data being considered are measured on a fine discrete scale, but are modelled using a continuous distribution.

Example: Taxi fares

Exercise 6 involves taxi fares. Any actual taxi fare is in dollars and cents, and hence the variable ‘taxi fare’ is discrete: it only takes values to the nearest five cents. To model the variation in taxi fares, however, we would typically use a continuous distribution.

Answers to exercises

Exercise 1

a On the interval $[0,1]$, the function $f(x)$ is a quadratic with a negative coefficient of x^2 and x-intercepts 0 and 1. Thus $f(x) \geq 0$, for x in $[0,1]$. By definition, we have $f(x)=0$, for x outside the interval $[0,1]$. Hence, the first property is satisfied: $f(x) \geq 0$, for all x.

To check the second property, we calculate:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_0^1 6x(1-x)\,dx \\ &= \int_0^1 6x - 6x^2\,dx \\ &= \left[3x^2 - 2x^3\right]_0^1 \\ &= 1.\end{aligned}$$

So the second property is also satisfied.

The graph of $f(x)$ is shown in the following figure.

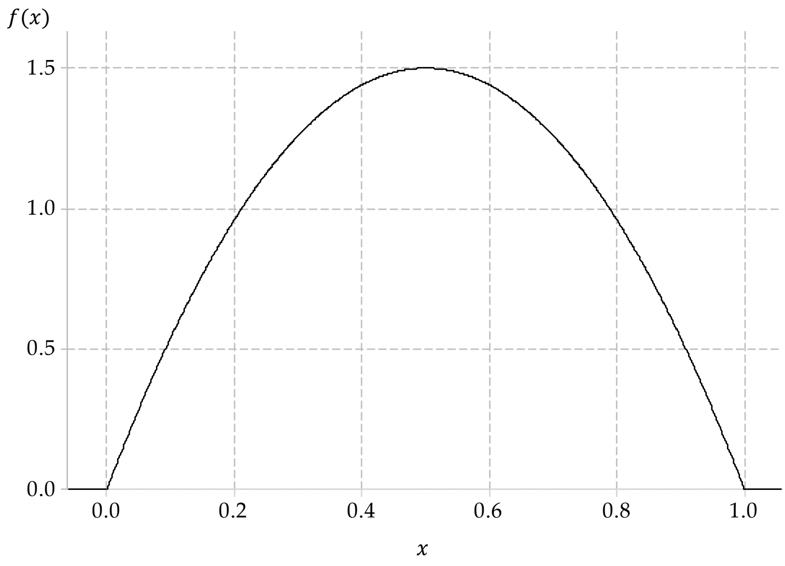

Figure 11: The probability density function $f(x)$.

b By considering the function or the graph, it is clear that:

i $\Pr(X \leq -3) = 0$

ii $\Pr(0 \leq X \leq 1) = 1$

iii $\Pr(0.5 \leq X \leq 1) = 0.5$, since the pdf is symmetric about $x = \frac{1}{2}$.

c To guess this probability, you need to estimate the area under the curve between 0.4 and 0.7. You can do this subjectively, keeping in mind that the total area under the curve is 1. Using the gridlines in figure 11 as a guide, we can make a slightly more informed guess:

- The area under the curve between 0.4 and 0.6 is nearly $0.2 \times 1.5 = 0.3$.
- The region under the curve between 0.6 and 0.7 is a rectangle of area $0.1 \times 1 = 0.1$ plus a small region whose area appears to be a bit less than $\frac{3}{4} \times 0.1 \times 0.5 = 0.0375$.

The 'guesstimate' for the area is therefore 'a bit less than $0.3 + 0.1 + 0.0375 = 0.4375$'. Calculating the probability gives

$$\begin{aligned}
\Pr(0.4 \le X \le 0.7) &= \int_{0.4}^{0.7} 6x(1-x)\,dx \\
&= \big[3x^2 - 2x^3\big]_{0.4}^{0.7} \\
&= 0.784 - 0.352 \\
&= 0.432.
\end{aligned}$$

d **i** $f(0.2) = 0.96$, $f(0.4) = 1.44$, so $\lambda = \dfrac{1.44}{0.96} = 1.5$.

ii $p_{0.2} = \Pr(0.15 \le X \le 0.25) = \displaystyle\int_{0.15}^{0.25} 6x(1-x)\,dx = 0.0955.$

iii $p_{0.4} = \Pr(0.35 \le X \le 0.45) = \displaystyle\int_{0.35}^{0.45} 6x(1-x)\,dx = 0.1435.$

iv $\dfrac{p_{0.4}}{p_{0.2}} = \dfrac{0.1435}{0.0955} = 1.503 \approx \lambda.$

Exercise 2

a By inspection of the graph in figure 7, we can see that $f_V(v) \ge 0$, for all v, and that the total area under the curve is $\frac{1}{2} \times 1 \times 2 = 1$. Hence, the function has the two properties of a pdf.

b The function $f_V(v)$ is made up of two lines. Formally:

$$f_V(v) = \begin{cases} 4v & \text{if } 0 \le v \le \frac{1}{2}, \\ 4 - 4v & \text{if } \frac{1}{2} < v \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

c For $v > 0$, we have $F_V(v) = \int_0^v f_V(t)\, dt$. We need to deal with the two parts of the pdf separately. For $0 < v \le \frac{1}{2}$, we have

$$F_V(v) = \int_0^v 4t\, dt = 2v^2.$$

For $\frac{1}{2} < v \le 1$, we have

$$F_V(v) = \Pr(V \le v) = \Pr(V \le \tfrac{1}{2}) + \Pr(\tfrac{1}{2} < V \le v),$$

and so

$$\begin{aligned} F_V(v) &= \tfrac{1}{2} + \int_{0.5}^v (4-4t)\, dt \\ &= \tfrac{1}{2} + \big[4t - 2t^2\big]_{0.5}^v \\ &= \tfrac{1}{2} + (4v - 2v^2) - \tfrac{3}{2} \\ &= 4v - 2v^2 - 1. \end{aligned}$$

Putting these two parts together, we can write, formally, that the cdf of V is given by

$$F_V(v) = \begin{cases} 0 & \text{if } v \le 0, \\ 2v^2 & \text{if } 0 < v \le \frac{1}{2}, \\ 4v - 2v^2 - 1 & \text{if } \frac{1}{2} < v \le 1, \\ 1 & \text{if } v > 1. \end{cases}$$

d $\Pr(0.2 \le V \le 0.3) = F_V(0.3) - F_V(0.2) = 0.1$. This can also be found from the pdf using the area of a trapezium.

e We have $f_V(0.3) = 1.2 > 0.8 = f_V(0.8)$, and so $V \approx 0.3$ is more likely than $V \approx 0.8$. (Here we write $V \approx v$ to mean that V lies in a small interval of given width around v.)

Exercise 3

a For the first pdf:

$$f(x) = \begin{cases} \frac{1}{5}x & \text{if } 0 \le x \le 1, \\ \frac{1}{45}(10-x) & \text{if } 1 < x \le 10, \\ 0 & \text{otherwise.} \end{cases}$$

For the second pdf:

$$f(y) = \begin{cases} \frac{1}{20}y & \text{if } 0 \le y \le 4, \\ \frac{1}{30}(10-y) & \text{if } 4 < y \le 10, \\ 0 & \text{otherwise.} \end{cases}$$

b You need to guess where the centre of gravity of the pdf is, that is, where you would need to put the pivot to make the pdf balance. Reasonable guesses would be 'between 3 and 4' for the first pdf, and 'between 4 and 5' for the second pdf.

If the continuous random variable X has the first pdf, then

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\
&= \int_0^1 \frac{1}{5}x^2\, dx + \int_1^{10} \frac{1}{45}\left(10x - x^2\right) dx \\
&= \left[\frac{1}{15}x^3\right]_0^1 + \frac{1}{45}\left[5x^2 - \frac{1}{3}x^3\right]_1^{10} \\
&= \left(\frac{1}{15} - 0\right) + \frac{1}{45}\left(\left(500 - \frac{1000}{3}\right) - \left(5 - \frac{1}{3}\right)\right) \\
&= \frac{1}{15} + \frac{162}{45} = \frac{11}{3}.
\end{aligned}$$

If the continuous random variable Y has the second pdf, then

$$\begin{aligned}
\mathrm{E}(Y) &= \int_{-\infty}^{\infty} y\, f(y)\, dy \\
&= \int_0^4 \frac{1}{20}y^2\, dy + \int_4^{10} \frac{1}{30}\left(10y - y^2\right) dy \\
&= \left[\frac{1}{60}y^3\right]_0^4 + \frac{1}{30}\left[5y^2 - \frac{1}{3}y^3\right]_4^{10} \\
&= \left(\frac{64}{60} - 0\right) + \frac{1}{30}\left(\left(500 - \frac{1000}{3}\right) - \left(80 - \frac{64}{3}\right)\right) \\
&= \frac{32}{30} + \frac{108}{30} = \frac{14}{3}.
\end{aligned}$$

c The total area under the curve is 1. Reasonable guesses might be 'about a quarter' and 'a bit more than one half'. The values are $\frac{2}{9} = 0.222$ and $\frac{5}{8} = 0.625$, obtained by either integration or simple calculations of areas.

Exercise 4

a The pdf of U is given by

$$f_U(u) = \begin{cases} 1 & \text{if } 0 \le u \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We know the mean of the distribution is $\mu_U = \frac{1}{2}$.

First we find the variance:

$$\begin{aligned}
\mathrm{var}(U) &= \int_{-\infty}^{\infty} (u-\mu_U)^2 f_U(u)\,du \\
&= \int_{-\infty}^{\infty} \Big(u-\frac{1}{2}\Big)^2 f_U(u)\,du \\
&= \int_0^1 \Big(u-\frac{1}{2}\Big)^2\,du \\
&= \int_0^1 \Big(u^2-u+\frac{1}{4}\Big)\,du \\
&= \Big[\frac{1}{3}u^3-\frac{1}{2}u^2+\frac{1}{4}u\Big]_0^1 \\
&= \frac{1}{3}-\frac{1}{2}+\frac{1}{4}=\frac{1}{12}.
\end{aligned}$$

Hence, the standard deviation of U is $\mathrm{sd}(U)=\sqrt{\frac{1}{12}}=0.289$.

b The pdf of V is given by

$$f_V(v)=\begin{cases} 4v & \text{if } 0\le v\le \frac{1}{2}, \\ 4-4v & \text{if } \frac{1}{2}< v\le 1, \\ 0 & \text{otherwise.}\end{cases}$$

We know the mean of the distribution is $\mu_V=\frac{1}{2}$.

First we find the variance:

$$\begin{aligned}
\mathrm{var}(V) &= \int_{-\infty}^{\infty} (v-\mu_V)^2 f_V(v)\,dv \\
&= \int_{-\infty}^{\infty} \Big(v-\frac{1}{2}\Big)^2 f_V(v)\,dv \\
&= \int_0^{\frac{1}{2}} \Big(v-\frac{1}{2}\Big)^2 4v\,dv + \int_{\frac{1}{2}}^1 \Big(v-\frac{1}{2}\Big)^2 (4-4v)\,dv \\
&= \int_0^{\frac{1}{2}} \Big(4v^3-4v^2+v\Big)\,dv + \int_{\frac{1}{2}}^1 \Big(-4v^3+8v^2-5v+1\Big)\,dv \\
&= \Big[v^4-\frac{4}{3}v^3+\frac{1}{2}v^2\Big]_0^{\frac{1}{2}} + \Big[-v^4+\frac{8}{3}v^3-\frac{5}{2}v^2+v\Big]_{\frac{1}{2}}^1 \\
&= \Big(\frac{1}{16}-\frac{1}{6}+\frac{1}{8}\Big)-0+\Big(-1+\frac{8}{3}-\frac{5}{2}+1\Big)-\Big(-\frac{1}{16}+\frac{1}{3}-\frac{5}{8}+\frac{1}{2}\Big)=\frac{1}{24}.
\end{aligned}$$

Hence, the standard deviation of V is $\mathrm{sd}(V)=\sqrt{\frac{1}{24}}=0.204$.

Exercise 5

From the previous exercise, we have $\mu_V = 0.5$ and $\sigma_V = 0.204$. Thus

$$\begin{aligned}\Pr\big(\mu_V - 2\sigma_V \le V \le \mu_V + 2\sigma_V\big) &= \Pr\big(0.5 - (2 \times 0.204) \le V \le 0.5 + (2 \times 0.204)\big)\\ &= 2 \times \Pr(0.5 - 0.408 \le V \le 0.5),\end{aligned}$$

because the pdf is symmetric about 0.5. Now

$$\begin{aligned}\Pr(0.5 - 0.408 \le V \le 0.5) &= \Pr(0.092 \le V \le 0.5)\\ &= \int_{0.092}^{0.5} 4v\,dv\\ &= \big[2v^2\big]_{0.092}^{0.5}\\ &= 0.5 - 0.0169\\ &= 0.483.\end{aligned}$$

Hence, $\Pr\big(\mu_V - 2\sigma_V \le V \le \mu_V + 2\sigma_V\big) = 2 \times 0.483 = 0.966$. This is reasonably close to the probability 0.95 indicated by the rule of thumb.

Exercise 6

We have $Y = 2.20X + 3.20$.

Recall that $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$. As $\mathrm{E}(X) = 15$, this gives $\mathrm{E}(Y) = 2.20 \times 15 + 3.20 = 36.20$. The average cost is \$36.20.

Recall that $\mathrm{sd}(aX + b) = |a|\,\mathrm{sd}(X)$. Since $\mathrm{sd}(X) = 50$, this gives $\mathrm{sd}(Y) = 2.20 \times 50 = 110$. The standard deviation of the cost is \$110.

0 1 2 3 4 5 6 7 8 9 10 11 12