6 Continuous probability models

6.1 Introduction

In this section we will deal with continuous sample spaces so that the random variables take values in the set of real numbers. Of course, continuous random variables are models for continuously varying quantities that we might measure, like weight, concentration, or an interval of time. The change from *discrete* to *continuous* sample spaces and random variables brings certain changes to the underlying maths.

6.2 Cumulative distribution functions and probability density functions

Previously, when X was a discrete random variable, we considered events such as $A = \{X = x\}$ (or more accurately, $A = \{s \in S : X(s) = x\}$). For *continuous* random variables, such events usually have zero probability for the following reason. Suppose X is someone's height, and using a metre ruler we measure $x = 170$cm. By this we really mean $169.5 \leq x < 170.5$cm because with this ruler we can only measure to the nearest cm. Suppose we now measure the height with a set of medical equipment to the nearest mm, and observe $x = 170.0$cm. Again, this really means $169.95 \leq x < 170.05$cm, due to the accuracy of the instrument. This can be repeated with ever more sophisticated measurements. So the statement $X = x$ means "X *is exactly* x up to arbitrarily precise measurement". The probability of observing someone with height *exactly* 170cm is zero. Instead, it is necessary to consider the probability of events such as $169.5 \leq X < 170.5$.[2]

It is therefore not possible to specify the distribution of a random variable via a probability mass function

$$p(x) = \Pr(X = x)\,.$$

However, our definition of the cumulative distribution function (sometimes simply referred to as the distribution function) still makes sense:

$$F(x) = \Pr(X \leq x)\,.$$

Using the cumulative distribution function we can work out the probability of being close to a value x through

$$\Pr(x < X \leq x + h) = F(x + h) - F(x).$$

Now consider the gradient (or slope) of F at x which, for small h, is approximately

$$\frac{F(x+h) - F(x)}{h} \approx \text{gradient of } F \text{ at } x.$$

As h approaches zero, this approximation becomes exact and we write

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \frac{d}{dx} F(x)$$

[2] Note that because $\Pr(X = x) = 0$, it follows that $\Pr(X \leq x) = \Pr(X < x)$ and $\Pr(X \geq x) = \Pr(X > x)$. This is not generally true with discrete random variables.

where $\frac{d}{dx}$ represents the *derivative* with respect to x. Normally we denote this derivative by $f_X(x)$ or $f(x)$, i.e.

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

and call $f(x)$ the *probability density function* (PDF). We can represent continuous probability models through either the cumulative distribution function $F(x)$ or the probability density function $f(x)$. From the way we have defined the PDF, note that

$$\Pr(x < X \leq x + h) \approx f(x) \times h$$

for small h and so the probability density function $f(x)$ is proportional to the probability of lying in a very small interval $(x, x + h]$.

Because $F(x)$ is an *increasing function* it follows that on a graph of $f(x)$, the PDF curve will always lie on or above the x-axis, i.e. $f(x) \geq 0$ for all x. Moreover, probabilties are represented by areas under the curve. For example, the probability $\Pr(169.5 \leq X < 170.5)$ would be the area above the x-axis and below the PDF curve between the limits at 169.5 and 170.5. More formally,

$$\Pr(a \leq X \leq b) = F(b) - F(a) = \text{area under PDF curve} = \int_a^b f(x)\,dx$$

where the expression on the far right-hand-side is the *integral* of f over the interval $[a, b]$. Also note that

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

and so the total area under the PDF curve is 1. Finally, to derive the distribution function from the PDF, we have

$$F(x) = \int_{-\infty}^{x} f(u)\,du.$$

Let us contrast some of the properties of the PDF for continuous random variables with properties of the probability mass function (pmf) for discrete random variables (see Section 5.1.2). For a pmf $p(x)$, we require that $\sum_{\text{all } x} p(x) = 1$. For a PDF $f(x)$, we require $\int_{-\infty}^{\infty} f(x)\,dx = 1$. To work out the probability that a discrete random variable X lies in some set A, we add up the probabilities for all the elements in that set, $\Pr(X \in A) = \sum_{x \in A} p(x)$. To work out the probability that a continuous random variable X lies in some interval $[a, b]$, we integrate $f(x)$ over the interval, $\Pr(X \in [a, b]) = \int_a^b f(x)\,dx$. So the PDF in the continuous case is analogous to the pmf in the discrete case: we simply replace sums with integrals.

6.2.1 Example

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment can be modelled by a continuous random variable X having the probability density function

$$f(x) = \begin{cases} x^2/3, & -1 < x < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

1. Verify that $f(x)$ is a density function;
2. Find $\Pr(0 < X < 1)$;
3. Find the CDF $F(x)$ and use it to verify your result in 2.

Solution.

3. For $x \leq -1$, we have

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

For $-1 < x < 2$, we have

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\,du \\ &= \int_{-\infty}^{-1} f(u)\,du + \int_{-1}^{x} f(u)\,du \\ &= \int_{-\infty}^{-1} 0\,du + \int_{-1}^{x} \frac{u^2}{3}\,du \\ &= 0 + \left[\frac{u^3}{9}\right]_{-1}^{x} \\ &= \frac{x^3}{9} + \frac{1}{9} = \frac{1}{9}(x^3+1). \end{aligned}$$

Finally for $x \geq 2$ we have

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\,du \\ &= \int_{-\infty}^{-1} f(u)\,du + \int_{-1}^{2} f(u)\,du + \int_{2}^{x} f(u)\,du \\ &= \int_{-\infty}^{-1} 0\,du + \int_{-1}^{2} \frac{u^2}{3}\,du + \int_{2}^{x} 0\,du \\ &= 0 + 1 + 0 = 1. \end{aligned}$$

Therefore the CDF is given by

$$F(x) = \begin{cases} 0, & x \leq -1, \\ (x^3+1)/9 & -1 < x < 2, \\ 1, & x \geq 2. \end{cases}$$

Using the CDF we can then compute

$$\Pr(0 < X < 1) = F(1) - F(0) = \frac{1^3+1}{9} - \frac{0^3+1}{9} = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

which, of course, agrees with the result obtained using the PDF.

6.3 Expectation

Since the definition of the probability mass function doesn't make sense for a continuous random variable, we need to re-define the expectation. For a continuous random variable X we have:

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

i.e. we weight each value x by the *density* rather than the mass and then integrate rather than sum. Equivalently, we can think of the expectation as the total area under the graph of the function $h(x) = x \times f(x)$. Similarly, if $g(X)$ is a function of X then its expectation is

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

The definition of the variance of a random variable is unchanged in the continuous case and the properties of expectations and variances listed in Section 5.5 also remain true.

6.3.1 Example

For the example in Section 6.2.1, compute the expectation $\mathrm{E}[X]$ and variance $\mathrm{Var}\,(X)$.

Solution. ✐

6.4 The uniform distribution

Suppose every value between two fixed limits a and b is equally likely.

Then the probability density function is constant and is defined by

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

The corresponding distribution function is

$$F(x) = \Pr(X \leq x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases} .$$

Derivation:

A random variable with this distribution is called a *uniform* random variable, and we write $X \sim \text{Unif}(a, b)$, or $X \sim U(a, b)$.

The PDF and CDF are obtained via the R commands `dunif` and `punif` respectively. Figure 1 shows the PDF and CDF for $\text{Unif}(0, 1)$. The figure was generated using the commands

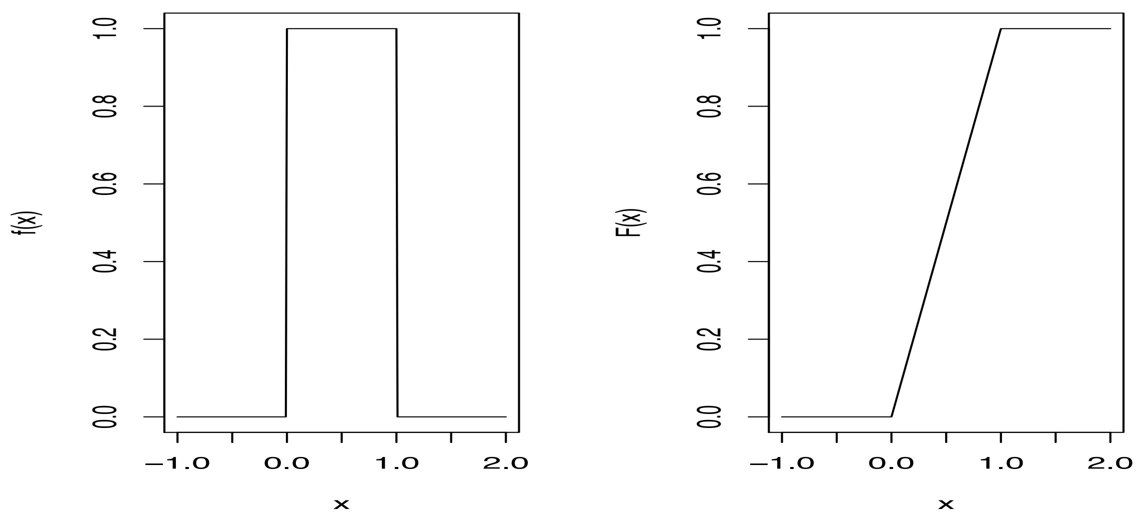

Figure 1: The PDF and CDF for Unif$(0, 1)$

```
a = 0; b = 1;
x = seq(a-1,b+1,0.01)
plot(x,dunif(x,a,b),type="l",ylab="f(x)")
plot(x,punif(x,a,b),type="l",ylab="F(x)")
```

The expectation and variance of $X \sim \text{Unif}(a, b)$ are

$$E[X] = \frac{a+b}{2} \qquad \text{and} \qquad \text{Var}\,(X) = \frac{(b-a)^2}{12}.$$

Derivation: ✐

Now, to compute the variance we need

$$\begin{aligned}
\mathrm{E}\big[X^2\big] &= \int_a^b \frac{x^2}{b-a}\,dx \\
&= \left[\frac{x^3}{3(b-a)}\right]_a^b \\
&= \frac{b^3-a^3}{3(b-a)} \\
&= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} \qquad \text{using the difference of two cubes} \\
&= \frac{a^2+ab+b^2}{3},
\end{aligned}$$

and so

$$\begin{aligned}
\mathrm{Var}\,(X) &= \mathrm{E}\big[X^2\big] - \mathrm{E}[X]^2 \\
&= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} \\
&= \frac{4(a^2+ab+b^2)-3(a^2+2ab+b^2)}{12} \\
&= \frac{b^2-2ab+a^2}{12} \\
&= \frac{(b-a)^2}{12}.
\end{aligned}$$

Examples:

1. Waiting time. At a particular train station, trains arrive every ℓ minutes. If you arrive at the station at a random time which is uniformly distributed between train times then the time T you have to wait until the next train would be $T \sim \mathrm{Unif}(0, \ell)$.

2. Random number generation. Uniform random numbers can be used to generate random numbers for any other distribution for which the CDF is known – and this is widely implemented on computers. If $X \sim \mathrm{Unif}(0, 1)$ then it can be shown that $Y = F^{-1}(X)$ has distribution function F, and so Y has the desired distribution. This is because

$$\begin{aligned}
\Pr(Y \le y) &= \Pr\big(F^{-1}(X) \le y\big) = \Pr(X \le F(y)) \\
&= F(y)
\end{aligned}$$

   This last line follows from the definition of the cumulative distribution function for X: $\Pr(X \le x) = x$ for $0 \le x \le 1$.

Exercise:
Suppose $Y \sim U(6, 18)$. Find

(a) $\Pr(Y \leq 14)$

(b) $\Pr(Y > 8)$

(c) $\Pr(8 < Y \leq 14)$

(d) $E[Y]$

(e) $SD(Y)$

(f) $\Pr(Y = 12)$

6.5 The exponential distribution

The random variable X has an *exponential distribution* with parameter $\lambda > 0$, written

$$X \sim \text{Exp}(\lambda),$$

if it has PDF

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

The corresponding CDF is given by

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

Derivation:

The PDF and CDF for an Exp(1) random variable are shown in Figure 2.

To generate the PDF and CDF in R, we use:

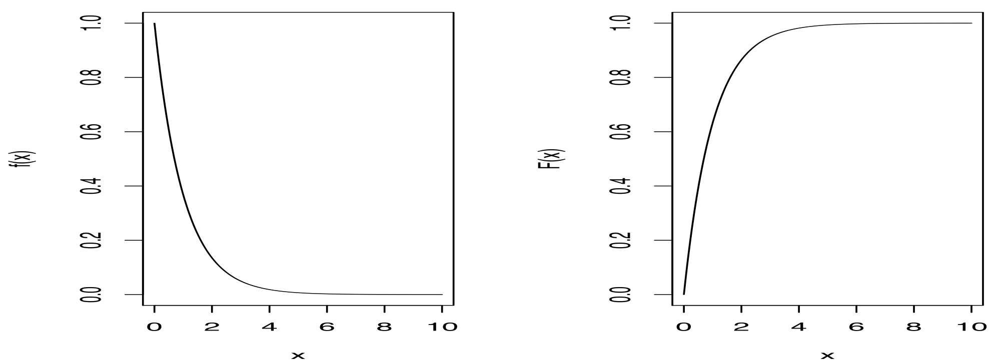

Figure 2: The Exp(1) PDF and CDF.

```
x = seq(0,10,0.01)
plot(x,dexp(x,1),type="l",ylab="f(x)")
plot(x,pexp(x,1),type="l",ylab="F(x)")
```

The expectation and variance of the exponential distribution are

$$\mathrm{E}[X] = \frac{1}{\lambda} \qquad \text{and} \qquad \mathrm{Var}\,(X) = \frac{1}{\lambda^2}.$$

Derivation: ✐

Exponential random variables are often used to model the times between events or the lifespan of organisms or mechanical parts. This is because the exponential distribution has the following property:

$$\Pr(X \geq t + s | X > t) = \Pr(X > s) \, .$$

So if you know the time interval X is greater than t then the clock is effectively 'restarted' at t. The geometric distribution also has this property, often called the 'memoryless' property.

Example:
Suppose an exponential distribution is used to model the time (in minutes) until cell division in a colony of bacteria, and that bacteria divide on average once every 30 minutes. Calculate the following:

1. The probability a bacterium takes more than 1 hour to divide.

2. The probability that a bacterium divides in the next 15 minutes, given that we observe it as undivided after 1 hour.

6.5.1 The Poisson process

Suppose we are modelling some events that occur along a continuous interval as a Poisson process with rate λ. For example, we might be modelling defects in a stretch of wire or phone calls to a call centre over time. Recall that the number of events up to time (or distance etc) t is

$$\begin{aligned} X_t &= (\text{number of events up to time } t) \\ &\sim \text{Po}(\lambda t). \end{aligned}$$

Instead of the *number* of events, we might consider the time (or distance etc) between successive events, often called the *inter-arrival time*. It can be shown that the inter-arrival time has an exponential distribution with parameter λ.

Although a proof is beyond the scope of this course, it is straightforward to verify the result for the time T_1 until the first event:

$$\Pr(T_1 \leq t) = 1 - \Pr(T_1 > t) = 1 - \Pr(X_t = 0) = 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = 1 - e^{-\lambda t},$$

which is the distribution function of an exponential random variable with parameter λ. Therefore

$$T_1 \sim \text{Exp}(\lambda).$$

6.6 The Normal Distribution

The Normal distribution is by far the most important distribution in the theory and practice of statistics. A Normally distributed random variable $X \sim N(\mu, \sigma^2)$ has PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

The distribution is characterized by the two constants μ and σ which determine the mean and variance:

$$\text{E}[X] = \mu \qquad \text{and} \qquad \text{Var}\,(X) = \sigma^2.$$

The PDF has a characteristic 'bell-shaped' curve, as illustrated by Figure 3. The figure also shows the effect of changing the two parameters μ and σ.

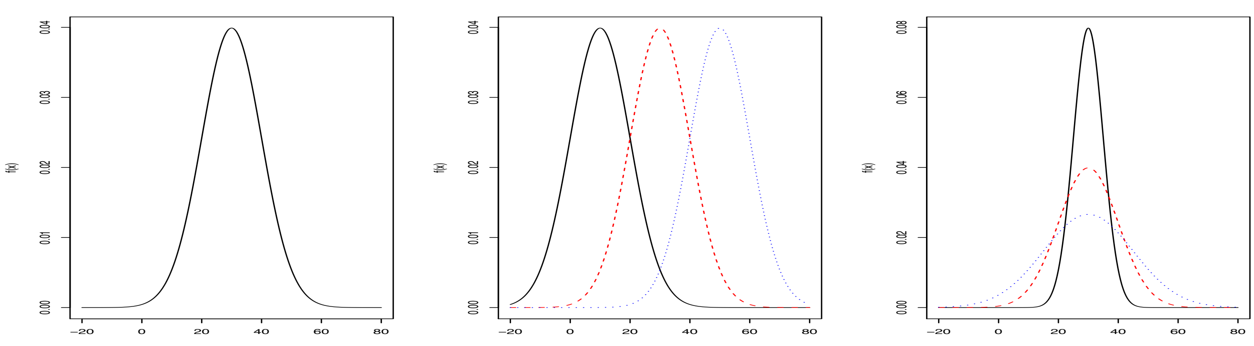

Figure 3: (a) Normal PDF with mean 30 and sd 10. (b) Normal PDFs with mean 10,30,50 and sd 10. (c) Normal PDF with mean 30 and sd 5,10,15.

The PDF can be obtained in R via the `dnorm` command. The CDF for the Normal distribution does not have a closed algebraic form – there is no exact algebraic equation we can write down that describes the area under the PDF. Numerical methods are necessary to evaluate $F(x)$ for each value of x. Fortunately, R does this automatically, and the CDF can be obtained via the `pnorm` command.

Example:
Suppose the weight of newborn babies (in kg) can be modelled by a Normal distribution with mean 3.5 and standard deviation 0.5. What is the probability that a baby weighs more than 4.5 kg?

We can use R to obtain the CDF at $x = 4.5$:

```
pnorm(4.5,mean=3.5,sd=0.5) # Gives 0.9772499
```

Then $\Pr(X > 4.5) = 1 - F(4.5) = 1 - 0.9772 = 0.0028$.

6.6.1 The Standard Normal distribution

A *standard* Normal random variable, $Z \sim N(0,1)$, is a Normal random variable with mean zero and variance equal to one. The PDF is denoted $\phi(z)$ and is

$$\phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}, \qquad -\infty < z < \infty,$$

which is symmetric about zero. The CDF is denoted $\Phi(z)$ and is given by

$$\Phi(z) = F_Z(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \phi(x)dx.$$

As noted above, there is no algebraic expression for $\Phi(z)$, so it is usually obtained from tables or using a package like R. However, the symmetry of the PDF ensures the following properties:

$$\Phi(-\infty) = 0,\ \Phi(\infty) = 1,\ \Phi(0) = \frac{1}{2},\ \Phi(-z) = 1 - \Phi(z).$$

6.6.2 Translation and scaling

Suppose that Z is a standard Normal random variable and a, b are fixed constants with $b > 0$. If we define $X = a + bZ$ it follows that

$$F_X(x) = \Pr(X \le x) = \Pr(a + bZ \le x) = \Pr\left(Z \le \frac{x-a}{b}\right) = \Phi\left(\frac{x-a}{b}\right).$$

By differentiating, we obtain

$$f_X(x) = \frac{1}{b}\phi\left(\frac{x-a}{b}\right) = \frac{1}{b\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}.$$

It follows that $X \sim N(a, b^2)$. Thus translating and scaling the standard Normal distribution gives rise to Normal random variables with arbitrary mean and standard deviation.

Example:
Suppose that the weight (in kg) of parcels sent by a particular online retailer can be modelled by $X \sim N(5.5, 1.5^2)$. Given that $\Phi(1.96) = 0.975$, find a range of weights that contains 95% of parcels.

Solution.

6.6.3 Quantiles

The result above (that $a + bZ$ is $N(a, b^2)$ when Z is standard Normal) implies that Normal distributions can be characterized in the following way.

- $\Pr(-1 < Z < 1) = 0.683$ when Z is standard Normal. It follows that for *any* Normal variable X, 68.3% of observations will lie between ± 1 standard deviations of the mean i.e. $\mu \pm \sigma$.
- Similarly, we saw above that $\Pr(-1.96 < Z < 1.96) = 0.95$ when Z is standard Normal. It follows that 95% of observations of X will lie between ± 1.96 standard deviations of the mean.
- In general the values q that determine any given probability p i.e. $\Pr(Z \leq q) = p$ are called *quantiles*. The set of quantiles for the standard Normal distribution fix the distribution of Normal variables with arbitrary mean and standard deviation.

This idea can be used to test whether a set of observations appears to come from a Normal distribution. A plot can be drawn of the data in the following way. For each observation you plot a point on the graph. For the ith smallest observation $x_{(i)}$ the vertical co-ordinate corresponds to the value of the observation, $x_{(i)}$; the horizontal co-ordinate corresponds to the standard Normal quantile q_i where $\Pr(Z \leq q_i) = (i - 0.5)/n$, in other words the quantile corresponding to how far through the ordered data the observation lies.[3] The resulting points will be scattered around a straight line when the data are Normally distributed. This kind of plot is called a Q-Q plot and is routinely used to check whether data are Normally distributed.

Example:
In R, generate a random sample of $n = 50$ observations from $X \sim N(10, 3^2)$ and produce a Q-Q plot with the following commands:

```
x = rnorm(50,10,3)
qqnorm(x)
qqline(x) # Adds a line which passes through the lower and upper quartiles.
```

[3] You might think that $x_{(i)}$ should be plotted against the quantile q_i such that $\Pr(Z \leq q_i) = i/n$ for $i = 1, \ldots, n$. But the last of these, n/n, corresponds to the 100th percentile of the Normal distribution, i.e. the maximum possible value, which is infinite. The shift by 0.5 addresses this problem.

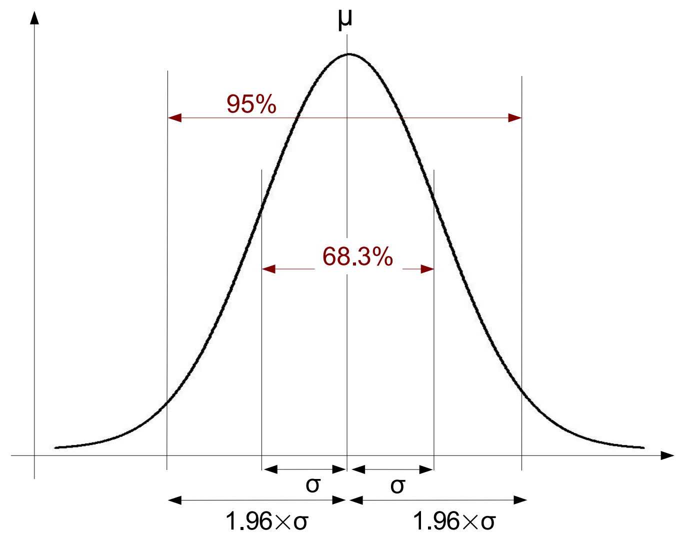

Figure 4: Characterizing the Normal distribution via quantiles.

The resulting plot is shown in Figure 5. As expected, the points cluster closely around the diagonal line. When the Normal distribution is a poor model for the data, we see marked departures from this diagonal line.

6.6.4 Sums of Normal random variables

If X and Y are two independent Normal random variables, $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then the sum $Z = X + Y$ is also Normal with

$$Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$

The proof requires mathematics beyond the scope of this course, but nonetheless, we'll use the result many times.

6.7 The sample mean and Central Limit Theorem

Let's reconsider the properties of the sample mean, originally seen in Section 5.6 of the notes. Suppose $X_1, X_2, \ldots, X_n$ are *independent and identically distributed* (IID) with common expectation μ and common variance σ^2. At this stage we make no kind of assumption about whether the X_i are discrete or continuous — our results will apply to both. An observed sample mean is a measurement of the random variable $\bar{X}$, where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We saw previously that

$$\mathrm{E}[\bar{X}] = \mu \qquad \text{and} \qquad \mathrm{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

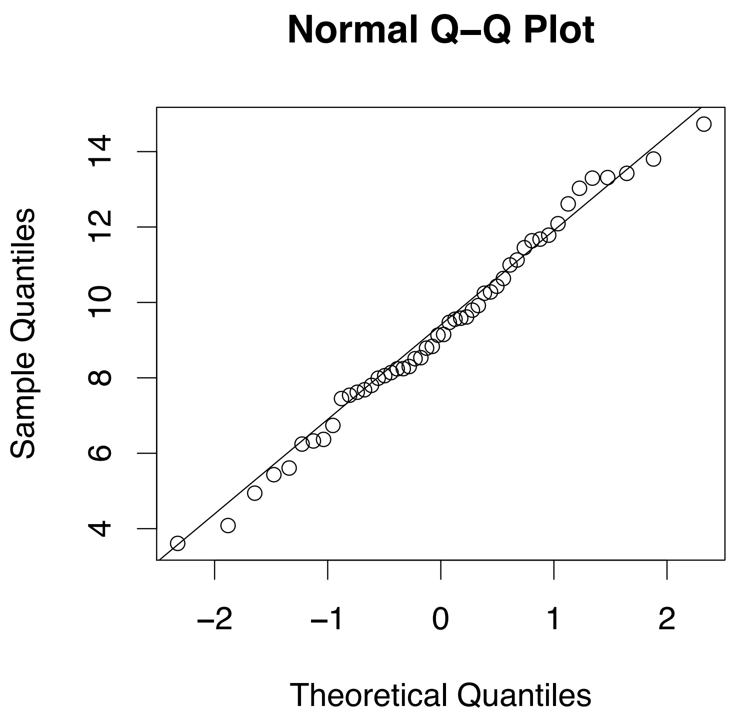

Figure 5: Normal Q-Q plot for simulated data.

As we discussed, $\bar{X}$ represents a model for the observed sample mean $\bar{x}$. Thus if we have a model X_i for one measurement from the population, the result above tells us that the sample mean will be the same as the population mean *on average*. It also tells us how much we can expect the sample mean to *vary* around the population mean: the larger the sample, the lower the variability. So if we take lots of random samples of size n and for each compute an observed sample mean $\bar{x}$, these sample means will cluster around the population mean. The clustering will be tighter for larger samples.

Example:
Let's illustrate this idea with some simulated data. The following table shows four random samples of size 10 simulated from the Unif$(0, 1)$ distribution. The sample mean and sample standard deviation for each sample have been computed and the data have been rounded to two decimal places for clarity.

	Sample 1	Sample 2	Sample 3	Sample 4
	0.41	0.80	0.54	0.98
	0.52	0.24	0.58	0.12
	0.52	0.50	0.41	0.18
	0.83	0.30	0.38	0.92
	0.88	0.32	0.42	0.65
	0.94	0.40	0.57	0.18
	0.52	0.20	0.04	0.67
	0.67	0.64	0.43	0.57
	0.07	0.32	0.89	0.14
	0.63	0.45	0.94	0.14
mean	0.60	0.42	0.52	0.45
SD	0.25	0.19	0.26	0.34

In effect, the table shows four realizations of the random variable $\bar{X}$. As we'd expect, these are distributed around the true value of the population mean, which in this case is $(a+b)/2 = 1/2$. Similarly, the sample standard deviations vary around the true population standard deviation of $\sqrt{(b-a)^2/12} = \sqrt{1/12} = 0.29$. How much do the observed sample means vary? Well, $\text{SD}\{0.60, 0.42, 0.52, 0.45\} = 0.08$. Our theory says that the sample mean has variance σ^2/n, so we would expect the standard deviation of the sample means to be $0.29/\sqrt{10} = 0.09$. That matches our example really well – even for these relatively small samples and numbers of replicates.

The data in this example were generated in R using the following commands:

```
samples=matrix(NA,10,4) # Create a matrix with 10 rows and
                        # 4 columns.
for(i in 1:4) samples[,i]=runif(10) # Simulate 4 random samples
                                    # of size 10 from Unif(0,1).
apply(samples,2,mean) # Compute the column (sample) means.
apply(samples,2,sd) # Compute the column (sample) std devs.
```

6.7.1 Distribution of the sample mean for Normal random variables

Suppose that $X_1, X_2, \dots, X_n$ are IID $N(\mu, \sigma^2)$ random variables. Recall that the sum of two Normal RVs is itself another Normal random variable. It follows that

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

In other words, the sample mean $\bar{X}$ itself is distributed as a Normal random variable, with the (population) mean and variance we'd expect.

Example:
Insulin resistance is a condition in which the body's cells don't react to insulin. Previous experiments using the "gold standard" measurement technique have revealed that insulin mediated glucose uptake can be modelled well by a Normal distribution with mean 8.2 mg/kg per minute and standard deviation 1.9 mg/kg per minute. In an experiment with a new,

cheaper measurement technique, 8 measurements are taken and a sample mean of 6.6 mg/kg per minute is obtained. Is this compatible with the original hypothesised distribution?

Solution.

6.7.2 The Central Limit Theorem

We have seen that for IID Normal random variables the sample mean itself is Normally distributed. The Central Limit Theorem is a very important and deep result about sample means when the variables $X_1, \ldots, X_n$ have *almost any distribution*. It says that no matter what the underlying distribution of the individual measurements is, the sample mean is *always* approximately Normal.

The theorem can be stated more formally as follows. Suppose that $X_1, \ldots, X_n$ are IID with mean μ and variance σ^2. Then (as we saw above) $\bar{X}$ has expectation μ and variance σ^2/n. Moreover, $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$:

$$\Pr\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq x\right) \longrightarrow \Phi(x)$$

as $n \to \infty$.

Why is this result important? It gives us an approximate distribution for the sample mean $\bar{X}$ independent of the underlying distribution for the individual measurements. If we have estimates of the population mean and standard deviation, then we have an approximate distribution for sample means (provided the sample size is large enough).

Example:
Suppose we observe 75 Unif$(0, 1)$ random variables. What is the probability that the sample mean lies between 0.45 and 0.55?

Solution.