

**Continuous random variables** In this Chapter we shall see how the concepts we have looked at for discrete random variables can be used with continuous random variables.

## 1 Continuous pdfs

We have seen if  $X$  is a **continuous random variable** then  $X$  takes values from the real line.

### Exercise 5.1

Buses run from the university underpass to Lancaster every 10 minutes, without fail. I arrive at random at the bus stop. Let  $X$  be the length of time in minutes I wait for the next bus. So  $X$  takes realisations  $x$  where  $0 < x < 10$ .

Find the probability I have to wait (i) no more than 10 minutes, (ii) no more than 5 minutes, (iii) no more than 1 minutes, (iv) no more than 1 second.

Sol: 5.1

□

As the interval shrinks the probability gets very small, in fact, it tends to 0. More generally  $P(x < X \leq x + h) = h/10$  in this example, so that  $P(X = x) = \lim_{h \rightarrow 0} h/10 = 0$ . However some value of  $x$  must occur! Thus specifying  $P(X = x)$  is problematic.

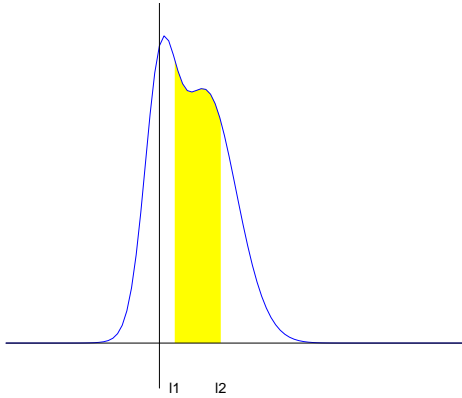
Better is to specify  $P(a < X \leq b)$ , the probability that  $X$  takes a value within a range. The **probability density function, pdf**,  $f(x)$ , of a continuous random variable,  $X$ , is defined implicitly by

$$P(a < X \leq b) = \int_a^b f(x) dx.$$

### Exercise 5.2

Indicate the probability that an observation on  $X$  lies between  $x_1$  and  $x_2$  as the area under the curve of  $f(x)$  between  $x = x_1$  and  $x = x_2$ .

Sol: 5.2



Generally, for a continuous random variable  $X$  with probability density function  $f(x)$ ,

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

□

End 05.19.4

## 2 Expectation and variance

The expected value (or mean) of a discrete random variable was given earlier as

$$E[R] = \sum_{r=0}^{\infty} r p(r).$$

We can think of this as an average of the different values that  $R$  may take, **weighted** according to their chance of occurrence. Evaluation of  $E[R]$  gives numerical value (e.g. 3.4) or a symbolic value (e.g.  $\lambda$ , or  $n\theta$ ).

The expected value (or mean) of a continuous random variable is similarly defined: The **expected value** of a continuous random variable  $X$  is

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

and the expected value of a real valued function  $g$

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

Summations in the discrete case go to integrals in the continuous case. The value of  $E[X]$  is often called  $\mu$ .

The variance of a discrete random variable was given earlier as

$$\text{var}(R) = E(R - E[R])^2 = \sum_{r=0}^{\infty} (r - E[R])^2 p(r).$$

We can think of this as a measure of the spread or dispersion of a random variable about the expectation. We define the variance of a continuous random variable similarly. Write  $\mu$  for  $E[X]$ . The **variance** of a continuous random variable  $X$  is

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

One can write this formula with  $E[X]$  instead of  $\mu$ , but it looks confusing.

Note that, in parallel to the discrete pmf,

$$\begin{aligned}\text{var}(X) &= E[(X - \mu)^2] \quad \text{def of } E \\ &= E[X^2 - 2X\mu + \mu^2], \quad \text{squareout} \\ &= E[X^2] + E[-2X\mu] + E[\mu^2], \quad \text{quadlin of } E \\ &= E[X^2] - 2\mu E[X] + E[\mu^2], \quad \text{lin of } E \\ &= E[X^2] - 2\mu E[X] + \mu^2, \quad \text{E of constants} \\ &= E[X^2] - \mu^2, \quad \text{def of } \mu \\ &= E[X^2] - (E[X])^2.\end{aligned}$$

### 3 The uniform pdf

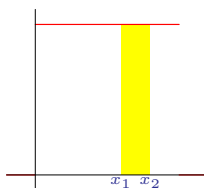
It is believed that the probability of a birth at a hospital does not depend on the time of day. A model for the time of birth,  $T$ , should give the same probability to being born at any time during 24 hours. Such a model is known as a continuous uniform distribution.

The function

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

is a probability density function since  $f(x) \geq 0$  and  $\int_{-\infty}^{\infty} f(x) dx = \int_0^1 1 dx = [x]_0^1 = 1$ . A random variable  $X$  is said to be **uniformly distributed** over the interval  $(a, b)$  if its pdf is given by

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



This is written as  $X \sim \text{Uniform}(a, b)$ .

The probability

$$\begin{aligned} P(x_1 < X \leq x_2) &= \int_{x_1}^{x_2} \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[ x \right]_{x_1}^{x_2} \\ &= \frac{x_2 - x_1}{b-a}. \end{aligned}$$

### Exercise 5.3

If  $X$  is uniformly distributed over  $(0, 10)$ , calculate the probability that

(a)  $X < 3$ , (b)  $X > 6$ , and (c)  $3 < X < 8$ .

Sol: 5.3

□

### Exercise 5.4

By considering a uniformly distributed random variable  $X$  show that the pdf  $f(x)$  is **not** a probability. [Unlike a pmf].

Sol: 5.4

□

### Exercise 5.5

Find the expected value of a uniform random variable.

Sol: 5.5

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \quad \text{def } E \\ &= \int_a^b x \frac{1}{b-a} dx \quad \text{non zero on } (a, b) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b && \text{integrate} \\
&= \frac{1}{2} \left[ \frac{b^2 - a^2}{b-a} \right] \\
&= \frac{a+b}{2}.
\end{aligned}$$

□

### Exercise 5.6

Find the expectation of the square of a uniform random variable.

Sol: 5.6

□

### Exercise 5.7

Find the variance of the uniform random variable.

Sol: 5.7

□

## 4 The exponential pdf

The exponential pdf often arises, in practice, as being the model for the amount of time until some specific event occurs. For example, the amount of time from now until an earthquake occurs, or until you receive the next telephone call which is a wrong number, or until the next flood tide happens.

*Example:* The following function is an example of an exponential pdf.

$$f(x) = \begin{cases} 2 \exp(-2x), & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

A random variable  $X$  is said to be **exponentially distributed** with rate parameter  $\lambda$ ,  $\lambda > 0$  if its pdf is given by

$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We say  $X \sim \text{Exp}(\lambda)$ .

### Exercise 5.8

Show that the the exponential pdf is well defined.

Sol: 5.8

**Have to show that**  $f(x) \geq 0$  **and**  $\int_x f(x)dx = 1$  **Now**  $\exp(t) \geq 0$  **whatever**  $t$  **so**  $\lambda \exp(-\lambda x) \geq 0$  **if**  $x \geq 0, \lambda \geq 0$

$$\begin{aligned} \int_{x=-\infty}^{\infty} f(x)dx &= 0 + \int_{x=0}^{\infty} \lambda \exp(-\lambda x)dx. \text{ Put } y = \lambda x \\ &= \int_{y=0}^{\infty} \exp(-y)dy, \\ &= [-\exp(-y)]_{y=0}^{\infty} \\ &= 0 - (-1) = 1. \end{aligned}$$

□

### Exercise 5.9

Find the probability that  $X \leq t$  when  $X \sim \text{Exp}(1)$ , for  $t = 0.5, 1, 2, 3$ .

R hint: `pexp(3,rate=1)`

Sol: 5.9

□

Exercise 5.10

Find the expectation of the exponential random variable.

Sol: 5.10

□

End 05.20.1

Exercise 5.11

Find the variance of the exponential random variable.

Sol: 5.11

$$\begin{aligned} E(X^2) &= \int_{x=0}^{\infty} x^2 \lambda \exp(-\lambda x) dx \\ &= \frac{1}{\lambda^2} \int_{y=0}^{\infty} y^2 \exp(-y) dy \\ &= \frac{2}{\lambda^2} \end{aligned}$$

using two integration by parts. Finally

$$\begin{aligned} \text{var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

□

The variance is the reciprocal of the rate squared.

Exercise 5.12

Assume that earthquakes occur world-wide at random but at an estimated average rate of once every 437 days. Assuming an exponential model for the inter-arrival times, find the probability that there will be no earthquakes in 2004-2006,

R hint: `1-pexp(1096,rate=1/437)`

Sol: 5.12



### Exercise 5.13

Find the median time between successive quakes. R hint: `qexp(0.5,rate=1/437)`

Sol: 5.13



### Exercise 5.14

Find the proportion of waiting times that are longer than expected.

R hint: `pexp(1/rate,rate=1/437)`

Sol: 5.14



## 5 The Normal pdf

The normal pdf was introduced by the French mathematician Abraham De Moivre in 1733. De Moivre used this distribution to approximate probabilities of winning in various games of chance involving coin tossing.

It was later used by the German mathematician Karl Gauss to predict the location of astronomical bodies and became known as the Gaussian distribution.

In the late nineteenth century statisticians started to believe that most data sets would have histograms with the Gaussian bell-shaped form and that all “normal” data sets would follow this form and so the curve came to be known as the **normal curve**. [gqview weplib/Gauss.html](#)

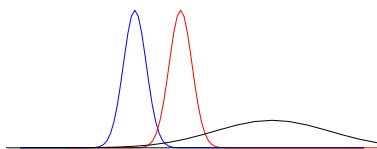
The normal pdf plays an important role in a result known as the **central limit theorem**. This theorem is one of the most important results in probability theory and gives a theoretical base to the empirical observation that many random phenomena seem to obey a normal distribution.

Examples of random phenomena which have a normal pdf are: soldiers’ chest measurements, babies’ birthweights, other measurement errors.

### Exercise 5.15

Three normal probability density functions.

Sol: 5.15



□

A random variable  $X$  is said to be **normally distributed** with parameters  $\mu$  and  $\sigma^2$  if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x - \mu)^2/2\sigma^2)$$

where  $-\infty < x < \infty$ . We write  $X \sim N(\mu, \sigma^2)$ . The **expectation** and the **variance** of the normal random variable can be shown, by integration, to be  $\mu$  and  $\sigma$  respectively: Result: If  $X \sim N(\mu, \sigma^2)$  then  $E[X] = \mu$  and  $\text{var}[X] = \sigma^2$ .

### Exercise 5.16

A normal model is proposed to model the variation in height  $H$  of women with parameters  $\mu = 160$  and  $\sigma^2 = 36$  measured in cm. Find the proportion of tall women, defined as over 180cm tall.

R hint: `1-pnorm(180,mean=160,sd=6)`

Sol: 5.16

$$\begin{aligned} P(180 < H < \infty) &= \int_{t=180}^{\infty} \frac{1}{\sqrt{2\pi}6} \exp(-(t - 160)^2/2 \times 36) dt \\ &= 0.000429 \end{aligned}$$

i.e.  $4/10000$  very small. Note that  $180 - 160 = 3.333\sigma$  □

In the above example we have expressed the proportion in terms of an integral and as the number of deviations from the mean.

The integral is impossible to calculate analytically. It is useful to express such probabilities in terms of a standardized random variable, with  $\mu = 0$  and  $\sigma = 1$ . Result: If  $X \sim N(\mu, \sigma^2)$ , then the random variable

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1),$$

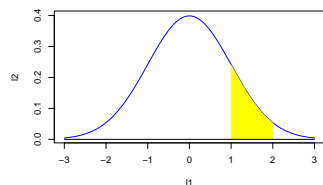
and conversely, if  $Z \sim N(0, 1)$ , then the random variable

$$X = \sigma Z + \mu \sim N(\mu, \sigma^2).$$

A random variable  $Z$  is said to have a **standard normal distribution** with mean 0 and standard deviation 1 if its pdf is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

where  $-\infty < z < \infty$ . We write  $Z \sim N(0, 1)$ . A sketch of the pdf of  $Z$  and  $P(1 < Z \leq 2)$ .



The area under the curve of the standard normal variable  $Z$  is given by

$$\Phi(z) = P(Z \leq z) = \int_{x=-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx.$$

Values of  $\Phi(z)$  are obtained from a table of standard normal probabilities or from computer software such as R:

$z$	-3.00	-2.33	-1.67	-1.00	-0.33	0.33	1.00	1.67	2.33	3.00
$\Phi(z)$	0.0013	0.0098	0.0478	0.1587	0.3694	0.6306	0.8413	0.9522	0.9902	0.9987

### Exercise 5.17

Find (a)  $P(Z \leq 1.00)$ , (b)  $P(Z \leq 1.96)$ , (c)  $P(Z \leq 3.333)$ .

R hint: `pnorm(1)`; `pnorm(1.96)`; `pnorm(3.333)`

Sol: 5.17

$$P(Z \leq 1.00) = \Phi(1) = 0.841$$

$$P(Z \leq 1.96) = \Phi(1.96) = 0.975$$

$$P(Z \leq 3.333) = \Phi(3.333) = 0.9995704$$

□

### Exercise 5.18

Find

(a)  $P(Z \geq 2.00)$ ,

(b)  $P(-1.33 \leq Z \leq 2.50)$ ,

R hint: `1-pnorm(2)`; `pnorm(2.50)-pnorm(-1.33)`; `pnorm(2.25)-pnorm(0.50)`

Sol: 5.18

$$P(Z \geq 2.00) = 1 - P(Z \leq 2.00) = 1 - \Phi(2) = 0.0228$$

$$\begin{aligned} P(-1.33 \leq Z \leq 2.50) &= P(Z \leq 2.50) - P(Z \leq -1.33) \\ &= \Phi(2.50) - \Phi(-1.33) = 0.902, \end{aligned}$$

using the additivity of probabilities;

□

### Exercise 5.19

Let  $X \sim N(10, 25)$ . Standardise this rv to find

(a)  $P(X \leq 20)$ , (b)  $P(0 \leq X \leq 15)$ .

R hint: `pnorm(2)` ; `pnorm(20, mean=10, sd=5)`; `pnorm(1)-pnorm(-2)`;  
`pnorm(15, mean=10, sd=5)-pnorm(0, mean=10, sd=5)`

Sol: 5.19

□

Exercise 5.20

If  $X$  is a Normal random variable with parameters  $\mu = 3$  and  $\sigma^2 = 4$  find  
(a)  $P(2 < X < 5)$  and (b)  $P(X > 0)$ .

R hint: `pnorm(2) ; pnorm(20, mean=10, sd=5); pnorm(1)-pnorm(-2);`  
`pnorm(15, mean=10, sd=5) - pnorm(0, mean=10, sd=5)`  
`pnorm(1)-pnorm(-0.5); 1- pnorm(-1.5);`

Sol: 5.20

