M347

Mathematical statistics

Univariate continuous distribution theory

Contents

This publication forms part of an Open University module. Details of this and other Open University modules can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)845 300 6090; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

To purchase a selection of Open University materials visit www.ouw.co.uk, or contact Open University Worldwide, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure (tel. +44 (0)1908 858793; fax +44 (0)1908 858787; email ouw-customer-services@open.ac.uk).

Note to reader
Mathematical/statistical content at the Open University is usually provided to students in printed books, with PDFs of the same online. This format ensures that mathematical notation is presented accurately and clearly. The PDF of this extract thus shows the content exactly as it would be seen by an Open University student. Please note that the PDF may contain references to other parts of the module and/or to software or audio-visual components of the module. Regrettably mathematical and statistical content in PDF files is unlikely to be accessible using a screenreader, and some OpenLearn units may have PDF files that are not searchable. You may need additional help to read these documents.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2015.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Martins the Printers, Berwick upon Tweed.

ISBN 978 1 4730 0341 5

1.1

1 Pdfs and cdfs

The distributions of continuous random variables are described by their 'probability density functions (pdfs)' and 'cumulative distribution functions (cdfs)'. Pdfs are the topic of Subsection 1.1, where their basic properties are described and examples are developed. Cdfs, including their properties, some examples and their relationship to pdfs, are considered in Subsection 1.2. Finally, you will be reminded how differences between cdf values give probabilities of a random variable lying within an interval in Subsection 1.3.

1.1 Densities and normalising constants

A continuous random variable X follows a distribution with **probability density function**, or **pdf** (or sometimes just **density function** or **density**), f say.

This pdf has nothing to do with the 'portable document format' PDF files on your computer.

The key properties of a pdf f are that it is non-negative and integrates to 1, that is,

$$f(x) \geq 0 \text{ for all } x \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

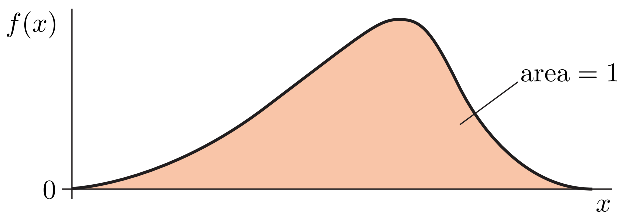


Figure 2.1 A pdf $f(x)$ plotted as a function of x

Note that in Figure 2.1, $f(x) \geq 0$ for all x. Also, its integral, which is the area of the shaded region under the pdf, is 1.

In fact, any mathematical function which is non-negative, positive on at least one interval of values of x, and has a finite integral can be made into a pdf. (A function whose integral does not exist cannot be made into a pdf.) Suppose that g is such a function with

$$\int_{-\infty}^{\infty} g(x)\,dx = C,$$

where $0 < C < \infty$. Then

$$f(x) = \frac{g(x)}{C}$$

is a pdf. To see this, note that

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{g(x)}{C}\,dx = \frac{1}{C}\int_{-\infty}^{\infty} g(x)\,dx = \frac{1}{C} \times C = 1.$$

C is called the **normalising** (or **normalisation**) **constant**.

Some statisticians use the term 'normalising constant' for the quantity $1/C$ instead of C.

The following nomenclature is not entirely standard in statistics, but can be useful. The part g of the pdf f that contains all the dependence of f on x will be referred to as the **density core**, or just **core** for short, in M347.

One thing to bear in mind is that a pdf is not a probability itself. In particular, $f(x) \neq P(X = x)$. Indeed, for a continuous distribution, $P(X = x)$ equals zero! However, there is a closely related probability. The probability that X lies within a specific interval $[x_0, x_0 + \varepsilon)$, where $\varepsilon > 0$ is small, is approximately equal to ε times the density at x_0 (see Figure 2.2). Formally,

$$\frac{P(x_0 \leq X < x_0 + \varepsilon)}{\varepsilon} \to f(x_0) \quad \text{as } \varepsilon \to 0.$$

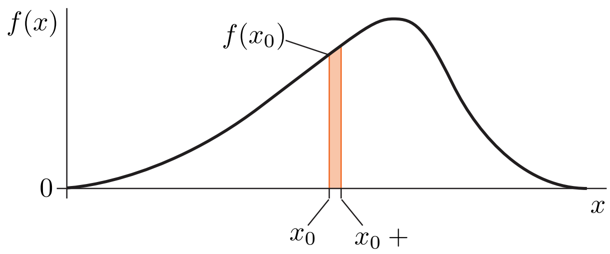


Figure 2.2 The pdf $f(x)$ of Figure 2.1 with the area under the pdf and between x_0 and $x_0 + \varepsilon$ shaded

1.1.1 Examples

Example 2.1 *The pdf of the uniform distribution on* $(0, 1)$

The distribution with density f which is constant on $0 < x < 1$ and zero otherwise is called the **uniform distribution on** $(\mathbf{0}, \mathbf{1})$, denoted $U(0, 1)$. The constant value, k say, must be positive for f to be a density. What is the value of k? Well, the density is 0 for $x \leq 0$, k for $0 < x < 1$ and 0 again for $x \geq 1$. So, splitting the range of integration into three parts gives

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx \\ &= \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} k\,dx + \int_{1}^{\infty} 0\,dx \\ &= 0 + \int_{0}^{1} k\,dx + 0 \\ &= \int_{0}^{1} k\,dx = [kx]_0^1 = k(1-0) = k.\end{aligned}$$

Integration of a constant

However, this integral should be 1, so it must be that $k = 1$. That is,

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

For short, it can be said that

$$f(x) = 1 \quad \text{on } 0 < x < 1.$$

The uniform pdf is shown in Figure 2.3.

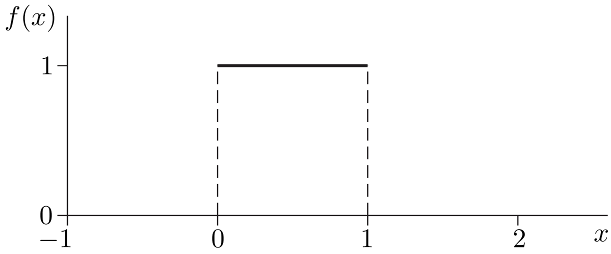


Figure 2.3 Graph of the pdf of the uniform distribution on $(0, 1)$

Notice that you could have replaced the integration step in calculating k by using the area of the square in Figure 2.3.

When a pdf only takes positive values on a subset of $\mathbb{R}$, then the limits of integration reduce to the limits of that subset. In Example 2.1, these limits were 0 and 1. In general, the domain on which f takes positive values is called the **support** of the distribution. Thus, in Example 2.1, the support of f is $(0, 1)$ and the density is written in the form

$$f(x) = \{\text{function of } x\} \quad \text{on } x \text{ in its support.}$$

And, in general, the integral of f is the integral of f over its support. In M347, the support of f will always be a single interval, although one or both endpoints of that interval may be $\pm\infty$.

Example 2.2 *The pdf of the exponential distribution*

The **exponential distribution**, $M(\lambda)$, with parameter $\lambda > 0$, is claimed to have density

$$f(x) = \lambda e^{-\lambda x} \quad \text{on } x > 0.$$

(Thus, implicitly, $f(x) = 0$ for $x \leq 0$.) Is this truly a density? Well the exponential function is non-negative everywhere, so f certainly is. Does it integrate to 1? Yes, it does: noting that the support of f is $(0, \infty)$,

Integration of an exponential

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_0^{\infty} \lambda e^{-\lambda x}\,dx \\ &= \lambda \int_0^{\infty} e^{-\lambda x}\,dx \\ &= \lambda \left[\left(-\frac{1}{\lambda}\right) e^{-\lambda x}\right]_0^{\infty} = -\left[e^{-\lambda x}\right]_0^{\infty} = 0 - (-1) = 1.\end{aligned}$$

The exponential pdf is shown in Figure 2.4 for $\lambda = 1$.

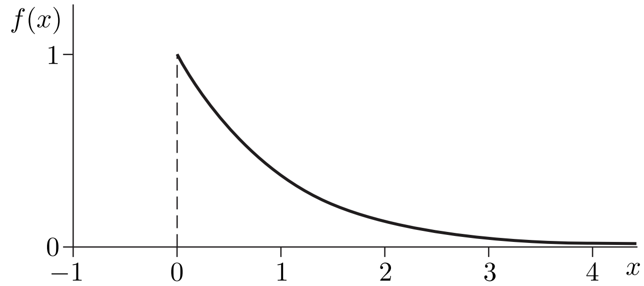


Figure 2.4 Graph of the pdf of the exponential distribution with $\lambda = 1$

Exercise 2.1

For each of the following four functions, decide whether or not they are, or can be made into, densities. If they are non-negative and integrable, calculate what their correct normalising constants should be.

(a) $f_{\mathrm{a}}(x) = x(1 - x) \quad \text{on } 0 < x < 1.$

(b) $f_{\mathrm{b}}(x) = e^{x}(1 - x) \quad \text{on } x > 0.$

(c) $f_{\mathrm{c}}(x) = \begin{cases} 1 + x & \text{on } -1 < x < 0, \\ 1 - x & \text{on } 0 \leq x < 1. \end{cases}$

(d) $f_{\mathrm{d}}(x) = 1/x \quad \text{on } x > 1.$

1.2 Distribution functions and densities

The **cumulative distribution function**, **distribution function** or **cdf**, F say, of the random variable X is defined as

$$F(x) = P(X \leq x).$$

That is, $F(x)$ is the probability that the random variable X takes any value less than or equal to the fixed value x.

Exercise 2.2

Explain why, in the continuous case, $F(x)$ is also equal to $P(X < x)$.

In this unit, the continuous case only is studied.

The key properties of F are:

- $0 \leq F(x) \leq 1$ (because $F(x)$ is a probability);
- for any $c < d$ in the support, $F(c) < F(d)$ (that is, the probability of being less than or equal to a particular value is less than the probability of being less than or equal to a larger value);
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$ (that is, X is certain to lie somewhere between $-\infty$ and ∞).

The cdf and pdf are intimately related as shown below.

$$F(x) = \int_{-\infty}^{x} f(y)\,dy; \qquad f(x) = \frac{d}{dx}F(x) = F'(x).$$

It is valid to think of this relationship as *defining* the pdf f.

The audio below accompanies Animation 2.1, and you should listen to this audio while using the animation.

Audio 2.1

Interactive content appears here. Please visit the website to use it.

Animation 2.1 Linked pair of plots of f and F

Interactive content appears here. Please visit the website to use it.

It is important to notice that in the relationship $F(x) = \int_{-\infty}^{x} f(y)\,dy$, x is the upper limit of integration, and something else – here, y – is the variable with respect to which you integrate. Any symbol will do for the variable of integration in place of y, except for x as this has to be the upper limit of integration. It is both wrong and confusing to write $F(x) = \int_{-\infty}^{x} f(x)\,dx$.

Exercise 2.3

Use the relationships between a pdf and cdf to check mathematically that:

(a) $\lim_{x\to\infty} F(x) = 1$;

(b) F is an increasing function of x on its support.

1.2.1 Examples

Example 2.3 *The cdf of the uniform distribution on* $(0,1)$

From Example 2.1,

$$f(x) = 1 \quad \text{on } 0 < x < 1.$$

For any value of $x \le 0$,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} 0\,dy = 0.$$

For any value of $0 < x < 1$,

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(y)\,dy &= \int_{-\infty}^{0} 0\,dy + \int_{0}^{x} 1\,dy \\ &= 0 + \int_{0}^{x} 1\,dy = [y]_0^x = x - 0 = x. \end{aligned}$$

For any value of $x \ge 1$,

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(y)\,dy &= \int_{-\infty}^{0} 0\,dy + \int_{0}^{1} 1\,dy + \int_{1}^{x} 0\,dy \\ &= 0 + \int_{0}^{1} 1\,dy + 0 = [y]_0^1 = 1 - 0 = 1. \end{aligned}$$

That is,

$$F(x) = \begin{cases} 0 & \text{if } x \le 0, \\ x & \text{if } 0 < x < 1, \\ 1 & \text{if } x \ge 1. \end{cases}$$

You can check this calculation by differentiating F to get $f(x) = F'(x) = 1$ if $0 < x < 1$, $f(x) = 0$ otherwise.

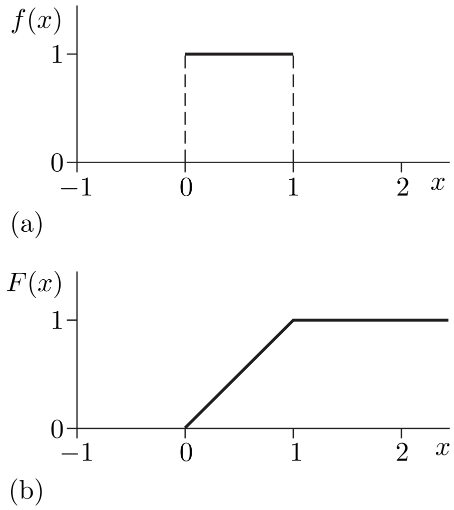


Figure 2.5 (a) Graph of the pdf of the uniform distribution on $(0,1)$; (b) graph of the cdf of the uniform distribution on $(0,1)$

It is always the case that if the support of f, (a,b) say, is not the whole of $\mathbb{R}$ but an interval subset thereof, then $F(x)=0$ for $x \leq a$ and $F(x)=1$ for $x \geq b$. This was illustrated for the uniform distribution with $a=0$ and $b=1$ in Figure 2.5. In such cases, for short, just write

$$F(x) = \{\text{function of } x\} \quad \text{on } a < x < b.$$

For example, for the uniform distribution of Example 2.3, write

$$F(x) = x \quad \text{on } 0 < x < 1.$$

When doing so, do not lose sight of the fact that F is zero 'below' its support and F is one 'above' its support. If $a=-\infty$, then there is no interval of values for which $F(x)=0$, and if $b=\infty$, then there is no interval of values with $F(x)=1$.

Exercise 2.4

From Example 2.2, the pdf of the exponential distribution with parameter $\lambda > 0$ is

$$f(x) = \lambda e^{-\lambda x} \quad \text{on } x > 0.$$

This exercise concerns the cdf of the exponential distribution.

(a) What is the value of $F(x)$ for $x \leq 0$?

(b) Show that the cdf of the exponential distribution is

$$F(x) = 1 - e^{-\lambda x} \quad \text{on } x > 0.$$

(c) Check this result by differentiation to obtain the corresponding pdf.

A distribution is introduced now that may be new to you but will be used as an example sufficiently often in M347 to warrant a name: the **power distribution**. It has density function

$$f(x) = \beta x^{\beta - 1} \quad \text{on } 0 < x < 1.$$

The parameter β can take any positive value.

Animation 2.2 Graph of the pdf of a power distribution, with a slider to change the value of $\beta > 0$

Interactive content appears here. Please visit the website to use it.

Exercise 2.5

Using Animation 2.2, describe the main qualitative behaviour of the power density for values of $\beta < 1$, $\beta = 1$ and $\beta > 1$, respectively.

Exercise 2.6

For each of the following two distributions, obtain their cdfs.

(a) The power distribution.

(b) The continuous **uniform distribution on** $(\boldsymbol{a}, \boldsymbol{b})$, which has density

$$f(x) = \frac{1}{b-a} \quad \text{on } a < x < b.$$

Notice that the uniform distribution on $(0, 1)$ is a special case of both of the distributions in Exercise 2.6. It is the special case of the uniform distribution on (a, b) obtained by setting $a = 0$, $b = 1$; and it is also the special case of the power distribution corresponding to setting $\beta = 1$ (as $x^0 = 1$ for any x).

1.3 Probabilities of lying within intervals

Let $c < d$. It is the case that

$$P(c \le X \le d) = F(d) - F(c) = \int_c^d f(x)\,dx.$$

Also,

$$P(c \le X \le d) = P(c \le X < d) = P(c < X \le d) = P(c < X < d)$$

since X is continuous so that $P(X = c) = P(X = d) = 0$. The formula in the box above is most easily seen with the aid of an animated figure:

Animation 2.3 Consecutive showing of images corresponding to $F(d)$, then $F(c)$, then $F(d) - F(c)$

Interactive content appears here. Please visit the website to use it.

Exercise 2.7

(a) For the distribution with cdf

$$F(x) = x^2(3 - 2x) \quad \text{on } 0 < x < 1,$$

calculate $P(\frac{1}{4} \le X \le \frac{5}{8})$.

(b) For the exponential distribution with $\lambda = 1$ which has cdf

$$F(x) = 1 - e^{-x} \quad \text{on } x > 0,$$

show that

$$P(\log 2 \le X \le \log 4) = \tfrac{1}{4}.$$

Exercise 2.8

(a) Write $P(x_0 \leq X < x_0 + \varepsilon)$, $\varepsilon > 0$, as a function of F.

(b) Hence explain why

$$\lim_{\varepsilon \to 0} \frac{P(x_0 \leq X < x_0 + \varepsilon)}{\varepsilon} = f(x_0).$$

2 Moments

The mean and variance are special cases of a concept known as the 'moments' of a distribution. These in turn are special cases of the more general concept of 'expectation', or 'expected value', which is described in Subsection 2.1. The familiar notion of the mean is then considered in Subsection 2.2. The mean is also known as the first moment of a distribution, which leads, in Subsections 2.3 and 2.4, to two general definitions of moments. Following these, another familiar notion, the variance, is studied in Subsection 2.5. Subsection 2.6 concerns random variables linked by 'linear transformation' and, in particular, the means, variances and densities thereof. Finally, in Subsection 2.7, you will see how to deal with many moments all in one go through the medium of something called the 'moment generating function'.

> 'Doing the best at this moment puts you in the best place for the next moment.'
>
> (Oprah Winfrey, American television personality, actress and producer)

2.1 Expectation

Informally, as you might guess, the **expected value**, or **expectation**, of a random variable X is some kind of average or mean or typical value of X. The formal definition of expected value is given here in general terms for the expected value of a general function h of X. This expected value is denoted by $E\{h(X)\}$ and is defined by another integral.

Let (a, b) denote the support of pdf f, remembering that a can be $-\infty$ and b can be ∞. Then

$$E\{h(X)\} = \int_a^b h(x)\, f(x)\, dx. \tag{2.1}$$

The specific type of brackets used is *not* part of the notation for expectation, which is just the E.

There is an implicit assumption here that the integral is finite, but this need not necessarily be so; if the integral is not finite, then the expected value of $h(X)$ is said not to exist.

Here is a first, particularly simple but useful, special case of the above. Suppose that h does not actually depend on X at all, that is, $h(X) = k$ where k is a constant. Then

$$E(k) = \int_a^b k\, f(x)\, dx = k \int_a^b f(x)\, dx = k \times 1 = k.$$

(This particular expectation always exists because f is a pdf.) This is very reasonable: the expected (or 'average') value of a constant is none other than the constant value itself (what else could it be?).

Some other simple choices of h are especially important in statistics, and these will be the principal focus in the following subsections. Some of them will be familiar to you already, for example, the mean and the variance. However, before looking at them, it will be useful to obtain a general result and some important consequences thereof.

Let h_1 and h_2 be two functions of X, and let k_1 and k_2 be two constants. Then the **linearity of expectation** property is that

$$E\{k_1\, h_1(X) + k_2\, h_2(X)\} = k_1\, E\{h_1(X)\} + k_2\, E\{h_2(X)\}. \qquad (2.2)$$

Exercise 2.9

Since $k_1\, h_1(X) + k_2\, h_2(X)$ is itself just a function of X, its expected value is

$$E\{k_1\, h_1(X) + k_2\, h_2(X)\} = \int_a^b \{k_1\, h_1(x) + k_2\, h_2(x)\}\, f(x)\, dx,$$

where (a, b) is the support of f. Show that Equation (2.2) holds.

Exercise 2.10

Write down a formula for $E\{k_1\, h(X) + k_2\}$ in terms of k_1, k_2 and $E\{h(X)\}$.

The result of Exercise 2.10 is a particularly important consequence of the linearity of expectation:

$$E\{k_1\, h(X) + k_2\} = k_1\, E\{h(X)\} + k_2. \qquad (2.3)$$

Exercise 2.11

Write down a formula for $E\{k\, h(X)\}$ in terms of k and $E\{h(X)\}$.

In fact, linearity of expectation holds for a linear combination of functions of X consisting of any *finite* number of terms (and not just two terms as above). The result is as follows.

$$E\left\{\sum_{i=1}^{n} k_i\, h_i(X)\right\} = \sum_{i=1}^{n} k_i\, E\{h_i(X)\}. \qquad (2.4)$$

This result follows from the linearity of integration in the same way that Equation (2.2) does, but no further details of the proof are given.

2.2 The mean

Perhaps the most important measure of the location of a distribution is the **mean**, or **arithmetic mean**, of X. This is nothing other than the expected value, or expectation, of the random variable X itself, i.e. $E\{h(X)\}$ where $h(X) = X$. The usual notation for the mean is μ.

If X is defined on (a, b), then

$$\mu = E(X) = \int_a^b x\, f(x)\, dx.$$

The mean does not exist if its defining integral does not.

Notice that this subsection concerns the **population mean** of a continuous distribution rather than the sample mean briefly reviewed in Subsection 1.2 of Unit 1 or the population mean of a discrete distribution briefly reviewed in Subsection 1.4 of Unit 1.

Example 2.4 *The mean of the uniform distribution*

The uniform distribution on (a, b) was introduced in Exercise 2.6(b). The mean of the uniform distribution on (a, b) is

$$\begin{aligned} E(X) &= \int_a^b x\left(\frac{1}{b-a}\right) dx \\ &= \frac{1}{b-a}\int_a^b x\, dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{1}{2(b-a)}\left(b^2 - a^2\right) = \frac{(b+a)(b-a)}{2(b-a)} = \tfrac{1}{2}(a+b). \end{aligned}$$

The mean of the uniform distribution on $(0, 1)$ is therefore $\frac{1}{2}(0+1) = \frac{1}{2}$.

The next example requires integration by parts as well as by substitution.

Example 2.5 *The mean of the exponential distribution*

Using the pdf of the exponential distribution given in Example 2.2,

$$\begin{aligned}E(X) &= \int_0^\infty x\lambda e^{-\lambda x}\,dx\\ &= \frac{1}{\lambda}\int_0^\infty ue^{-u}\,du\\ &\quad\text{(using the substitution } u=\lambda x,\ du=\lambda\,dx)\\ &= \frac{1}{\lambda}\left\{\left[-ue^{-u}\right]_0^\infty - \left(-\int_0^\infty e^{-u}\,du\right)\right\}\\ &\quad\text{(integrating by parts using } f(u)=u,\ g'(u)=e^{-u}\\ &\quad\text{so that } f'(u)=1,\ g(u)=-e^{-u})\\ &= \frac{1}{\lambda}\left\{0-0-\left[e^{-u}\right]_0^\infty\right\} = \frac{1}{\lambda}\{-(0-1)\} = \frac{1}{\lambda}\times 1 = \frac{1}{\lambda}.\end{aligned}$$

Integration by substitution
Integration by parts

Exercise 2.12

Calculate the mean of the distribution with density

$$f(x) = 6x(1-x) \quad \text{on } 0<x<1.$$

2.3 Raw moments

The mean is also known as the first moment of a distribution. The second and third moments are defined to be $E(X^2)$ and $E(X^3)$, respectively, and so on.

Generally, the rth **moment** is defined to be $E(X^r)$, $r=1,2,\ldots,$ where

$$E(X^r) = \int_a^b x^r\, f(x)\,dx.$$

Since $X^0=1$, $E(X^0)=E(1)=1$, although one rarely refers to the zeroth moment.

As with the expected value, these moments are subject to the existence of their defining integrals.

To distinguish them from other versions of moments, one of which will be introduced in the next section, these moments are sometimes called the **raw moments**. Broadly speaking, the lower the value of r, the more important the corresponding moment is in statistics!

For some distributions, it is as easy to determine the general rth moment as it is the mean, so all the (raw) moments might as well be calculated at once.

Example 2.6 *Raw moments of the uniform distribution on* $(0, 1)$

The rth moment of the uniform distribution on $(0, 1)$ is

$$E(X^r) = \int_0^1 x^r \, dx = \left[\frac{x^{r+1}}{r+1}\right]_0^1 = \frac{1}{r+1}(1-0) = \frac{1}{r+1}.$$

(Remember, $f(x) = 1$ on $0 < x < 1$.) In particular, setting $r = 1$ gives $E(X) = \frac{1}{2}$, as previously shown in Example 2.4.

Uniform distribution on $(0, 1)$

Exercise 2.13

Calculate the rth moment of the power distribution, which has density $f(x) = \beta x^{\beta-1}$ on $(0, 1)$. Can you deduce the formula for the rth moment of the uniform distribution on $(0, 1)$ (Example 2.6) from your result?

Power distribution

2.4 Central moments

For $r = 2, 3, \ldots$, it is sometimes useful to consider the **central moments**, or **moments about the mean**, defined by $\mu_r = E\{(X-\mu)^r\}$.

It will make life easier below to also define $\mu_0 = E\{(X-\mu)^0\}$ and $\mu_1 = E\{(X-\mu)^1\}$, but these both take particular constant values.

Exercise 2.14

What are the values of μ_0 and μ_1?

The rth central moment can be written in terms of the mean μ and the raw moments up to and including the rth, and the rth raw moment can be written in terms of the mean and the central moments up to the rth. The latter formula will be derived next. The derivation starts with a handy little trick: write $X = X - \mu + \mu$ so that

$$X^r = (X - \mu + \mu)^r = \{(X - \mu) + \mu\}^r.$$

Now the binomial expansion

$$(a + b)^r = \sum_{i=0}^{r} \binom{r}{i} a^i b^{r-i}$$

can be used. Set $a = X - \mu$, $b = \mu$ to get

$$X^r = \sum_{i=0}^{r} \binom{r}{i} (X - \mu)^i \mu^{r-i}$$

so that, by the linearity of expectation and the fact that μ is a constant,

$$E(X^r) = E\left\{\sum_{i=0}^{r}\binom{r}{i}(X-\mu)^i\mu^{r-i}\right\} = \sum_{i=0}^{r}\binom{r}{i}E\{(X-\mu)^i\}\,\mu^{r-i}$$

Linearity of expectation

or

$$E(X^r) = \sum_{i=0}^{r}\binom{r}{i}\mu_i\mu^{r-i}.$$

Thus the rth raw moment $E(X^r)$ has been written in terms of the mean μ, $\mu_0 = 1$, $\mu_1 = 0$ and the central moments $\mu_2, \mu_3, \ldots, \mu_r$.

Exercise 2.15

Show that the rth central moment μ_r can be written in terms of the mean μ and the raw moments up to the rth as follows:

$$\mu_r = \sum_{i=0}^{r}\binom{r}{i}(-1)^{r-i}\,E(X^i)\,\mu^{r-i}.$$

2.5 The variance

In the same way that the mean holds a special place in statistics as a leading measure of the location of a distribution, so the **variance** – and its very close relation the **standard deviation** – holds an equivalent place as the leading measure(s) of the **scale** or **spread** of a distribution. You may well have recognised the variance as being precisely the same as the second central moment, μ_2, in Subsection 2.4, although this notation is rarely used for the variance. Instead, the variance is denoted either by σ^2 or by $V(X)$.

The variance is given by

$$\sigma^2 = V(X) = \mu_2 = E\{(X-\mu)^2\}.$$

The standard deviation, σ, is the square root of the variance,

$$\sigma = \sqrt{V(X)}.$$

Again, this is the population, rather than the sample, variance.

Example 2.7 *Variance and standard deviation of the uniform distribution on* $(0, 1)$

In Example 2.4, it was shown that for the uniform distribution on $(0, 1)$, $\mu = \frac{1}{2}$. This can be used in calculating the variance of the uniform distribution:

$$\begin{aligned} V(X) = E\{(X-\mu)^2\} = E\{(X-\tfrac{1}{2})^2\} &= \int_0^1 (x-\tfrac{1}{2})^2\,dx \\ &= \int_0^1 (x^2 - x + \tfrac{1}{4})\,dx \\ &= \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 \\ &= (\tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{4}) - 0 \\ &= \tfrac{1}{12}. \end{aligned}$$

(Remember, $f(x) = 1$ on $0 < x < 1$.)

The standard deviation is therefore

$$\sigma = \sqrt{V(X)} = \sqrt{\tfrac{1}{12}} = 0.29 \text{ (correct to two decimal places).}$$

If, having worked out the variance directly in Example 2.7, you were concerned that such calculations could quickly start to become tricky, you will be pleased to find that there is an alternative, easier, route to the same answer. It depends on the following simple link between the second central moment, the variance, and the first and second raw moments, μ and $E(X^2)$.

$$V(X) = E(X^2) - \{E(X)\}^2 = E(X^2) - \mu^2. \tag{2.5}$$

This result is, of course, obtainable from the relationship derived in Exercise 2.15, although it is instructive to verify its truth from scratch:

$$\begin{aligned} V(X) &= E\{(X-\mu)^2\} \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \quad \text{(by linearity of expectation)} \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

Linearity of expectation

Example 2.8 *Variance of the uniform distribution revisited*

For the uniform distribution on $(0,1)$, $\mu = \frac{1}{2}$ and, from Example 2.6,

$$E(X^2) = \frac{1}{2+1} = \frac{1}{3}.$$

Using Equation (2.5), it is therefore the case that

$$V(X) = E(X^2) - \mu^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Note that this is the same result as that calculated earlier in Example 2.7 by directly evaluating the second central moment.

Exercise 2.16

In Exercise 2.13 you showed that for the power distribution,

$$E(X^r) = \frac{\beta}{r+\beta}.$$

Calculate the variance of the power distribution.

Power distribution

The variance and standard deviation are positive quantities. This is because the variance is an expectation, or average, of positive quantities, namely the squared values $h(X) = (X-\mu)^2$. Equivalently, the integral $\int h(x)\,f(x)\,dx$ involves positive quantities only and hence must be positive (the integral is also the area under a positive function). So if your calculations result in a negative variance you know you must have gone wrong!

Basic ideas of integration

2.6 Linear transformation

Let X be a random variable and let Y be another random variable defined through the **linear transformation** $Y = aX + b$. This subsection is split into two parts: the first concerns the mean and variance of Y, the second concerns the density of Y.

2.6.1 Expectation and variance

Let X be a random variable and let L be a new random variable defined by

$$L = X + b, \quad b \in \mathbb{R}.$$

Then Equation (2.3) tells us immediately that

$$E(L) = E(X+b) = E(X) + b = \mu + b.$$

Also,

$$\begin{aligned} V(L) &= E\{(L - E(L))^2\} \\ &= E\{(X + b - (\mu + b))^2\} \\ &= E\{(X-\mu)^2\} = V(X). \end{aligned}$$

On the other hand, let S be another random variable defined in terms of X by

$$S = aX, \quad a \in \mathbb{R}.$$

Then, again by Equation (2.3),

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

and

$$\begin{aligned} V(S) &= E\{(S - E(S))^2\} \\ &= E\{(aX - a\mu)^2\} \\ &= E\{a^2(X-\mu)^2\} = a^2\,E\{(X-\mu)^2\} = a^2\,V(X). \end{aligned}$$

If the standard deviation of S is denoted by σ_S and that of X by σ_X, this means that

$$\sigma_S = \sqrt{V(S)} = \sqrt{a^2\,V(X)} = |a|\,\sigma_X.$$

Note the absolute value signs on a: this is because σ_S must be positive, and σ_X is positive but there is no sign restriction on a.

Exercise 2.17

Write $Y = aX + b$.

(a) What is $E(Y)$ in terms of a, b, $E(X)$ and $V(X)$?

(b) What is $V(Y)$ in terms of a, b, $E(X)$ and $V(X)$?

The results of Exercise 2.17 are summarised in the following box.

> If $Y = aX + b$, then
>
> $$E(Y) = a\,E(X) + b \quad \text{and} \quad V(Y) = a^2\,V(X). \tag{2.6}$$

2.6.2 Effects on the pdf

Subsection 2.6.1 specifically concerned the effect of the linear transformation $Y = aX + b$ on the mean and the variance of Y as compared with those of X. More generally, it can be said that this linear transformation effects a **location and scale change**: the quantity b moves X along to a new location, to the right if $b > 0$, to the left if $b < 0$; the quantity $|a|$ changes the scale of X by expanding it if $|a| > 1$ or by contracting it if $|a| < 1$. This is illustrated in the following animation.

For the remainder of this subsection, restrict attention to cases where $a > 0$.

Animation 2.4 Exploring the effects of changing the location b and the scale a

Interactive content appears here. Please visit the website to use it.

Suppose now that X has a continuous distribution with pdf f_X and cdf F_X, and let $Y = aX + b$ once more, with $a > 0$.

The distribution of Y can be figured out in terms of the distribution of X. The cdf, F_Y, of Y is given in terms of F_X by

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(aX + b \le y) \\ &= P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Note that the manipulation of the inequality is correct because $a > 0$.

Differentiating the first and last terms of the above equation gives the density, f_Y, of Y in terms of f_X:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\ &= \frac{d}{dy}\left(\frac{y-b}{a}\right) \times \frac{d}{dx} F_X(x)\bigg|_{(y-b)/a} \\ &= \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Chain rule

(by the chain rule of differentiation). This relationship between the densities of linearly transformed (Y) and original (X) random variables is an important one.

If $Y = aX + b$, $a > 0$, then

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right). \qquad (2.7)$$

If f_X has support (c, d), then f_Y has support $(ac + b, ad + b)$.

Exercise 2.18

Suppose now that X follows the standard normal distribution and define $Y = \sigma X + \mu$.

(a) What are $E(Y)$ and $V(Y)$?

(b) Let the pdf of the standard normal distribution be designated $\phi(x)$. What, in terms of μ, σ and ϕ, is the pdf of Y?

(c) Use the result of part (b) and the formula for $\phi(x)$ to show that the density of Y is that of the general normal distribution with mean μ and variance σ^2 (a fact with which you should already be familiar).

Standard normal distribution

Normal distribution

The density of $Y = aX + b$ in Equation (2.7) is said to be written in **location–scale** form. The quantity b is then said to be a **location parameter** of the distribution of Y, and a is said to be its **scale parameter**. Often, the density of X will be in **standardised** form: this means that $E(X) = 0$ and $V(X) = 1$. In such cases $E(Y) = b$ and $\sigma_Y = a$, so the location and scale parameters coincide with the mean and standard deviation of the distribution.

These ideas were illustrated in Exercise 2.18 for the normal distribution where the density of X – the standard normal distribution – is indeed the standardised form of the normal distribution with mean 0 and standard deviation 1. As in Subsection 1.5.2 of Unit 1, $Y = \sigma X + \mu$ has the general normal distribution with mean μ and variance σ^2. But the point here is that the location–scale notion is quite general and by no means confined to the normal distribution.

2.7 The moment generating function

A single formula for the moments for all r is one way of providing a whole host of moment information at once. Another way, and one that proves to be especially useful in statistics (and in parts of the rest of M347), is to summarise all the moment information by calculating the **moment generating function**, usually abbreviated to **mgf**.

If $E(e^{tX})$ exists for all t in some interval $(-\delta, \delta)$, $\delta > 0$, then the function

$$M_X(t) = E(e^{tX}), \quad t \in (-\delta, \delta),$$

is called the moment generating function of X.

For fixed t, e^{tX} is just a particular choice of function $h(X)$ for which expectation can be considered as in Subsection 2.1.

Why the name? Well, here is a mathematically somewhat informal explanation. First, the Maclaurin expansion of e^{tX} – see Subsection 2.4 of Unit 1 – is employed:

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E\left\{1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots + \frac{(tX)^r}{r!} + \cdots\right\} \\ &= E\left\{1 + Xt + X^2\frac{t^2}{2!} + X^3\frac{t^3}{3!} + \cdots + X^r\frac{t^r}{r!} + \cdots\right\}. \end{aligned}$$

Now invoke the linearity of expectation result (2.4), noting that it was claimed to be true only for finite sums. Here there is an infinite series and so such a result cannot be applied trivially. However, it turns out that under fairly general conditions, which you need not worry about in M347, it is valid to write the moment generating function in the following form.

$$\begin{aligned} M_X(t) = 1 + E(X)\,t + E(X^2)\,\frac{t^2}{2!} + E(X^3)\,\frac{t^3}{3!} + \cdots \\ + E(X^r)\,\frac{t^r}{r!} + \cdots. \end{aligned} \qquad (2.8)$$

The reason for the name is now apparent: the function $M_X(t)$ *generates* all the (raw) moments of a distribution by attaching them as coefficients of $t^r/r!$ in a power series expansion of itself.

Moreover, there is a neat way to recover the individual raw moments from the mgf. First, differentiate (2.8) with respect to t to get

$$M_X'(t) = E(X) + E(X^2)\,t + E(X^3)\,\frac{t^2}{2!} + \cdots$$

Each prime, $'$, denotes differentiating once with respect to t.

and again to get

$$M_X''(t) = E(X^2) + E(X^3)\,t + \cdots.$$

Then set $t = 0$:

$$M_X'(0) = E(X), \quad M_X''(0) = E(X^2).$$

In general, the rth derivative of the mgf at $t=0$ is the rth raw moment, i.e.

$$E(X^r) = M_X^{(r)}(0) = M_X^{(r)}(t)\,|_{t=0}\,,$$

where $M_X^{(r)}(t)$ means differentiation r times with respect to t.

2.7.1 Examples

The mgf is particularly straightforward to obtain for the standard normal and exponential distributions.

Example 2.9 *The mgf of the standard normal distribution*

The standard normal distribution, $N(0,1)$, has density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right) \quad \text{on } \mathbb{R}.$$

Using Equation (2.1), the mgf is therefore

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} \exp(tx)\, f(x)\, dx \\ &= \int_{-\infty}^{\infty} \exp(tx)\, \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}x^2 + tx\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}(x^2 - 2tx)\right\} dx. \end{aligned}$$

Completing the square and then integrating by substitution gives

Completing the square

Integration by substitution

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}(x-t)^2 + \tfrac{1}{2}t^2\right\} dx \\ &= \exp(\tfrac{1}{2}t^2)\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}(x-t)^2\right\} dx \\ &= \exp(\tfrac{1}{2}t^2)\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}z^2\right) dz \\ &\qquad \text{(using the substitution } z = x - t,\ dz = dx) \\ &= \exp(\tfrac{1}{2}t^2) \int_{-\infty}^{\infty} f(x)\, dx \\ &= \exp(\tfrac{1}{2}t^2) \times 1 = \exp(\tfrac{1}{2}t^2). \end{aligned}$$

Now, the derivative of $M_X(t) = \exp(\frac{1}{2}t^2)$ with respect to t is

$$M'_X(t) = t\exp(\tfrac{1}{2}t^2),$$

so $\mu = E(X) = M'_X(0) = 0$, confirming the value of the mean of the standard normal distribution.

Also,

Product rule

$$M''_X(t) = \frac{d}{dt} M'_X(t) = \exp(\tfrac{1}{2}t^2) + t^2 \exp(\tfrac{1}{2}t^2) = (1+t^2)\exp(\tfrac{1}{2}t^2),$$

so $E(X^2) = M''_X(0) = 1$. In addition, it follows that the variance of the standard normal distribution is $E(X^2) - \mu^2 = 1 - 0^2 = 1$.

Exercise 2.19

If X follows the standard normal distribution, then $E(X^4) = 3$. Use the mgf of the standard normal distribution to prove this result.

Exercise 2.20

Exponential distribution

(a) Calculate the mgf of the exponential distribution. (You may assume that $t < \lambda$, which turns out to be the condition necessary for the exponential mgf to exist.)

(b) Hence verify that the mean and variance of the exponential distribution are $1/\lambda$ and $1/\lambda^2$, respectively.

(c) The formula for the $(r-1)$th derivative of $M_X(t)$, denoted $M_X^{(r-1)}(t)$, is

$$M_X^{(r-1)}(t) = \frac{(r-1)!\lambda}{(\lambda - t)^r}, \quad t < \lambda.$$

Hence obtain the formula for $M_X^{(r)}(t)$, $t < \lambda$.

(d) Using the result of part (c), verify that, for the exponential distribution,

$$E(X^r) = \frac{r!}{\lambda^r}.$$

Solutions

Solution 2.1

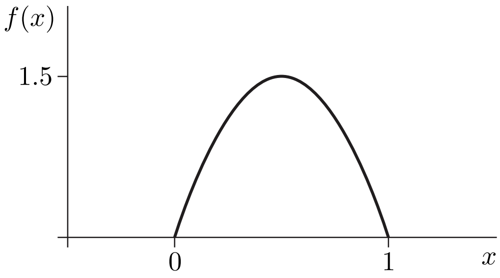
(a) $f_{\mathrm{a}}(x) \geq 0$ for all $x \in (0,1)$. Also,

$$\begin{aligned}\int_0^1 x(1-x)\,dx &= \int_0^1 (x-x^2)\,dx \\ &= \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 \\ &= \left(\tfrac{1}{2} - \tfrac{1}{3}\right) - (0-0) = \tfrac{1}{6}.\end{aligned}$$

Integration of a power

The correct density associated with this function is therefore

$$\frac{f_a(x)}{1/6} = \frac{x(1-x)}{1/6} = 6x(1-x) \quad \text{on } 0 < x < 1.$$



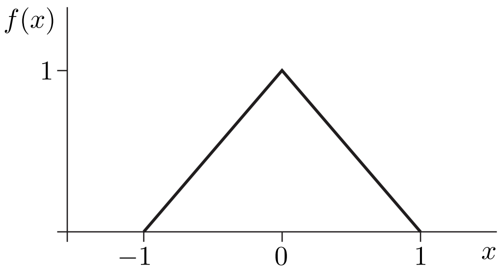
Graph of the pdf $6x(1-x)$ on $0 < x < 1$

(b) The exponential function is positive for all $x > 0$, but $1 - x$ is not: it is negative when $x > 1$. It follows that $f_{\mathrm{b}}(x)$ is negative when $x > 1$ and hence is not a probability density function on $x > 0$.

(c) This function is non-negative for all $x \in (-1,1)$. Its integral is

$$\begin{aligned}\int_{-1}^0 (1+x)\,dx + \int_0^1 (1-x)\,dx &= \left[x + \frac{x^2}{2}\right]_{-1}^0 + \left[x - \frac{x^2}{2}\right]_0^1 \\ &= 0 - \left((-1) + \tfrac{1}{2}\right) + \left(\left(1 - \tfrac{1}{2}\right) - 0\right) \\ &= \tfrac{1}{2} + \tfrac{1}{2} = 1.\end{aligned}$$

Alternatively, you might recognise f_{c} as the triangular function shown below. If so, the integral, which is the area under the triangle, is $\frac{1}{2} + \frac{1}{2} = 1$. This is because the area under the left-hand half of the triangle is half that of the unit square (which is $\frac{1}{2} \times 1$), and similarly for the right-hand half.



Graph of the function f_c

Either way, being non-negative and integrating to 1, this function is a density function as it stands.

(d) $f_{\text{d}}(x)$ is non-negative for all $x > 1$. However,

$$\int_1^\infty \frac{1}{x}\,dx = [\log x]_1^\infty = \log\infty - \log 1 = \infty - 0 = \infty.$$

This function is not, and cannot be made into, a density, because its integral is infinity.

Solution 2.2

$$F(x) = P(X \leq x) = P(X < x) + P(X = x) = P(X < x)$$

in the continuous case because then $P(X = x) = 0$.

Solution 2.3

(a) As f is a pdf and must therefore integrate to 1,

$$\begin{aligned}\lim_{x\to\infty} F(x) &= \lim_{x\to\infty}\int_{-\infty}^{x} f(y)\,dy \\ &= \int_{-\infty}^{\infty} f(y)\,dy = 1.\end{aligned}$$

(b) The derivative of $F(x)$ is the density $f(x)$, which is positive for all x on its support; it follows that F is increasing on its support.

Solution 2.4

(a) $F(x) = 0$ for all $x \leq 0$. Explicitly, for $x \leq 0$,

$$F(x) = \int_{-\infty}^{x} 0\,dy = 0.$$

(b) For $x > 0$,

Integration of an exponential

$$\begin{aligned}\int_{-\infty}^{x} f(y)\,dy &= \int_{-\infty}^{x} \lambda e^{-\lambda y}\,dy \\ &= \int_0^x \lambda e^{-\lambda y}\,dy \\ &= \lambda\left[\left(-\frac{1}{\lambda}\right)e^{-\lambda y}\right]_0^x \\ &= -\left[e^{-\lambda y}\right]_0^x \\ &= -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}.\end{aligned}$$

So

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 - e^{-\lambda x} & \text{if } x > 0, \end{cases}$$

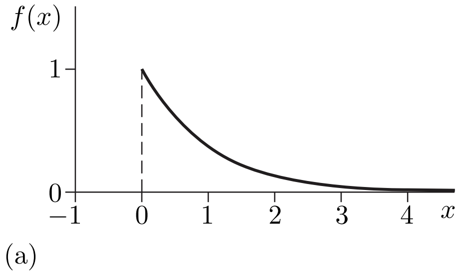
which can be written

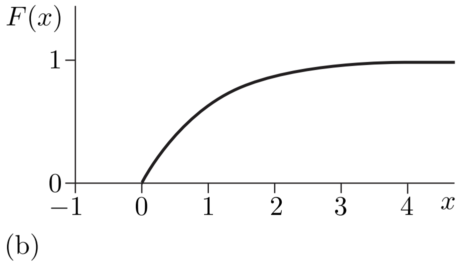
$$F(x) = 1 - e^{-\lambda x} \quad \text{on } x > 0.$$

(c) For $x > 0$,

$$F'(x) = \frac{d}{dx}(1 - e^{-\lambda x}) = -(-\lambda)e^{-\lambda x} = \lambda e^{-\lambda x} = f(x).$$

The pdf and cdf of the exponential distribution are shown below (for reference).





(a) Graph of pdf of exponential distribution with $\lambda = 1$;
(b) graph of cdf of exponential distribution with $\lambda = 1$

Solution 2.5

For $\beta < 1$, the power density is decreasing. For $\beta = 1$, the power density is flat/constant. For $\beta > 1$, the power density is increasing.

Solution 2.6

(a) On $0 < x < 1$,

$$F(x) = \int_0^x \beta y^{\beta-1} dy = \beta \left[\frac{y^\beta}{\beta}\right]_0^x = x^\beta - 0 = x^\beta.$$

Integration of a power

(b) On $a < x < b$,

$$F(x) = \int_a^x \frac{1}{b-a}\, dy = \frac{1}{b-a}\, [y]_a^x = \frac{x-a}{b-a}$$

(or you could have evaluated the area of the appropriate rectangle).

Solution 2.7

(a)
$$\begin{aligned} P(\tfrac{1}{4} \le X \le \tfrac{5}{8}) &= F(\tfrac{5}{8}) - F(\tfrac{1}{4}) \\ &= (\tfrac{5}{8})^2(3 - 2(\tfrac{5}{8})) - (\tfrac{1}{4})^2(3 - 2(\tfrac{1}{4})) \\ &= \tfrac{25}{64}(3 - \tfrac{5}{4}) - \tfrac{1}{16}(3 - \tfrac{1}{2}) \\ &= (\tfrac{25}{64})(\tfrac{7}{4}) - (\tfrac{1}{16})(\tfrac{5}{2}) \\ &= \tfrac{175-40}{256} = \tfrac{135}{256} = 0.53 \end{aligned}$$

(correct to two decimal places).

(b)
$$\begin{aligned} P(\log 2 \le X \le \log 4) &= F(\log 4) - F(\log 2) \\ &= (1 - e^{-\log 4}) - (1 - e^{-\log 2}) \\ &= (1 - \tfrac{1}{4}) - (1 - \tfrac{1}{2}) \\ &= \tfrac{3}{4} - \tfrac{1}{2} = \tfrac{1}{4}. \end{aligned}$$

Exponentials and logarithms

Solution 2.8

(a)
$$\begin{aligned} P(x_0 \le X < x_0 + \varepsilon) &= P(x_0 \le X \le x_0 + \varepsilon) \\ &= F(x_0 + \varepsilon) - F(x_0). \end{aligned}$$

(b) $$\lim_{\varepsilon \to 0} \frac{P(x_0 \leq X < x_0 + \varepsilon)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{F(x_0 + \varepsilon) - F(x_0)}{\varepsilon}$$

which equals the derivative of F at x_0 by the definition of differentiation given in Subsection 2.2 of Unit 1. And the derivative of F at x_0 is $f(x_0)$.

Solution 2.9

$$\begin{aligned} E\{k_1 h_1(X) + k_2 h_2(X)\} &= \int_a^b \{k_1 h_1(x) + k_2 h_2(x)\} f(x)\, dx \\ &= \int_a^b k_1 h_1(x) f(x)\, dx + \int_a^b k_2 h_2(x) f(x)\, dx \\ &= k_1 \int_a^b h_1(x) f(x)\, dx + k_2 \int_a^b h_2(x) f(x)\, dx \\ &= k_1 E\{h_1(X)\} + k_2 E\{h_2(X)\}. \end{aligned}$$

Linearity of integration

Solution 2.10

Set $h_1(X) = h(X)$ and $h_2(X) = 1$. Then Equation (2.2) shows that

$$E\{k_1 h(X) + k_2\} = k_1 E\{h(X)\} + k_2 E(1) = k_1 E\{h(X)\} + k_2$$

since $E(1) = 1$.

Solution 2.11

$E\{k\, h(X)\} = k\, E\{h(X)\}$ by setting $k_1 = k$ and $k_2 = 0$ in Equation (2.3).

Solution 2.12

$$\begin{aligned} E(X) &= \int_0^1 6x^2(1-x)\, dx \\ &= 6 \int_0^1 (x^2 - x^3)\, dx \\ &= 6 \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = 6((\tfrac{1}{3} - \tfrac{1}{4}) - (0-0)) = 6 \times \tfrac{1}{12} = \tfrac{1}{2}. \end{aligned}$$

Solution 2.13

$$\begin{aligned} E(X^r) &= \int_0^1 x^r \beta x^{\beta - 1}\, dx \\ &= \beta \int_0^1 x^{r+\beta-1}\, dx \\ &= \beta \left[\frac{x^{r+\beta}}{r+\beta}\right]_0^1 = \frac{\beta}{(r+\beta)}(1-0) = \frac{\beta}{r+\beta}. \end{aligned}$$

The formula for the uniform distribution arises when $\beta = 1$, namely $1/(r+1)$.

Solution 2.14

$$\mu_0 = E\{(X-\mu)^0\} = E(1) = 1.$$

By Equation (2.3),

$$\mu_1 = E(X-\mu) = E(X) - \mu = \mu - \mu = 0.$$

Solution 2.15

$$\mu_r = E\{(X-\mu)^r\} = E\left\{\sum_{i=0}^{r}\binom{r}{i}X^i(-\mu)^{r-i}\right\}$$

(by setting $a = X$ and $b = -\mu$ in the binomial expansion)

$$= \sum_{i=0}^{r}\binom{r}{i}(-1)^{r-i}\,E(X^i)\,\mu^{r-i}$$

(by the linearity of expectation).

Binomial expansion

Linearity of expectation

Solution 2.16

$$E(X) = \frac{\beta}{1+\beta}, \quad E(X^2) = \frac{\beta}{2+\beta},$$

and so

$$\begin{aligned} V(X) &= E(X^2) - \{E(X)\}^2 \\ &= \frac{\beta}{2+\beta} - \left\{\frac{\beta}{1+\beta}\right\}^2 \\ &= \frac{\beta(1+\beta)^2 - \beta^2(2+\beta)}{(2+\beta)(1+\beta)^2} \\ &= \frac{\beta + 2\beta^2 + \beta^3 - 2\beta^2 - \beta^3}{(2+\beta)(1+\beta)^2} \\ &= \frac{\beta}{(2+\beta)(1+\beta)^2}. \end{aligned}$$

Solution 2.17

(a) From Equation (2.3), $E(Y) = E(aX+b) = a\,E(X) + b$.

(b) The calculation of $V(Y)$ is similar to the calculation of $V(L)$ and $V(S)$ above:

$$\begin{aligned} V(Y) &= E\{(Y - E(Y))^2\} \\ &= E\{(aX + b - (a\mu + b))^2\} \\ &= E\{(aX - a\mu)^2\} \\ &= E\{a^2(X-\mu)^2\} = a^2\,E\{(X-\mu)^2\} = a^2\,V(X). \end{aligned}$$

Solution 2.18

(a) By Equations (2.6),

$$\begin{aligned} E(Y) &= \sigma\,E(X) + \mu = \sigma \times 0 + \mu = \mu, \\ V(Y) &= \sigma^2\,V(X) = \sigma^2 \times 1 = \sigma^2. \end{aligned}$$

(b) By Equation (2.7),

$$f_Y(y) = \frac{1}{\sigma}\,\phi\left(\frac{y-\mu}{\sigma}\right).$$

(c) $$\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}x^2\right)$$

gives

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\},$$

which is indeed the pdf of the normal distribution with mean μ and variance σ^2.

Solution 2.19

Starting from $M''_X(t) = (1+t^2)\exp(\frac{1}{2}t^2)$ as in Example 2.9, you should find that

Product rule

$$\begin{aligned} M_X^{(3)}(t) &= \frac{d}{dt}M''_X(t) \\ &= 2t\exp(\tfrac{1}{2}t^2) + (1+t^2)t\exp(\tfrac{1}{2}t^2) \\ &= (3t+t^3)\exp(\tfrac{1}{2}t^2) \end{aligned}$$

and

$$\begin{aligned} M_X^{(4)}(t) &= \frac{d}{dt}M_X^{(3)}(t) \\ &= (3+3t^2)\exp(\tfrac{1}{2}t^2) + (3t+t^3)t\exp(\tfrac{1}{2}t^2) \\ &= (3+6t^2+t^4)\exp(\tfrac{1}{2}t^2). \end{aligned}$$

Thus

$$E(X^4) = M_X^{(4)}(0) = (3+0+0)\times 1 = 3.$$

Solution 2.20

(a)

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \int_0^\infty \exp(tx)\, f(x)\, dx \\ &= \int_0^\infty \exp(tx)\, \lambda\exp(-\lambda x)\, dx \\ &= \lambda\int_0^\infty \exp\{(t-\lambda)x\}\, dx \\ &= \lambda\left[\frac{1}{t-\lambda}\exp\{(t-\lambda)x\}\right]_0^\infty \\ &= \frac{\lambda}{t-\lambda}(0-1) = \frac{\lambda}{\lambda-t} \end{aligned}$$

Integration of an exponential

(note that $t-\lambda<0$ gives $e^{(t-\lambda)x}=0$ for $x=\infty$).

(b)

$$M'_X(t) = \frac{\lambda}{(\lambda-t)^2}, \text{ so } E(X) = M'_X(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda},$$

$$M''_X(t) = \frac{2\lambda}{(\lambda-t)^3}, \text{ so } E(X^2) = M''_X(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2},$$

and therefore

$$V(X) = E(X^2) - \{E(X)\}^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

(c)

$$M_X^{(r)}(t) = \frac{d}{dt}M_X^{(r-1)}(t) = \frac{r(r-1)!\lambda}{(\lambda-t)^{r+1}} = \frac{r!\lambda}{(\lambda-t)^{r+1}}, \quad t<\lambda.$$

(d)

$$E(X^r) = M_X^{(r)}(0) = \frac{r!\lambda}{\lambda^{r+1}} = \frac{r!}{\lambda^r},$$

as required.