Chapter 3
Models for Data II

In this chapter, more models for random variation within a population are described and characteristics of the models are identified. In particular, the notions of population moments and population quantiles are introduced.

In Chapter 1, some straightforward ways of examining data were described. The methods were applied to samples of data to shed light on particular questions. The techniques introduced included a variety of graphical methods for picturing the data, and some numerical quantities for summarizing the data in different ways—such as the sample mean and sample standard deviation. Very often, however, the sample itself is not of any special interest. What we really want to do is make some statement about the entire population from which the sample is drawn. Much of statistics is concerned with making inferences from a sample to the population: how to estimate population parameters from samples, how to evaluate the confidence we should have in the results, how to design experiments or draw samples so that we can obtain the most accurate estimates, and so on.

In Chapter 2, probability models were introduced. In particular, probability distributions were introduced as models for random variation. Some sections of that chapter focused on general properties (such as probability mass functions, probability density functions and cumulative distribution functions), while others concentrated on specific models and their properties. The distinction between discrete and continuous random variables was drawn. The ideas were illustrated by the Bernoulli and binomial distributions (both discrete), and the normal distribution (continuous). This chapter continues in the same vein: some sections introduce general properties of population models, while others deal with particular models.

General properties are introduced in Sections 3.1, 3.2 and 3.5. In Section 3.1, population analogues of the sample mean and sample variance introduced in Chapter 1 are discussed. Chapter 1 also introduced sample quartiles and the sample median. There are again corresponding population measures and these are discussed in Section 3.5. Just as sample statistics can be used to describe aspects of the shape of the distribution of values in the sample, so the corresponding population parameters describe features of the population models.

Section 3.2 concerns a fundamental concept, introduced briefly in Chapter 2, namely the concept of independence. The idea of independence—that the outcome of one trial (measurement, score, observation, and so on) does not influence the outcome of another—is vitally important. A great many statisti-
cal techniques assume that observations are independent (for example, that the outcome of two tosses of a coin are independent, or that the responses of two patients in a test of a new medicine are independent). If this assumption turns out to be false then more sophisticated statistical techniques have to be devised.

The two particular models developed in this chapter are the geometric distribution and the uniform distribution. The former is outlined in Section 3.3 and arises, like the binomial distribution, from a statistical experiment involving a sequence of independent trials each of which has a binary outcome (either success or failure). Here, however, interest focuses not on the total number of successes in a sequence of trials, but on the number of trials needed until the first success is obtained.

The uniform distribution has already been used in Chapter 2, page 62 as a model for the outcomes observed when a perfect die is rolled. In Section 3.4, the uniform distribution is discussed in more detail.

### 3.1 Population means and variances

In Chapter 1, we saw how certain important characteristics of samples of data can be encapsulated by various numerical summaries. Examples included the sample mean, the sample median and the sample standard deviation. If a bar chart or histogram, as appropriate, were used to represent the variation in the data, then these numerical summaries could be seen to be describing various aspects of the shape of the bar chart or histogram. For example, the sample mean and median tell us about the ‘location’ of the sample in certain senses; the ‘spread’ or ‘dispersion’ of the data can be measured by the sample standard deviation, or alternatively by the sample interquartile range; and the sample skewness measures asymmetry of the data.

Given a sample of values, we know how to compute numerical descriptors such as those mentioned above: you had some practice at doing so in Chapter 1. However, if we were to assume a probability model as adequate for the variation within a population, how might we define and calculate similar numerical summaries for the population such as the population mean and population standard deviation? These are sometimes required for a statistical analysis.

Here is an example of the sort of context within which we might need to perform a comparative test, and for which we therefore first need a model.

**Example 3.1 Origins of the Etruscan empire**

The origins of the Etruscan empire remain something of a mystery to anthropologists. A particular question is whether Etruscans were native Italians or immigrants from elsewhere. In an anthropometric study, observations on the maximum head breadth (measured in mm) were taken on 84 skulls of Etruscan males. These data were compared with the same skull dimensions for a sample of 70 modern Italian males. The data are summarized in Table 3.1.
Table 3.1  Maximum head breadth (mm)

<table>
<thead>
<tr>
<th>84 Etruscan skulls</th>
<th>70 modern Italian skulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>141 148 132 138 154 142 150 146 155 158 150</td>
<td>133 138 130 138 134 127 128 138 136 131 126</td>
</tr>
<tr>
<td>140 147 148 144 150 149 145 149 158 143 141</td>
<td>120 124 132 132 125 139 127 129 132 116</td>
</tr>
<tr>
<td>144 144 126 140 144 142 141 140 145 135 147</td>
<td>125 130 129 125 136 131 132 127 129 132 116</td>
</tr>
<tr>
<td>146 141 136 140 146 142 137 148 154 137 139</td>
<td>134 125 128 139 132 130 132 128 139 135 133</td>
</tr>
<tr>
<td>143 140 131 143 141 149 148 135 148 152 143</td>
<td>128 130 130 143 144 137 140 136 135 128 139</td>
</tr>
<tr>
<td>144 141 143 147 146 150 132 142 142 143 153</td>
<td>151 133 138 133 137 140 130 137 134 130 148</td>
</tr>
<tr>
<td>149 146 149 138 142 149 142 137 134 144 146</td>
<td>135 138 135 138</td>
</tr>
</tbody>
</table>

The statistical procedures for such a comparison will be described in Chapter 8. (In fact, a simple comparative boxplot as shown in Figure 3.1 suggests marked differences between the Etruscan skulls and those of modern Italian males.)

Figure 3.1  Comparative boxplot, two skull samples

Histograms for the two skull samples are shown in Figure 3.2. These suggest that in either case a normal model as described in Chapter 2, Section 2.4, might be adequate for the purposes of a comparative test, but with different indexing parameters in the two cases.

Figure 3.2

Histograms, two skull samples: (a) 84 Etruscan males (b) 70 modern Italian males
The two usual sample summaries for location and dispersion are the sample mean and sample standard deviation. For these data, we have the following summary statistics:

84 Etruscan skulls: \( \bar{x} = 143.8, \ s_x = 6.0; \)
70 Italian skulls: \( \bar{y} = 132.4, \ s_y = 5.7. \)

These values suggest that the skulls of the modern Italian male are, on average, both narrower and slightly less variable than those of the ancient Etruscan. If the breadth of Etruscan skulls may be adequately modelled by a normally distributed random variable \( X \) and the breadth of modern Italian skulls by a normally distributed random variable \( Y \), then differences in the indexing parameters should reflect the differences observed in the two samples.

In this section, we shall be concerned chiefly with two of the main numerical summaries of population probability models, the population analogues of the sample mean and sample variance (or sample standard deviation). Other important concepts, such as the population median and other quantiles, will be discussed in Section 3.5.

Let us begin with a look at how the notion of the sample mean can be developed to help us define its population counterpart, the population mean.

### 3.1.1 The population mean: discrete random variables

**Example 3.2 Rolls of a fair die**

The outcomes of 30 rolls of a fair die are given in Table 3.2.

<table>
<thead>
<tr>
<th>Frequency Table for the 30 rolls of the die</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome ((j))</strong></td>
</tr>
<tr>
<td><strong>Frequency ((f_j))</strong></td>
</tr>
</tbody>
</table>

One way of calculating the sample mean for the 30 rolls (if you had to do this by hand) is simply to obtain the sample total and divide by the sample size:

\[
\bar{x} = \frac{4 + 3 + 2 + \cdots + 2 + 5}{30} = \frac{96}{30} = 3.2.
\]

Alternatively, one could start by summarizing the original data in the form of a frequency table. This is shown in Table 3.3.

There are five 1s (a total of \(1 \times 5 = 5\)), seven 2s (a total of \(2 \times 7 = 14\)), and so on. Then the sample mean could be computed as

\[
\bar{x} = \frac{(1 \times 5) + (2 \times 7) + (3 \times 6) + (4 \times 4) + (5 \times 5) + (6 \times 3)}{30} = \frac{96}{30} = 3.2,
\]

achieving the same result but by a different method.

This example shows that, given a sample of data from a discrete distribution, there are two equivalent ways of calculating the sample mean. If \( x_1, x_2, \cdots, x_{30} \)
denote the values in the sample in Example 3.2, in the first method we obtain the sample mean by adding all the values together and dividing by 30. In this case we are using the familiar definition of the sample mean,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \quad (3.1)$$

The idea behind the second approach is to count how many of each particular outcome there are in the sample in order to obtain a frequency table. If we denote the number of occurrences of outcome $j$ in the sample by $f_j$ (for instance, $f_2 = 7$, in Example 3.2), the contribution made to the total by each of the outcomes $j$ is $j \times f_j$ (for example, $2 \times 7 = 14$). Adding all these up and then dividing by $n$ (the sample size, equal to 30 in Example 3.2), we obtain the sample mean in the form

$$\bar{x} = \frac{1}{n} \sum_{j} jf_j. \quad (3.2)$$

Here, the sum is over all possible outcomes $j$ (1, 2, ..., 6, for the die). The two formulas (3.1) and (3.2) for $\bar{x}$ give the same answer.

However, in Chapter 2, Subsection 2.1.2, you saw how a discrete probability $p(j)$ could be defined by taking the limiting value of the sample relative frequency of the observation $j$ as the sample size $n$ gets larger and larger. By writing $\bar{x}$ in the form (3.2), or perhaps even more clearly in the form

$$\bar{x} = \sum_{j} j \frac{f_j}{n},$$

the same idea can be employed here to produce a definition of the population mean, at least for a discrete random variable. By simply replacing the sample relative frequency $f_j/n$ by its limiting value, the probability $p(j)$, we obtain the following definition.

For a discrete random variable taking the value $j$ with probability $p(j)$, the **population mean** is given by

$$\mu = \sum_{j} jp(j),$$

where the sum is over the set of possible observed values, that is, over the range of the random variable.

**Example 3.3 The average score when a fair die is rolled**

For a die that is assumed to be fair, each of the six possible outcomes $j = 1, 2, ..., 6$ occurs with probability $p(j) = \frac{1}{6}$. The mean outcome in a single roll of a fair die is therefore

$$\mu = \sum_{j} jp(j)$$

$$= (1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) + (3 \times \frac{1}{6}) + (4 \times \frac{1}{6}) + (5 \times \frac{1}{6}) + (6 \times \frac{1}{6})$$

$$= \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5. \quad \blacksquare$$

We shall discuss the case of continuous distributions in Subsection 3.1.2.

The lower-case Greek letter $\mu$ is often used to denote a population mean.
Notice that it is not necessary for the mean of a random variable \( X \) to be in the range of \( X \): 3.5 is not a possible outcome when a die is rolled. As another example, the ‘average family size’ in Britain is not an integer.

**Exercise 3.1**

In Chapter 2, you were also introduced to certain unfair dice called *Tops* and you saw that, for a *Double-Five*, the probability distribution for the outcome of a single roll is as given in Table 3.4.

<table>
<thead>
<tr>
<th>( j )</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(j) )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{6} )</td>
</tr>
</tbody>
</table>

What is the mean of this distribution? How does it compare with the mean outcome for a fair die?

One very important application of statistics is to epidemiology—the study of health and illness in human populations. Many different models have been developed for the transmission of infectious diseases, a few of them simple, but most of them rather complicated. In small communities (for instance, families and schools) one variable of interest is the total number of people who catch a disease, given that initially one member of the community becomes infected. In a family of 4, say, that number could be 1, 2, 3 or 4. This number is a random variable because epidemic dynamics are, to a great extent, a matter of chance—whether or not you catch your brother’s cold, for instance, is not a predetermined event.

**Exercise 3.2**

One model for a certain disease within a particular family of 6 gave the probability distribution for the number \( X \), who eventually suffer from the disease, as shown in Table 3.5.

<table>
<thead>
<tr>
<th>( x )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x) )</td>
<td>( \frac{3}{60} )</td>
<td>( \frac{8}{60} )</td>
<td>( \frac{15}{60} )</td>
<td>( \frac{20}{60} )</td>
<td>( \frac{24}{60} )</td>
<td>( \frac{20}{60} )</td>
</tr>
</tbody>
</table>

What is the mean of the distribution of \( X \)?

An alternative terminology for the mean of a random variable \( X \) is the expected value of \( X \) or simply the expectation of \( X \). This is written \( E(X) \) as an alternative to \( \mu \).

We can summarize the foregoing as follows.

**The mean of a discrete random variable**

For a discrete random variable \( X \) with probability mass function

\[
P(X = x) = p(x)
\]

over a specified range, the mean of \( X \) or the expected value of \( X \) or the expectation of \( X \) is given by

\[
\mu = E(X) = \sum_{x} xp(x),
\]

where the sum is taken over the range of \( X \).
The notation $\mu$ for the mean of a random variable $X$ is sometimes modified to include the subscript $X$, that is, $\mu_X$. This notation is particularly useful where a model involves more than one variable, as the means of different random variables cannot then be confused. However, in this course the subscript will not usually be included except where it is necessary to avoid ambiguity.

**Example 3.4 Using the $E(\cdot)$ notation**

If $X$ denotes the score on a single roll of a fair die (Example 3.3), then the expected value of $X$ is

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5.$$

On the other hand, if the random variable $Y$ denotes the score on a single roll of a Double-Five (Exercise 3.1), then the expected value of $Y$ is

$$E(Y) = 4.$$

For the family of 6 (Exercise 3.2), the expected number $E(Z)$ of family members who eventually suffer from the disease is

$$E(Z) = 4.3.$$

Occasionally, the phrase 'expected value' or 'expected number' is more natural than 'mean'. But notice that 'the value you would expect to get' is not usually a valid interpretation: you could not actually have 4.3 ill people in a family, because the number 4.3 is not an integer.

The examples above are for probability distributions where the probability mass function could be specified exactly. Earlier, the idea of indexing distributions by means of some unspecified quantity or quantities called the parameter(s) of the model was introduced. It will be particularly useful if some simple link can be established between the indexing parameter(s) and some summary measure, such as the mean: this would aid the interpretation of model parameters.

Let us take a look, then, at the first 'parametric family' of discrete distributions to which you were introduced in Chapter 2, Section 2.3. This was the Bernoulli family of models: each distribution in this family allows only the two possible outcomes 0 or 1, and the probability mass function of a Bernoulli distribution is $p(1) = p$, $p(0) = 1 - p$. (Here, $p$ is the indexing parameter.) The population mean of any Bernoulli distribution can therefore be found in terms of $p$. In fact,

$$\mu = \sum_{j=0}^{1} j p(j)$$

$$= 0 \times p(0) + 1 \times p(1) = 0 \times (1 - p) + 1 \times p = p.$$

That is, the population mean of a Bernoulli distribution is the parameter $p$ used to index this family of models.

**Exercise 3.3**

(a) What is the mean score resulting from a toss of a fair coin, if we score 1 for Heads and 0 for Tails?
(b) Suppose that a random variable is defined to take the value 1 when a fair die shows a 3 or a 6, and 0 otherwise. What is the mean value for this Bernoulli distribution?

The second discrete probability distribution introduced in Chapter 2 was the binomial distribution or family. If the random variable $X$ has a binomial distribution $B(n, p)$ then the probability mass function of $X$ is

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$ 

So, using the definition in (3.3), the mean of $X$ (or the expected value of $X$) is given by the sum

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1 - p)^{n-x}.$$ 

In fact, this calculation is not quite as unwieldy as it looks, and a small amount of algebraic manipulation would give us the answer we need. However, no algebra is necessary if we think about what the binomial random variable $X$ represents. It is the number of successes in a series of $n$ trials, where the probability of a successful outcome at each trial is $p$. So the problem posed is this: what is the expected number of successes in such a series of $n$ trials? No formal mathematics is required to provide an answer to this question (though it could be used to provide a formal proof). If, for instance, 100 trials are performed and for each the probability of success is $\frac{1}{4}$, then the expected number of successes is $100 \times \frac{1}{4} = 25$ and the expected number of failures is $100 \times \frac{3}{4} = 75$. In general, the mean of a binomial random variable $X$ indexed by the two parameters $n$ and $p$ is given by the product

$$E(X) = np. \quad (3.5)$$

Notice, as in the case of a fair die, that the mean may be some value not in the range of $X$ (for instance, take $n = 100$ and $p = \frac{1}{3}$; the number $np = 33\frac{1}{3}$ is not an integer). Nevertheless, the mean or expectation is a statement about the ‘long-term’ average number of successes in sequences of Bernoulli trials.

**Example 3.5 The mean of a binomial random variable: two methods of calculation**

If $X$ is binomial $B(4, 0.4)$, then its probability mass function is given by

$$p(x) = \binom{4}{x} (0.4)^x (0.6)^{4-x}, \quad x = 0, 1, 2, 3, 4.$$ 

The individual probabilities are as follows.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1296</td>
</tr>
<tr>
<td>1</td>
<td>0.3456</td>
</tr>
<tr>
<td>2</td>
<td>0.3456</td>
</tr>
<tr>
<td>3</td>
<td>0.1536</td>
</tr>
<tr>
<td>4</td>
<td>0.0256</td>
</tr>
</tbody>
</table>

The mean of $X$ is then

$$E(X) = \sum_{x=0}^{4} xp(x)$$

$$= (0 \times 0.1296) + (1 \times 0.3456) + \cdots + (4 \times 0.0256)$$

$$= 0 + 0.3456 + 0.6912 + 0.4608 + 0.1024 = 1.6.$$
This is much more easily obtained using (3.5):
\[ E(X) = np = 4 \times 0.4 = 1.6. \]
This may be shown graphically as in Figure 3.3. The point \( x = 1.6 \) (the mean of the distribution) is shown as an arrowhead on a sketch of the probability mass function of \( X \).

![Figure 3.3 The mean of \( X \) when \( X \sim B(4,0.4) \)](image)

**Example 3.6 Spores of the fungus Sordaria**

The spores of the fungus *Sordaria* are produced in chains of eight. Any chain may break (at any of the seven joints) and the spores are thus projected in chainlets varying in length from one to eight. It turns out that it is reasonable to model the breakages occurring at the joints as independent; and each joint has the same probability \( p \) of breaking.

So, for instance, an original chain of eight might survive unbroken with probability \((1 - p)^7\). However, it might break at all seven joints, so that eight chainlets all of length one are projected. This occurs with probability \( p^7 \). Chainlets of length one are called 'singletons'.

The number of singletons, \( X \), produced by a chain of eight spores is a random variable taking values \( 0, 1, 2, \ldots, 8 \). (Actually, a count of seven is not possible, for the eighth spore would itself necessarily be a singleton too.) The probability distribution of \( X \) is given in Table 3.6. (The distribution derives directly from a complete enumeration of possible cases; you should not bother to check these results.)

In this case there is no particularly obvious way of deducing the average number of singletons produced, other than by applying the formula. This gives

\[
E(X) = \sum_{x=0}^{8} xp(x) \\
= 0 \times (1 - p)^4(1 + 2p - p^2 - p^3) + \cdots + 8 \times p^7 \\
= 2p(3p + 1),
\]

after considerable simplification. For instance, if \( p \) is 0.8 (the original chain is very fragile) the expected number of resulting singletons is 5.44. If \( p \) is as low as 0.1 (the original chain is robust) the expected number of resulting singletons is only 0.26.

The mean has the following physical interpretation (rather idealized): imagine lead weights of mass 0.1296, 0.3456, 0.3456, 0.1536 and 0.0256 units placed at equal intervals on a thin plank of zero mass. If it is represented by the horizontal axis in Figure 3.3, the plank will balance at the point indicated by the arrowhead—at a point just to the right of the midpoint between the two largest weights.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>((1 - p)^4(1 + 2p - p^2 - p^3))</td>
</tr>
<tr>
<td>1</td>
<td>(2p(1 - p)^3(1 + 4p + p^2))</td>
</tr>
<tr>
<td>2</td>
<td>(p^5(1 - p)^3(3 + 12p - 5p^2))</td>
</tr>
<tr>
<td>3</td>
<td>(4p^3(1 - p)^3(1 + 4p))</td>
</tr>
<tr>
<td>4</td>
<td>(5p^4(1 - p)^3(1 + 2p))</td>
</tr>
<tr>
<td>5</td>
<td>(6(1 - p)^3p^3)</td>
</tr>
<tr>
<td>6</td>
<td>(7(1 - p)p^5)</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>(p^7)</td>
</tr>
</tbody>
</table>

The probability of a conjunction of independent events is found from the product of the probabilities of the component events. This result was used in our derivation of the binomial probability distribution in Chapter 2, Subsection 2.3.2.
However, in this case there is no immediate interpretation of the mean in terms of the indexing parameter \( p \).

These arithmetic calculations are illustrated in Figure 3.4. Notice the interesting multimodal nature of the probability distribution of the number of singletons when \( p \) is 0.8.

![Figure 3.4](image)

**Figure 3.4** The probability distribution of the number of singletons (a) for \( p = 0.8 \); (b) for \( p = 0.1 \).

**Exercise 3.4**

A chainlet of length four is called a ‘quad’. The original chain will result in two quads being projected, for instance, only if the middle joint breaks and the other six do not: the probability of this is \( p(1-p)^6 \). The probability distribution of the number of quads projected is given in Table 3.7. Again, do not worry about the algebraic details.

<table>
<thead>
<tr>
<th>( Y )</th>
<th>( p(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( 1 - p(1-p)^2 (3p^3 - 2p^3 + 11p^2 - 4p + 1) )</td>
</tr>
<tr>
<td>1</td>
<td>( p^3(1-p)^2 (2p^2 + 2p + 5) )</td>
</tr>
<tr>
<td>2</td>
<td>( p(1-p)^6 )</td>
</tr>
</tbody>
</table>

Find the expected value of \( Y \) when

(a) \( p = 0.1 \);   (b) \( p = 0.4 \);   (c) \( p = 0.6 \);   (d) \( p = 0.8 \).

**3.1.2 The population mean: continuous random variables**

The variation that might be observed in measurements on a discrete random variable is expressed through its probability mass function, and you have seen how to use the p.m.f. to calculate the mean or expected value of a discrete random variable. Similarly, variation observed in measurements on a continu-
ous random variable may be expressed by writing down a probability density function. If this density function is a good model then we should be able to use it not just for forecasts about the likelihood of different future measurements but, as in the case of discrete random variables, to provide information about the long-term average in repeated measurements.

As in the case of the mean of a discrete random variable, the value of the expectation $\mu = E(X)$ of a continuous random variable $X$ has a physical interpretation. It is the point about which a physical model of the density would balance, if such a model were constructed (say, of tin plate). Here are some examples.

**Example 3.7 Means of continuous random variables**

(a) In Chapter 2, Example 2.8, the triangular density $\text{Triangular}(20)$ was used as a model for the waiting time (in seconds) between vehicles, for traffic using the Kwinana Freeway in Perth, Western Australia. The density is shown in Figure 3.5. Also shown is the point at which a tin triangle with these dimensions would balance. This point occurs one-third of the way along the base of the triangle, at the point $t = 6 \frac{2}{3}$. This is the mean of the triangular distribution with parameter 20.

![Figure 3.5](image)

**Figure 3.5** The triangular density $\text{Triangular}(20)$, showing the mean $\mu = 6 \frac{2}{3}$

(b) The density of the random variable $X \sim N(\mu, \sigma^2)$ is shown in Figure 3.6. (Observations on $X$ much below $\mu - 3\sigma$ or much above $\mu + 3\sigma$ are possible, but unlikely.) The normal density is symmetric: if it were possible to construct a tin-plate model of the density, the model would balance at the point $x = \mu$.

![Figure 3.6](image)

**Figure 3.6** The p.d.f. of the normal distribution $N(\mu, \sigma^2)$
So, in fact, one of the two indexing parameters for the normal family is the mean of the distribution. Observations are no more likely to be above the central value \( \mu \) than below it. The long-term observed average is bound to be \( \mu \) itself.

In the following exercise you are asked to simulate random observations from the triangular family, and use your findings to suggest a general formula for the mean of the triangular distribution \( \text{Triangular}(\theta) \).

**Exercise 3.5**

(a) Generate a random sample of size 10 from the triangular distribution with parameter \( \theta = 20 \). List the elements of your sample and calculate the sample mean.

(b) Find the mean of a random sample of size 1000 from the triangular distribution with parameter \( \theta = 20 \).

Repeat the sampling procedure a further nine times and list the ten sample means you obtained. (You should find that your list offers supporting evidence for the result stated in Example 3.7(a): that the mean of the triangular distribution \( \text{Triangular}(20) \) is \( 6 \frac{2}{3} \).

(c) Find the mean of random samples of size 1000 from triangular populations with parameters (i) \( \theta = 30 \); (ii) \( \theta = 300 \); (iii) \( \theta = 600 \).

(d) Use your results in this exercise to hazard a guess at the mean of the continuous random variable \( T \) when \( T \sim \text{Triangular}(\theta) \).

Most probability density functions possess neither the symmetry of the normal density function nor the rather convenient geometrical form of the triangular density function. How, in general, is the mean of a continuous random variable calculated?

The result (3.3) tells us that for a discrete random variable \( X \) with probability mass function \( p(x) \), the mean of \( X \) is given by the formula

\[
\mu = E(X) = \sum_X x p(x),
\]

where the summation is taken over the range of \( X \). This represents an average of the different values that \( X \) may take, according to their chance of occurrence. The definition of the mean of a continuous random variable is analogous to that of a discrete random variable.

**The mean of a continuous random variable**

For a continuous random variable \( X \) with probability density function \( f(x) \) over a specified range, the mean of \( X \) or the expected value of \( X \) is given by

\[
\mu = E(X) = \int_X x f(x) \, dx,
\]

where the integral is taken over the range of \( X \).
As you can see, the technique of integration is required for the calculation of the mean of a continuous random variable. If you are familiar with the technique, you might like to confirm for yourself some of the standard results that follow now and in the rest of the course. However, these are standard results and you would not be expected to obtain them from first principles.

**Example 3.8 The mean of the triangular distribution**

(a) You saw in Chapter 2, Subsection 2.2.2, that the p.d.f. of the continuous random variable \( T \sim \text{Triangular}(20) \) is given by

\[
  f(t) = \frac{20 - t}{200}, \quad 0 \leq t \leq 20.
\]

Using (3.6), the mean of \( T \) is

\[
  \mu = E(T) = \int_T tf(t) \, dt
  = \int_0^{20} t \left( \frac{20 - t}{200} \right) \, dt
  = \frac{1}{200} \int_0^{20} t(20 - t) \, dt
  = \frac{1}{200} \int_0^{20} (20t - t^2) \, dt
  = \frac{1}{200} \left[ 10t^2 - \frac{1}{3} t^3 \right]_0^{20}
  = \frac{1}{200} \left( 4000 - \frac{8000}{3} \right)
  = \frac{20}{3}
  = 6\frac{2}{3}.
\]

(b) In Exercise 3.5(d) you guessed the mean of the triangular distribution \( \text{Triangular}(0) \). Now your guess can be confirmed (or otherwise). The p.d.f. of the continuous random variable \( T \sim \text{Triangular}(\theta) \) is given by

\[
  f(t) = \frac{2(\theta - t)}{\theta^2}, \quad 0 \leq t \leq \theta.
\]

So the mean of the distribution is given by

\[
  \mu = E(T) = \int_T tf(t) \, dt
  = \int_0^\theta t \left( \frac{2(\theta - t)}{\theta^2} \right) \, dt
  = \frac{2}{\theta^2} \int_0^\theta (\theta t - t^2) \, dt
  = \frac{2}{\theta^2} \left[ \frac{1}{2} \theta t^2 - \frac{1}{3} t^3 \right]_0^\theta
  = \frac{2}{\theta^2} \left[ \frac{1}{2} \theta^3 - \frac{1}{3} \theta^3 \right]
  = \frac{2}{\theta^2} \left[ \frac{1}{6} \theta^3 \right]
  = \frac{1}{3} \theta.
\]

The c.d.f. of \( T \),

\[
  F(t) = 1 - \left( 1 - \frac{t}{\theta} \right)^2, \quad 0 \leq t \leq \theta,
\]

was given in Chapter 2, Section 2.4. Just as a c.d.f. is obtained from a p.d.f. by integration, so a p.d.f. is obtained from a c.d.f. by the inverse operation, known as differentiation. If you are familiar with this technique, you can check that the triangular p.d.f. is as stated. However, it is used here only as a means to an end: to find the mean of the triangular distribution.
Example 3.9 The mean of the normal distribution

Were you to calculate the mean of the normal distribution $N(\mu, \sigma^2)$ directly from (3.6),

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right) dx,$$

you would find that the parameter $\sigma$ vanished and you were left with $E(X) = \mu$. ■

Now try the following exercise. (Notice that the two results you need to answer the question are given to you: you do not need to obtain them from first principles.)

Exercise 3.6

For the triangular distribution with parameter $\theta$, the c.d.f. is given by

$$F(t) = 1 - \left( 1 - \frac{t}{\theta} \right)^2, \quad 0 \leq t \leq \theta,$$

and the mean of the distribution is $\mu = \frac{1}{3}\theta$.

It is assumed in a traffic monitoring context that the triangular density provides an adequate model for the waiting time between successive vehicles, but the value of the parameter $\theta$ is unknown. (It could be one of the aims of a sampling experiment to estimate the value of $\theta$: see Chapter 6 for more on the topic of estimation.)

What proportion of waiting times is longer than average?

3.1.3 The population variance

In Subsection 3.1.2 the idea of a sample mean was extended to the mean of a theoretical model for the observed variation, which was denoted by $\mu$ or $E(X)$ (the expected value of $X$). Now we require a measure of dispersion for a population, analogous to the sample variance or sample standard deviation. The following example illustrates a typical context in which knowledge of a population variance is essential to answering a scientific question.

Example 3.10 Measuring intelligence

A psychologist assessing intellectual ability decides to use the revised Wechsler Adult Intelligence Scale—WAIS-R—to measure IQ. She finds that one subject has a score of 110. This is above the population mean of 100. But how far above the mean is it? Should she expect many people to score as high as this, or is the difference of 10 points a large difference? To answer this question she needs to know something about the spread or dispersion of IQ scores in the population, and she might, for example, choose to measure this spread using the population analogue of the sample standard deviation encountered in Chapter 1. ■
As defined in Chapter 1, the sample variance is given by

\[ s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2. \]

This measure of dispersion gives the *average squared deviation* of each item in the sample from the sample mean, with the small distinction that the average is obtained through division by \( n - 1 \) rather than by \( n \). The analogous measure for a probability model is the *expected squared deviation* of a random variable \( X \) from the mean of \( X \). This may be written using the ‘expectation’ notation \( E(\cdot) \) as

\[ E[(X - \mu)^2]. \]

So now we require not simply ‘the expected value of \( X \)’, but ‘the expected value of a function of \( X \)’. We need to calculate the value of that function \((x - \mu)^2\) for each value of \( x \) in the range of \( X \), and then average the squared deviations obtained over the probability distribution of \( X \).

**Example 3.2 continued**

In Example 3.2 we looked at the results of 30 rolls of a fair die. The sample mean was found to be \( \bar{x} = 3.2 \). For this sample, the sum of squared deviations from the mean is given by

\[
\sum_{i=1}^{30} (x_i - \bar{x})^2
= (1 - 3.2)^2 + (1 - 3.2)^2 + (1 - 3.2)^2 + (1 - 3.2)^2 + (1 - 3.2)^2
+ (2 - 3.2)^2 + (2 - 3.2)^2 + (2 - 3.2)^2 + (2 - 3.2)^2 + (2 - 3.2)^2
+ \ldots
+ (6 - 3.2)^2 + (6 - 3.2)^2 + (6 - 3.2)^2.
\]

This can more conveniently be written as

\[
\sum_{x=1}^{6} (x - \bar{x})^2 f_x
= (1 - 3.2)^2 \times 5 + (2 - 3.2)^2 \times 7 + (3 - 3.2)^2 \times 6 + \cdots + (6 - 3.2)^2 \times 3
= 76.8;
\]

and so the sample variance (dividing by \( n - 1 = 29 \)) is

\[ s^2 = \frac{76.8}{29} = 2.65. \]

However, a *theoretical probability model* for the outcome of rolls of a fair die is provided by the random variable \( X \) with probability mass function

\[ P(X = x) = p(x) = \frac{1}{6}, \quad x = 1, 2, \ldots, 6. \]

The mean of \( X \) is given by

\[ \mu = E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5. \]
The expected value of \((X - \mu)^2\) is found by averaging the values obtained for \((x - \mu)^2\) over the probability distribution of \(X\):

\[
E[(X - \mu)^2] = (1 - 3.5)^2 \times \frac{1}{6} + (2 - 3.5)^2 \times \frac{1}{6} + (3 - 3.5)^2 \times \frac{1}{6} + \cdots + (6 - 3.5)^2 \times \frac{1}{6}
\]

\[
= \frac{6.25}{6} + \frac{2.25}{6} + \frac{0.25}{6} + \cdots + \frac{6.25}{6}
\]

\[
= 2.92.
\]

So we see that our sample of 30 rolls was, on that occasion, a little less variable than theory would have suggested.

For discrete probability distributions, then, the variance is given by (3.7).

**The variance of a discrete random variable**

For a discrete random variable \(X\) with probability mass function

\[
P(X = x) = p(x)
\]

over a specified range, with mean \(\mu = E(X)\), the variance of \(X\) is given by

\[
\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 p(x),
\]

where the sum is taken over the range of \(X\).

The standard deviation of \(X\) is given by the square root of the variance:

\[
SD(X) = \sqrt{V(X)} = \sigma = \sqrt{\sum_x (x - \mu)^2 p(x)}.
\]

For instance, the standard deviation in the outcome of rolls of a fair die is \(\sqrt{2.92} \approx 1.71\).

**Exercise 3.7**

What is the variance of the outcome of throwing a Double-Five? How does this compare with the variance for a fair die?

We can work out the variance of any Bernoulli distribution with parameter \(p\). Recall that the possible outcomes are just 0 and 1; so, from the definition,

\[
\sigma^2 = (0 - \mu)^2 p(0) + (1 - \mu)^2 p(1).
\]
Since \( p(0) = 1 - p \), \( p(1) = p \) and, from (3.4), the mean is \( p \), it follows that

\[
\sigma^2 = (-p)^2(1 - p) + (1 - p)^2p = p(1 - p)(p + 1 - p) = p(1 - p).
\]

Calculation of the variance of the binomial distribution \( B(n, p) \) will be postponed to Chapter 4, Section 4.3.

We can move from the case of discrete distributions to that of continuous distributions just as we did for means. In short, we replace the p.m.f. by a p.d.f., and the sum by an integral.

---

**The variance of a continuous random variable**

For a continuous random variable \( X \) with probability density function \( f(x) \) over a specified range and with mean \( \mu = E(X) \), the variance of \( X \) is given by

\[
\sigma^2 = V(X) = E[(X - \mu)^2] = \int_X (x - \mu)^2 f(x) \, dx, \tag{3.8}
\]

where the integral is taken over the range of \( X \).

The formula for the variance of a random variable will shortly be rewritten in a way that leads to some easement of the algebra. However, for common models it is not usually necessary to go through the algebra at all—the results are standard and well-known. At this stage, it is useful to notice that regardless of whether the random variable \( X \) is discrete or continuous, its variance can be written as

\[
V(X) = E[(X - \mu)^2]. \tag{3.9}
\]

The normal probability distribution is discussed in more detail in Chapter 5. However, just as the parameter \( \mu \) indexing the normal distribution is the mean of the normal distribution, so the second parameter \( \sigma^2 \) indexing the normal distribution is the variance of the distribution (and \( \sigma \) is its standard deviation).

Now recall the question posed in Example 3.10: is a score of 110 on an IQ test, where the population average score is 100, unusually high? If the underlying probability distribution of WAIS-R scores were known, this question could be answered.

In fact, the IQ scale is designed to take account of the variability in responses from within the population in such a way that the resulting scores are normally distributed with mean \( \mu = 100 \) and standard deviation \( \sigma = 15 \) (variance 225).
We can plot the corresponding normal density and identify the location of the particular score of interest, 110. This is shown in Figure 3.7.

\[
\begin{align*}
\text{Figure 3.7} & \quad \text{An observed score of 110 when } X \sim N(100, 225) \\
\end{align*}
\]

The observed score is higher than average; nevertheless it looks as though a considerable minority of the population (the proportion represented by the shaded area in the diagram) would score as well or better. You will see in Chapter 5 how to compute this proportion. It is given by the following integral:

\[
P(X \geq 110) = \int_{110}^{\infty} \frac{1}{15\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - 100}{15} \right)^2 \right] dx.
\]

In practice, evaluation of the integral is unnecessary since, for the normal distribution, such probabilities can be obtained using tables or a computer package. This is discussed further in Chapter 5. (The area of the shaded region in Figure 3.7 is 0.252: more than a quarter of the population would score 110 or more on the WAIS-R intelligence test.)

### 3.2 Independence of random variables

In Chapter 2, Subsection 2.3.1, the Bernoulli distribution was introduced as a model for a particular class of random variable. In Chapter 2, Subsection 2.3.2, we proceeded to the binomial distribution which is a model for an extended class of random variable obtained by adding together a number of separate Bernoulli random variables. However, the components of the sum all had to involve the same value of the parameter \( p \), and also the different Bernoulli random variables had to be independent. In such a case, the components of the sum are described as independent identically distributed (i.i.d.) random variables. This notion of independence recurs through so much of statistics that the whole of this fairly short section will be devoted to it. The idea is a particularly important one for Chapters 6 and 11.

If the random variable \( X \) represents the number showing on a rolled red die, and \( Y \) represents the number showing on a rolled white die, the value \( x \) taken by \( X \) has no bearing on the value \( y \) taken by \( Y \), and vice versa. Likewise, knowledge of the number of newspapers sold in a particular newsagent’s shop in Wrexham one morning will not tell us much about the number of babies born in a maternity hospital in Luton that day, and vice versa.
However, Seal (1964) describes an experiment in which frogs' skull lengths $X$ and skull breadths $Y$ were measured. One might reasonably expect a large reading on one variable to indicate a similarly large reading on the other, though there will still be random variation evident in both variables. A sample of paired measurements will yield information on the nature and degree of the association between the two variables.

There is a clear distinction between the first two of these three situations and the third. In the first two, the random variables have no bearing on each other; conversely, in the frogs' skulls example, they are, in some way, related to one another. The distinction is to do with the degree of dependence between them. In this section, we are concerned with random variables of the former type that do not have any effect on each other. Dependent random variables, like skull length and skull breadth, will be dealt with in later chapters (particularly in Chapter 11).

A particular property of independent random variables was used in the derivation of the binomial probability distribution in Chapter 2, Section 2.3. There, each component in a sequence of Bernoulli trials is assumed to be independent of every other component. We used the result that the probability of obtaining a particular sequence of responses is found by multiplying together the individual probabilities for each response. Thus, for instance, if $p$ is the probability of a Yes response, then

$$P(\text{Yes Yes No}) = P(\text{Yes})P(\text{Yes})P(\text{No}) = p^2(1 - p)$$

and

$$P(\text{No No No}) = (1 - p)^3.$$

If the assumption of independence breaks down, then this approach fails too. In an extreme case, the second and third respondents might simply repeat the previous response, so that

$$P(\text{Yes Yes Yes}) = p$$

$$P(\text{No No No}) = 1 - p,$$

and these are the only two possible sequences of responses. Here, the degree of dependence between the responses is very high.

The idea that the probability of a conjunction of independent events is found by multiplying together the individual probabilities for each component event may be extended naturally to independent random variables in the following way.

**Independence of random variables**

If $X$ and $Y$ are discrete random variables that are independent, then the probability that $X$ takes the value $x$ and $Y$ takes the value $y$, simultaneously, is given by

$$P(X = x, Y = y) = P(X = x)P(Y = y), \quad (3.10)$$

for all $x$ in the range of $X$ and all $y$ in the range of $Y$. The comma on the left-hand side of the expression may be read as 'and'.
Example 3.11 Scores on two different dice

If \( X \) is the score when a fair die is rolled and \( Y \) the score when a Double-Five is rolled, and you throw one after the other, the outcomes are clearly independent (for by what mechanical influence could the scores be related?). The probability of scoring two 5s is

\[
P(X = 5, Y = 5) = P(X = 5)P(Y = 5) = \frac{1}{6} \times \frac{2}{6} = \frac{1}{18}.
\]

Example 3.12 Leaves of Indian creeper plants

The leaves of Indian creeper plants \( Pharbitis nil \) can be variegated or unvariegated and, at the same time, faded or unfaded. Are the two characteristics independent? In one experiment, plants of a particular type were crossed. They were such that the offspring plants would have leaves that were variegated with probability \( \frac{3}{4} \) (and unvariegated with probability \( \frac{1}{4} \)); and also the leaves would be faded with probability \( \frac{1}{4} \) (and unfaded with probability \( \frac{3}{4} \)). If the theory of independence is correct, then one would expect to observe unvariegated unfaded leaves in \( \frac{3}{4} \times \frac{1}{4} = \frac{3}{16} \) of the offspring plants, unvariegated faded leaves in \( \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \) of them, and so on.

As it turned out, of 290 offspring plants observed, 187 had unvariegated unfaded leaves, 35 had unvariegated faded leaves, 37 had variegated unfaded leaves and 31 had variegated faded leaves.

The observed sample relative frequencies occurred in the ratios

\[
\frac{187}{290} : \frac{35}{290} : \frac{37}{290} : \frac{31}{290} = 0.64 : 0.12 : 0.13 : 0.11.
\]

These differ somewhat from the forecast proportions of

\[
\frac{9}{16} : \frac{3}{16} : \frac{1}{16} = 0.56 : 0.19 : 0.19 : 0.06.
\]

In fact, even allowing for random variation in the observed experimental results, the theory of independence is resoundingly rejected on the basis of the results.

When discussing two or more random variables some reduction in notation can be usefully achieved by using the notation \( p(x) \) for a probability mass function, and distinguishing between p.m.f.s by including the names of the random variables as subscripts. Thus we write

\[
P(X = x) = p_X(x),
\]

\[
P(Y = y) = p_Y(y).
\]

In the same way, we write

\[
P(X = x, Y = y) = p_{X,Y}(x, y).
\]

The function \( p_{X,Y}(x, y) \) is called the joint probability mass function for the random variables \( X \) and \( Y \). Then condition (3.10) for independence between random variables \( X \) and \( Y \) can be rewritten as

\[
p_{X,Y}(x, y) = p_X(x)p_Y(y), \tag{3.11}
\]

for all \( x \) in the range of \( X \) and all \( y \) in the range of \( Y \).
Exercise 3.8

The entries in Table 3.8 are values of the joint probability mass function \( p_{X,Y}(x,y) \) of random variables \( X \) and \( Y \). The random variable \( X \) takes the values 0, 1 and 2, and \( Y \) takes the values -1 or 1. The columns of the table correspond to the random variable \( X \) with p.m.f. \( p_X(0) = 0.4 \), \( p_X(1) = 0.4 \) and \( p_X(2) = 0.2 \). The rows are associated with the random variable \( Y \) with p.m.f. \( p_Y(-1) = 0.3 \) and \( p_Y(1) = 0.7 \). Notice that the separate probabilities for \( X \) are found by adding within columns, and the separate probabilities for \( Y \) by adding within rows.

Are \( X \) and \( Y \) independent random variables?

<table>
<thead>
<tr>
<th>( y )</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.12</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>0.28</td>
<td>0.30</td>
<td>0.12</td>
</tr>
</tbody>
</table>

The subject of dependent random variables is discussed in detail in Chapter 11.

3.3 The geometric probability model

In Chapter 2, Subsection 2.1.2, you learned about the Bernoulli trial, a statistical experiment where exactly one of only two possible outcomes occurs. The main features of the model were described in Chapter 2 and are as follows. The outcomes are usually something like Success–Failure, Yes–No, On–Off, Male–Female. In order to standardize the associated Bernoulli distribution
so that it becomes a useful statistical model with wide application, it is usual to identify one of the outcomes with the number 1 and the other with the number 0. Then, for example, a sequence of twenty engineering trials run on identical pieces of equipment as part of a quality control exercise, resulting in the outcomes Pass, Pass, Fail, \ldots, Fail, Fail could be written more easily as

\[ X_1 = 1, \quad X_2 = 1, \quad X_3 = 0, \quad \ldots, \quad X_{19} = 0, \quad X_{20} = 0. \]

In this case the number 1 was used as a label to indicate that a piece of equipment had passed the quality test. The numbers could easily have been arranged the other way round: a score of 1 for a piece of defective equipment.

The Bernoulli distribution has associated with it a single parameter, the number \( p \). This number is a probability (so \( 0 < p < 1 \)): it represents the probability that any single trial results in the outcome 1.

In this section we shall consider two particular examples based on the Bernoulli model.

**Example 3.13 The sex of consecutive children**

The pattern of boys and girls in a family is one that has received a lot of scientific attention, both biological and statistical, for nearly three centuries. In 1710 the philosopher John Arbuthnot, having examined parish records and noted that for 82 consecutive years more boys than girls had been christened and (reasonably) deducing that for 82 years more boys than girls had been born, proposed that sex determination was not a simple matter of chance, akin to the result of the toss of a fair coin. (Actually, he proposed a degree of divine intervention: that while the coin was in mid-air, God temporarily suspended the laws of chance.) The philosopher and probabilist Nicholas Bernoulli (1687–1759) had also noted the imbalance in the sexes, and commented that sex determination was like rolling a 35-sided die, with 18 faces marked 'boy' and 17 'girl'.

The biologists, statisticians, genealogists, philosophers, social scientists and demographers who have all at one time or another made their various investigations in this area, would have found their research greatly hampered had they not developed various models (and an easy notation) for the ideas and theories involved. The easiest model is to assume independence from child to child, write 1 for a son and 0 for a daughter, use Bernoulli's estimate for the probability of a boy of \( p = 18/35 = 0.514 \) and write

\[ P(X = 0) = 0.486, \quad P(X = 1) = 0.514, \]

where \( X \) is the random variable denoting the sex of a child.

Actually, nearly all the very considerable mass of data collected on the sex of children in families suggests that the Bernoulli model would be a very bad model to adopt, for at least two reasons. First, one theory suggests that the 'boy' probability \( p \), even if it averages 0.514 over a very large population, is probably not the same from family to family, with some couples seeming to have a preponderance for boys, and others for girls, greater than would be suggested simply by sampling variation on the binomial distribution. Second, and even more tantalizingly, some statistical analyses seem to show that the independence assumption of the Bernoulli model breaks down: that is, that Nature has a kind of memory, and the sex of a previous child affects to some degree the probability distribution for the sex of a subsequent child. Never-
theless, for the rest of this chapter, we shall assume that a Bernoulli model does provide, for our purposes, an adequate fit to the observed numbers of the sexes in families.

Statistical model testing is one very important way of testing biological and other scientific theories: if what actually happens is wildly different from what one might reasonably expect to happen if the scientific theory were true, then that scientific theory probably does not hold. It is the application of the science of statistics that helps one decide how different ‘different’ can reasonably be if due to chance alone. These tests are called tests of ‘goodness of fit’ and are dealt with in Chapter 9.

**Example 3.14 Silicon chips**

The manufacture of silicon chips is an extremely sensitive operation, requiring engineering accuracies many orders of magnitude greater than those required in most other manufacturing contexts, and a working environment that is clinically ‘clean’. (At the time of writing, only the production of compact discs requires higher standards.) In the early days of chip technology, most chips were ‘defective’—they did not work properly (or often, they did not work at all). Either they were dirty (a mote 0.5 microns across can cause havoc on a circuit board where the tracks carrying current are only 0.3 microns across); or not all the connections were correctly made. At all the stages of slicing, lapping, etching, cleaning and polishing involved in the manufacture of a chip, defective units are identified and removed. Even so, possibly as many as one chip in twenty is faulty.

During the manufacturing process, there are probably ‘runs of rough’, intervals during which nothing seems to go very well and the product defective rate is rather high. These periods will alternate with intervals where things go rather better than average. However, we shall assume for the rest of this chapter that chip quality can be regarded as invariant and independent from chip to chip, unrealistic though this might seem.

We have just seen that the Bernoulli model does not fit exactly in either Example 3.13 or Example 3.14, and yet we are going to go on and apply it in what follows! This is the way things often are in statistics (and in science, in general): these models are merely supposed to be adequate reflections of reality, not perfect ones. For many purposes, the Bernoulli model is adequate for both these situations.

### 3.3.1 The geometric distribution

After studying Chapter 2, you can answer the following types of questions. The following questions relate to Example 3.13. In families of five children, what proportion of families have all boys (take $p$ equal to 0.514)? What is the probability that in a family of four children, all the children will be girls? In what proportion of families of three children do the boys outnumber the girls? The next question relates to Example 3.14. If silicon chips are boxed in sealed batches of one hundred, what is the probability that a purchaser of a box will find he has bought more than ten defectives?

But now consider the following extensions to these problems.
Example 3.13 continued

One of the factors complicating the development of a satisfactory statistical model for family size and structure is that parents often impose their own 'stopping rules' for family limitation, depending on the number or distribution of boys and girls obtained so far. For instance, among completed two-child families in a recent issue of *Who’s Who in America* there was a striking excess of boy-girl and girl-boy sets, more than would be suggested by a simple binomial model: parents (apparently) prefer to contrive their families to include at least one of each sex. Family limitation rules may be more extreme than this: for instance, ‘keep going until the first son is born, then stop’. Under this rule, completed families (M for a son, F for a daughter) would show the structure M, FM, FFM, FFFM, .... The number of children in a completed family of this type is evidently a random variable: what is its probability distribution? ■

Example 3.14 continued

A quality inspector at a silicon chip factory introduces a new quality test. At random times he will sample completed chips from the assembly line. He makes a note of the number of chips sampled up to and including the first defective he finds. If this number reaches or exceeds some predetermined tolerance limit, then he will assume that factory procedures are running efficiently. Otherwise (defectives are occurring too frequently) the production process is stopped for assessment and readjustment. ■

You will have noticed that in both these examples the same type of random variable is being counted: essentially, the number of trials from the start of a sequence to the first success. Notice that the trial at which that success occurs is included in the count.

The assumptions of a sequence of Bernoulli trials are that the outcomes of successive trials are independent, and that the probability of success remains the same from trial to trial. If these twin assumptions hold, and if the number of trials to the first success is denoted by $N$, then we can say

- $N = 1$ if the first trial is a success,
- $N = 2$ if the first trial is a failure, the second a success,
- $N = 3$ if the first two trials are failures, the third a success,

and so on.

Exercise 3.9

Evidently, the number $N$ is a random variable: it is impossible at the start of the sequence to forecast with certainty the number of the trial at which success will first occur. Assume independence and that the probability of success at any trial is $p$, $0 < p < 1$, in the following.

(a) Write down the probability $P(N = 1)$.
(b) Write down the probability $P(N = 2)$.
(c) Write down the probability $P(N = 3)$. 

118
(d) Using your answers to parts (a), (b) and (c), try to find a general formula for the probability \( P(N = n) \).

(e) State the range of possible values for the random variable \( N \).

The results of Exercise 3.9 lead to the following definition. Like the Bernoulli, the binomial and the normal distributions, the probability model whose definition follows is one of the 'standard' probability models. It is the third to be associated with the fundamental notion of a Bernoulli trial.

**The geometric distribution**

If in a sequence of independent trials the probability of success is constant from trial to trial and equal to \( p \), \( 0 < p < 1 \), then the number of trials up to and including the first success is a random variable \( N \), with probability function given by

\[
P(N = n) = p_N(n) = q^{n-1}p, \quad n = 1, 2, \ldots, \tag{3.13}
\]

where \( q = 1 - p \).

The random variable \( N \) is said to follow a geometric distribution with parameter \( p \) and this is written \( N \sim G(p) \).

The reason for the name 'geometric' is that the sequence of probabilities \( P(N = 1), P(N = 2), \ldots \), form a geometric progression: each term is a constant multiple (in this case, \( q \)) of the preceding term. That multiple is less than 1 (since it is a probability), so successive terms of the probability function of \( N \) become smaller and smaller. This is illustrated in Figure 3.8. In (a), the parameter \( p \) is equal to 0.8—that is quite high: you would not have to wait long for the first successful trial. In (b), the parameter \( p \) is much lower, with the probability of success equal to only 0.3. In this case, you could find you have to wait for quite a time for the first success to occur.

It is very common in the context of Bernoulli trials to write the probability of failure as \( q \), where \( q = 1 - p \).

![Figure 3.8](image)

**Figure 3.8** (a) \( N \sim G(0.8) \) (b) \( N \sim G(0.3) \)

Notice that whatever the value of \( p \), the most likely value of \( N \) is 1.
**Example 3.15 Family size with a stopping rule**

One of the difficulties in the collection and analysis of data on family structures is in the definition of a 'completed' family. Parents do not always know whether or not more children will appear. Table 3.9 contains simulated data on family size (the number of children) for 1000 completed families under a hypothetical rule of 'stop after the first son', where $p$ is taken to be $18/35$ (Bernoulli's estimate for the probability of a boy). The table gives the observed frequency for each family size. Also shown are the theoretical frequencies, obtained by multiplying the probabilities $p_N(n)$ by 1000. For instance, the probability that a family is of size 3 is

$$p_N(3) = P(FFM) = \left(\frac{17}{35}\right)^2 \left(\frac{18}{35}\right) = 0.121329,$$

and multiplying this by 1000 gives the theoretical frequency for families of size 3 in a sample of 1000.

<table>
<thead>
<tr>
<th>Family structure</th>
<th>Family size</th>
<th>Observed frequency</th>
<th>Theoretical frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1</td>
<td>508</td>
<td>514.3</td>
</tr>
<tr>
<td>FM</td>
<td>2</td>
<td>255</td>
<td>249.8</td>
</tr>
<tr>
<td>FFM</td>
<td>3</td>
<td>138</td>
<td>121.3</td>
</tr>
<tr>
<td>FFFM</td>
<td>4</td>
<td>53</td>
<td>58.9</td>
</tr>
<tr>
<td>FFFFFM</td>
<td>5</td>
<td>24</td>
<td>28.6</td>
</tr>
<tr>
<td>FFFFFFM</td>
<td>6</td>
<td>12</td>
<td>13.9</td>
</tr>
<tr>
<td>FFFFFFFM</td>
<td>7</td>
<td>4</td>
<td>6.8</td>
</tr>
<tr>
<td>FFFFFFFFM</td>
<td>8</td>
<td>3</td>
<td>3.3</td>
</tr>
<tr>
<td>FFFFFFFFFM</td>
<td>9</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>$\geq 10$</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

In this case the geometric 'fit' to the simulated data seems quite good; but this does not prove anything, since the data were generated in the first place from a precisely stated geometric model.

The next example is necessarily sparse, but gives data on birth order taken from an investigation reported in 1963.

**Example 3.16 Salt Lake City data**

Details were obtained on the sequence of the sexes of children in 116458 families recorded in the archives of the Genealogical Society of the Church of Jesus Christ of Latter Day Saints at Salt Lake City, Utah. The records were examined to find the stage at which the first daughter was born in 7745 families where there was at least one daughter. The data are summarized in Table 3.10.

<table>
<thead>
<tr>
<th>First daughter</th>
<th>Family structure</th>
<th>Family size</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firstborn</td>
<td>F</td>
<td>1</td>
<td>3684</td>
</tr>
<tr>
<td>Secondborn</td>
<td>MF</td>
<td>2</td>
<td>1964</td>
</tr>
<tr>
<td>Thirdborn</td>
<td>MMF</td>
<td>3</td>
<td>1011</td>
</tr>
<tr>
<td>Fourthborn</td>
<td>MMMF</td>
<td>4</td>
<td>549</td>
</tr>
<tr>
<td>Later than fourth</td>
<td>MMMMF</td>
<td>$\geq 5$</td>
<td>537</td>
</tr>
</tbody>
</table>

Here, the data are really too sparse to provide evidence for or against a geometric model; but it is worth noting that successive frequencies are in the ratio

\[
\frac{1964}{3684} = 0.53, \quad \frac{1011}{1964} = 0.51, \quad \frac{549}{1011} = 0.54; \\
\]

all approximately one-half. The frequencies themselves form a geometric progression (roughly speaking).

The geometric probability function \( p_N(n) = q^{n-1}p \) is a much simpler formula than, for instance, the binomial probability function, and nobody would bother to publish tables listing the value of \( p_N(n) \) for different values of \( n \) and \( p \). The exercise that follows is quite straightforward: it only requires a few key-presses on your calculator.

**Exercise 3.10**

(a) Suppose that Nicholas Bernoulli was right, and his hypothesis that the sex of children may be modelled as independent rolls of a 35-sided die, with 18 faces marked ‘boy’ and the other 17 marked ‘girl’, was a correct one. Under a stopping rule ‘stop after the first son’, what proportion of completed families comprise at least four children?

(b) Suppose that the proportion of defective chips leaving a chip factory’s assembly line is only 0.012: manufacturing standards are quite high. A quality inspector collects daily a random sample. He examines sampled chips in the order in which they came off the assembly line until he finds one that is defective. He decides that he will halt production if fewer than six chips need examining. What proportion of his daily visits result in a halt in production?

As we have seen, there is no convenient formula for the sum

\[
P(X \leq x) = \sum_{j=0}^{x} P(X = j) = \sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j},
\]

when \( X \) is binomial \( B(n,p) \); so its calculation could be very time-consuming (and error-prone). If these probabilities exist in printed form, the problem reduces to one of using the tables correctly; alternatively, they may be obtained more or less directly from your computer.

In the case of the geometric random variable \( N \), it is possible to deduce a formula for the probability \( P(N \leq n) \) by writing down successive individual probabilities and then finding their sum. Actually, it is rather easier to return to the original context where the model arose and argue straight from there. An argument ‘from first principles’ goes as follows.

Remember that we are interested in counting the number of trials necessary to record the first success. If we are rather unlucky, the sequence could go on for a long time. If we start with 7 failures in the first 7 trials, for instance, then we know that \( N \) has to be at least 8 (that is, more than 7). The probability of recording 7 failures in 7 trials is \( q^7 \), so \( P(N > 7) = q^7 \). If after 20 trials we have recorded 20 failures, then we know that \( N \) has to be more than 20. The probability of recording 20 failures in 20 trials is \( q^{20} \), so \( P(N > 20) = q^{20} \). In general, the probability of recording \( n \) failures in \( n \) trials is \( q^n \), so

\[
P(N > n) = q^n.
\]
It follows that the cumulative distribution function \( F_N(n) \) for the geometric random variable \( N \) is given by

\[
F_N(n) = P(N \leq n) = 1 - P(N > n) = 1 - q^n, \tag{3.14}
\]

where \( n = 1, 2, 3, \ldots \).

The argument that led to (3.14), the statement of the c.d.f. of the random variable \( N \), is quite watertight; but you might like to see the following mathematical check. The distribution function \( F_N(n) \) can be obtained directly from the probability function \( p_N(n) \) by summing probabilities:

\[
P(N \leq n) = P(N = 1) + P(N = 2) + \cdots + P(N = n),
\]

so

\[
P(N \leq n) = p + qp + q^2p + q^3p + \cdots + q^{n-1}p. \tag{3.15}
\]

The terms in this series form a geometric progression (sometimes abbreviated to g.p.). Perhaps you already know a formula for the sum of the first \( n \) terms in a geometric progression. If not, we can proceed in this case as follows.

Multiplying both sides of (3.15) by \( q \), we obtain

\[
qP(N \leq n) = qp + q^2p + q^3p + \cdots + q^{n}p.
\tag{3.16}
\]

Subtracting (3.16) from (3.15), we obtain (since most terms vanish) the identity

\[
(1 - q)P(N \leq n) = p - q^n p.
\]

So replacing \( 1 - q \) by \( p \) and dividing by \( p \), we have

\[
P(N \leq n) = 1 - q^n, \quad n = 1, 2, 3, \ldots \tag{3.17}
\]

This is the answer that was obtained in (3.14) by a more direct argument.

**Exercise 3.11**

The proportion of defective products in a battery factory is 0.02. A quality control inspector tests batteries drawn at random from the assembly line. What is the probability that he will have to examine more than 20 to obtain a faulty one? What is the probability that he will have to examine at least 50?

**3.3.2 The mean and variance of the geometric distribution**

As with all models for random variables, two useful measures exist: one to give an idea of what value to expect, and the other to suggest how far away actual values might be from that expected value. These are the mean and standard deviation of the probability distribution.
Exercise 3.12

Without doing any arithmetic, try to answer the following questions. (Just jot down your first reaction.)

(a) The probability that a fair coin shows Heads when it is tossed is \( \frac{1}{2} \). How many times, on average, do you think a coin would need to be tossed to come up Heads?

(b) The probability that a Double-Five comes up showing 5 when it is rolled is \( \frac{2}{6} = \frac{1}{3} \). On average, how many times do you think a Double-Five would need to be rolled to show a 5?

(c) One car in six on British roads is white. If you stand by the side of the road and start counting, how many cars, on average, do you think you would have to count to record your first white one?

(d) The probability that a car starts first time in the morning is \( p \), where \( 0 < p < 1 \). Assuming that attempts to start it can be modelled as a sequence of Bernoulli trials, guess in terms of \( p \) the average number of attempts necessary to get it going each morning. (The assumption that a Bernoulli model will be a useful one is probably not reasonable in this case: most cars just need 'encouraging'. If it does not start first time, then it will almost certainly start at the second or third; and if not then, then not at all.)

What did you write down in Exercise 3.12(d)? Intuitively, you might feel that the mean of the geometric distribution with parameter \( p \) is the reciprocal of \( p \), \( \frac{1}{p} \). In this case your intuition is not misplaced. A proof of this result is as follows.

The mean of a geometric random variable \( N \) with parameter \( p \) is given by

\[
E(N) = \sum_{n=1}^{\infty} np_N(n) = 1p + 2qp + 3q^2p + 4q^3p + \cdots. \tag{3.18}
\]

This series for \( E(N) \) is not itself a geometric progression, since terms are not obtained from the preceding term by multiplying by a constant factor—the coefficients 1, 2, 3, \ldots are complicating features. However, proceeding as before, let us try multiplying both sides by \( q \). This gives

\[
qE(N) = 1qp + 2q^2p + 3q^3p + 4q^4p + \cdots.
\]

If we now subtract this expression from (3.18), on the left-hand side we have \( E(N) - qE(N) = pE(N) \); while on the right-hand side, terms almost vanish as they did before, but not quite. We are left with

\[
pE(N) = 1p + (2 - 1)qp + (3 - 2)q^2p + (4 - 3)q^3p + \cdots
\]

\[
= p + qp + q^2p + q^3p + \cdots,
\]

and the series on the right-hand side we know sums to 1: it is just a list of all the terms of the geometric probability mass function, so it must sum to 1.

Hence we have \( pE(N) = 1 \).

Dividing by \( p \) gives the final result:

\[
E(N) = \frac{1}{p}. \tag{3.19}
\]

The less likely an event is, the longer one should expect to wait for it to happen.
For a probability distribution as skewed (i.e. as asymmetric—see Figure 3.8) as the geometric distribution, knowledge of the value of the standard deviation is not as useful as it is in the case of the normal distribution, where (as you will see in Chapter 5) all probability statements may be made in terms of the number of standard deviations an observation is from its expected value. However, the following result will be useful for future work: we shall see in Chapter 5 that knowledge of the variance of a random variable can be put to other uses. This result for the variance of the geometric random variable \( N \sim G(p) \) is included without proof:

\[
V(N) = \frac{q}{p^2}. \tag{3.20}
\]

So the standard deviation of \( N \) is \( SD(N) = \sqrt{q/p} \).

The next exercise summarizes the work of this section.

**Exercise 3.13**

In some board games, progress round the board is dictated by the score from a roll of a six-sided die. In some games, you cannot start playing until you have obtained your first six (and then you move accordingly). If you score some other number, you have to wait until your next turn and then make another attempt.

(a) What is the probability that you can start playing with your first roll of the die?

(b) What is the probability that you can start playing only at your second roll? At your third?

(c) What is the probability that you will need at least six rolls to get started?

(d) Find the expected number of rolls required to get you started, and calculate the standard deviation.

### 3.4 Two models for uniformity

#### 3.4.1 The discrete uniform probability distribution

In Chapter 2, page 54 we considered the theoretical ‘perfect die’, which when rolled would land displaying any one of its six faces with equal probability. At any given roll of the die, the outcome is a random variable—one cannot forecast precisely what will happen. The probability mass function for the random variable (\( X \), say) is given by

\[
p(x) = \frac{1}{6}, \quad x = 1, 2, \ldots, 6.
\]

This is an example of a random variable following a discrete uniform distribution. The list of possible values that \( X \) can take (the range of \( X \)) is given as a set of integers with stated lower and upper limits; and no possible value is more probable than any other possible value.
Example 3.17 Digit frequencies

Most computers include a 'random number generator' designed to print out in no predictable order but with equal likelihood the digits 0, 1, 2, ..., 9 for as long as the user requires. (Many printed books of statistical tables contain at least a page of such random digits. A list of random digits is given in Table A1.) The way the computer generates successive digits is to follow some complicated rule involving earlier digits (and possibly the date or time as well, if the computer has an internal clock). So what is printed out is not random (in the sense that if you knew the rule you could predict the sequence exactly), but merely indistinguishable from random, or at least similar in certain key respects to the output of a random device such as a ten-sided die.

In an experiment using four different computer programs, the following digit frequencies in sequences of 1000 digits were observed.

Table 3.11 Digit frequencies (four programs)

<table>
<thead>
<tr>
<th>Digit</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC v.1.09</td>
<td>92</td>
<td>107</td>
<td>85</td>
<td>85</td>
<td>109</td>
<td>95</td>
<td>104</td>
<td>95</td>
<td>113</td>
<td>115</td>
</tr>
<tr>
<td>GW-Basic v.3.23</td>
<td>85</td>
<td>110</td>
<td>91</td>
<td>95</td>
<td>106</td>
<td>110</td>
<td>92</td>
<td>106</td>
<td>101</td>
<td>104</td>
</tr>
<tr>
<td>Spida v.5.50</td>
<td>110</td>
<td>94</td>
<td>86</td>
<td>97</td>
<td>101</td>
<td>94</td>
<td>113</td>
<td>133</td>
<td>84</td>
<td>88</td>
</tr>
<tr>
<td>Minitab v.7.20</td>
<td>112</td>
<td>93</td>
<td>96</td>
<td>87</td>
<td>108</td>
<td>84</td>
<td>103</td>
<td>120</td>
<td>111</td>
<td>86</td>
</tr>
</tbody>
</table>

The four bar charts in Figure 3.9 show the sample relative frequencies for each of the ten digits and for the four programs. The sample relative frequencies may be compared with the theoretical proportions: $\frac{1}{10}$ in each case. There is some evidence of variability, as one must expect. However, none of the four programs manifests serious departures from the theoretical uniform distribution.

A different, faulty, program gave the frequencies listed in Table 3.12.

Table 3.12 Digit frequencies (a faulty program)

<table>
<thead>
<tr>
<th>Digit</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
</tr>
</tbody>
</table>

Data provided by F. Daly, The Open University. The Minitab data were supplied by K.J. McConway, The Open University.
So this program resulted in sample relative frequencies of exactly \( \frac{100}{1000} = \frac{1}{10} \) for each of the ten digits 0, 1, 2, \ldots, 9. The uniform 'fit' is perfect!

This example demonstrates that there is more to the generation of random digits than that each should occur with equal likelihood. The faulty program was generating the sequence

\[
0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 0 \ 1 \ \ldots \ 8 \ 9,
\]

and so gave a perfectly uniform distribution of digits. So a 'good' uniform fit does not imply that the random number generator is satisfactory. A 'bad' fit on the other hand might suggest that the generator is unsatisfactory.

**Example 3.18 Month of death of royal descendants**

The data in Table 3.13 give the month of death (January = 1, February = 2, \ldots, December = 12) for 82 descendants of Queen Victoria who died of natural causes.

<table>
<thead>
<tr>
<th>Month</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>13</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

A straight test of uniformity here would not reflect the fact that, leap year or not, February is a short month; but at a first glance the data certainly seem to suggest that the summer months (6, 7, 8, 9) are less likely to include a death than winter months.

(A formal test of uniformity for the data in Table 3.13 should really reflect the fact that the months of January and December are adjacent. The data would be represented not as a bar chart or histogram in the usual sense (see Figure 3.10(a)) but as a circular histogram (see Figure 3.10(b)). Several tests for circular uniformity have been developed, but they will not be explored in this course.)

![Figure 3.10 Representations of the death data](image)

A definition of the discrete uniform distribution is as follows. Notice the phrase 'a definition' rather than 'the definition'. Here, the stated range of the
discrete random variable $X$ is $1, 2, \ldots, n$. Elsewhere, you might see an alternative definition where the list of possible values includes zero: $0, 1, 2, \ldots, m$, say. The essential feature is that each of the possible values occurs with equal probability.

The discrete uniform distribution
The random variable $X$ is said to follow a discrete uniform distribution if it has probability mass function
\[ p(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n. \]

Again, there is a whole family of discrete uniform probability distributions: the indexing parameter in this case is $n$, the maximum attainable value.

The c.d.f. of $X$ is found from
\[ P(X \leq x) = p(1) + p(2) + \cdots + p(x) = \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = \frac{x}{n}, \quad x = 1, 2, \ldots, n. \]

That is,
\[ F(x) = \frac{x}{n}, \quad x = 1, 2, \ldots, n. \]

The mean of the random variable $X$ following a discrete uniform distribution with parameter $n$ is given by
\[ E(X) = \sum_{j=1}^{n} j p(j) = \sum_{j=1}^{n} j \left( \frac{1}{n} \right) = \frac{1}{n} (1 + 2 + 3 + \cdots + n) = \frac{1}{n} \left( \frac{1}{2} n(n + 1) \right) = \frac{1}{2} (n + 1) \]
(which is, as you might expect, the middle of the range of $X$).

The variance of $X$ is given by
\[ V(X) = \frac{n^2 - 1}{12}. \]
This result is obtained by a straightforward but time-consuming application of the variance formula, and you need not bother with the details.

Here, the result
\[ 1 + 2 + \cdots + n = \frac{1}{2} n(n + 1) \]
is used. You can see this by writing
\[ S = 1 + 2 + \cdots + n; \]
then write the right-hand side last to first, as
\[ S = n + (n - 1) + \cdots + 1. \]
Now sum each side of the two expressions, term by term. This gives
\[ 2S = (n + 1) + (n + 1) + \cdots + (n + 1) = n(n + 1). \]
Dividing by 2 gives
\[ S = \frac{1}{2} n(n + 1). \]
Exercise 3.14

Generate 1200 rolls of a fair six-sided die.
(a) Obtain a frequency table for your data.
(b) Plot a bar chart of your data.
(c) Evaluate the sample mean and sample standard deviation, and compare these two statistics with the corresponding population moments.

Essentially, that is all that needs to be said about the discrete uniform distribution. There is a continuous analogue that we shall now examine.

3.4.2 The continuous uniform probability distribution

Here are some simple examples of situations for which a useful statistical model has not yet been developed.

Example 3.19 Admissions to an intensive care unit

In Chapter 2, Table 2.22, data were given on the waiting time (in hours) between the first 41 admissions to an intensive care unit. The full data set lists 254 admissions over a period of about 13 months. Information is given on the date and time of day of admission. It would be interesting, and helpful to planners, to explore whether admissions were more frequent at some times of day than others, or whether any time of admission is as likely as any other. The times of admission over two-hour intervals are summarized in Table 3.14.

Actually, the data are sufficiently sparse over the early hours 4 am–10 am (and dense over the early evening period 4 pm–6 pm) to provide strong statistical evidence that frequency of admission does depend on the time of day. We shall look in Chapter 9 at statistical procedures enabling this sort of evidence to be examined. The important point is that an essential requirement for such procedures is the formulation of a model to describe (possibly rather badly) the variation observed.

Example 3.20 Faulty cable

Faults in underground television cable cause degradation, or even complete loss, of the signal. When this happens, the cable needs to be repaired. In the absence of any indication of where the fault might be, the repair company has to search the cable until the fault is located—this is just as likely to be near the end, near the beginning or in the middle of the cable. The distance searched to locate the fault is a random variable, and a factor in the cost of the repair.

Example 3.21 Green-haired Martians

Suppose you were required to guess the unknown proportion \( p \) of inhabitants of Mars with green hair (assuming you knew that they had hair). One way of expressing your (assumed) total ignorance on this matter might be to say \( p = \frac{1}{2} \), on the principle that as far as you are concerned a Martian's hair is as likely as not to be green.


Table 3.14 Time of day of 254 admissions to an intensive care unit

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>midnight–2 am</td>
<td>14</td>
</tr>
<tr>
<td>2 am–4 am</td>
<td>17</td>
</tr>
<tr>
<td>4 am–6 am</td>
<td>5</td>
</tr>
<tr>
<td>6 am–8 am</td>
<td>8</td>
</tr>
<tr>
<td>8 am–10 am</td>
<td>5</td>
</tr>
<tr>
<td>10 am–midday</td>
<td>25</td>
</tr>
<tr>
<td>midnight–2 pm</td>
<td>31</td>
</tr>
<tr>
<td>2 pm–4 pm</td>
<td>30</td>
</tr>
<tr>
<td>4 pm–6 pm</td>
<td>36</td>
</tr>
<tr>
<td>6 pm–8 pm</td>
<td>29</td>
</tr>
<tr>
<td>8 pm–10 pm</td>
<td>31</td>
</tr>
<tr>
<td>10 pm–midnight</td>
<td>23</td>
</tr>
</tbody>
</table>
Another way to express your ignorance is to say that the proportion \( p \) could be anywhere between 0 and 1, and that you have no evidence that any value is any more likely than any other. Then you express your guess in terms of a probability distribution, rather than giving any particular value. ■

Example 3.21, though somewhat contrived, actually raises some very important questions. Much of the science of statistics that we shall be exploring in this course is directed at the problem of ‘estimation’: making a guess, and preferably a good one, at the value of some model parameter. This parameter is a constant; the problem is that we do not know its value.

An entirely different approach is to express your uncertainty about the value of such descriptive parameters by giving them a probability distribution which says ‘this value is credible but unlikely, this value is (to me) the most credible, this one is not at all plausible, . . . ’ and so on. The data you collect then allow you to modify or update this probability distribution. The whole question of this kind of inference, called Bayesian inference, is a very interesting one, but unfortunately it is beyond the scope of this course.

The central idea in these examples has been that of ‘no preferred value’. The probability model that permits left- and right-hand bounds to be stated, and that carries the sense that between those bounds there is no preferred value, is known as the continuous uniform distribution. Since there is no preferred value in the range of a continuous uniform random variable \( X \), the height of the p.d.f. of \( X \) on a sketch must be constant over the range. For instance, suppose that \( X \) can take values between \( a \) and \( b \) (\( a < b \)); then a sketch of its p.d.f. will look like that in Figure 3.11.

![Figure 3.11](image)

The p.d.f. is of the form \( f(x) = h \), for \( a \leq x \leq b \). Since the total area under the p.d.f. must be 1, the area of the rectangle in Figure 3.11 must be 1, so \((b - a)h = 1\). It follows that \( h = 1/(b - a)\). Hence the following definition can be made.

**The continuous uniform distribution**

The continuous random variable \( X \), equally likely to take any value between two stated bounds \( a \) and \( b \) (\( a < b \)), is said to be uniformly distributed over the interval \( a \leq x \leq b \) and has probability density function

\[
f(x) = \frac{1}{b - a}, \quad a \leq x \leq b.
\]  

This is written \( X \sim U(a, b) \).
A sketch of the density is shown in Figure 3.12.

\[
\begin{array}{c}
\text{f(x)} \\
\frac{1}{b-a} \\
0 \quad a \quad b \quad x
\end{array}
\]

**Figure 3.12** The uniform density, \( U(a, b) \)

In Figure 3.12 it is assumed that the bounds \( a \) and \( b \) are both positive. This is not a necessary condition. Depending on the situation being modelled, one bound or both bounds could be negative. The only constraint is that \( a \) should be less than \( b \).

**Example 3.19 continued**

In this example, one possibility (in the absence of data) is that times of admission might be independent of the time of day. A model for the time \( T \) of admission reflecting this suggestion is \( T \sim U(0, 24) \) (using the 24-hour clock). Once suggested, a model can be examined for the adequacy of its fit to a data set. (In this case, a uniform model provides a very bad fit.)

**Example 3.21 continued**

In this case the unknown proportion \( p \) of green-haired Martians is given a uniform probability distribution with \( a = 0 \) and \( b = 1 \).

**Exercise 3.15**

Use a sketch of the p.d.f. of \( X \) to find the following for the uniform random variable \( X \sim U(a, b) \):

(a) the mean (use a symmetry argument);
(b) the cumulative distribution function.

The mean of the uniform distribution \( U(a, b) \) is given by

\[
E(X) = \mu = \frac{a + b}{2}
\]

(in other words, the midpoint of the range) and the cumulative distribution function is given by

\[
F(x) = \frac{x - a}{b - a}, \quad a \leq x \leq b.
\]

Calculation of the variance involves an exercise in integration. It is given by

\[
V(X) = \sigma^2 = \frac{(b - a)^2}{12}.
\]

Notice that the value of the variance depends only on the difference \( b - a \) and not on the actual values of \( a \) and \( b \).
3.4.3 The standard continuous uniform distribution

One special case of the uniform distribution is the member of the family that starts at 0 and ends at 1; this has applications to Bayesian inference and to simulation (amongst others).

The probability density function for the uniform distribution on (0,1) is given (setting \( a = 0, b = 1 \) in (3.23)) by

\[
f(x) = 1, \quad 0 \leq x \leq 1.
\]  

(3.27)

The uniform distribution \( U(0,1) \) is known as the standard uniform distribution. Its graph is shown in Figure 3.13.

\[
\begin{align*}
f(x) \\
0 & \quad 1 & x
\end{align*}
\]

Figure 3.13 The uniform density, \( U(0,1) \)

The cumulative distribution function for the standard uniform distribution is particularly easy to find: the probability \( P(X \leq x) \) turns out to be \( x \) itself. Formally,

\[
P(X \leq x) = F(x) = \int_0^x f(v) \, dv = \int_0^x 1 \, dv = [v]_0^x = x.
\]  

(3.28)

However a geometrical argument like that used in Exercise 3.15(b) is easier!

The mean of the standard continuous uniform random variable \( X \) is

\[
E(X) = \mu = \frac{1}{2}.
\]

Exercise 3.16

Write down the variance of the standard uniform random variable \( X \sim U(0,1) \), and hence calculate its standard deviation.

This section ends with one final example.

Example 3.22 Traffic wardens

Many parking zones in city centres have notices that read: Waiting limited to 60 minutes. Return prohibited within 1 hour. One way that traffic wardens keep track of which cars are parked where, and when, is to note the position of the valve dust-cap on one (or, in some cases, all) of the four wheels. A typical record is shown in Figure 3.14.

A motorist accused of overstaying in a parking zone, claimed that he had been away and returned only after a proper interval as the law required. Counsel for the authority taking him to court produced evidence that, if this was so,
then the rather unlikely event had occurred that all four valve dust-caps had ended up in identical locations about the wheel (at least, to within recording variation) as they had been earlier. Expert statistical advisors suggested that after a car journey of anything more than a negligible distance, even given the four starting orientations, there would at the end of the journey be no orientation for the dust-caps more likely than any other. It may also be assumed that even after a short journey, and given the four starting orientations, the final orientations of the four dust-caps are independent. (In cornering, the wheels travel different distances; differential axles preventing skidding.) Assuming that the four dust-cap positions had been recorded to an accuracy no greater than one twelfth of a turn (like hours on a clock-face—accuracy rather greater than this is not difficult to attain) then the motorist's version of events meant that an event with probability $(1/12)^4 = 0.000048$ had occurred. The magistrates found these odds (more than 20,000 to 1 against) less than credible.

### 3.5 Population quantiles

In Sections 3.3 and 3.4, three particular probability models have been described: the geometric distribution, and both discrete and continuous uniform distributions. To finish the chapter, let us, in this section, consider a topic generally applicable to any probability distribution. As in Section 3.1, where the population analogues of the sample mean, variance and standard deviation were described, here we shall consider the population analogues of some more quantities you met in Chapter 1, specifically the sample median and sample quartiles (and the sample interquartile range). More generally, it will prove to be useful to introduce a new idea—which covers population medians and quartiles—namely that of population quantiles.

This is another instance where it is best to treat discrete and continuous situations separately. It is slightly more straightforward to deal with the continuous case, so that is where we shall start.

#### 3.5.1 Quantiles for continuous probability distributions

**The median**

In Chapter 1, the sample median was introduced as the middle value of a set of values (where we had to be careful about what to do with samples of even size). The middle value, by definition, splits the data into two portions either side of it, with equal numbers of data points in each. The idea is to split the data so that one half has values smaller than the sample median and the other half has values greater than it (although for samples of odd size this is not exact, since we 'lose' the median value itself so that the subsets are of size $(n - 1)/2$ rather than $n/2$).

In the modelling context, there is a natural analogue of this notion of splitting the data in half: we should like to define our population median $m$, say,
to split the distribution into two halves in the sense that, if the random variable concerned is denoted by \( X \) and its median by \( m \), \( P(X \leq m) = \frac{1}{2} \) and \( P(X \geq m) = \frac{1}{2} \). In terms of the c.d.f., this may be written \( F(m) = 1 - F(m) = \frac{1}{2} \).

For a continuous random variable, the **population median** is the value \( x \) which is the solution to the equation

\[
F(x) = \frac{1}{2};
\]

this solution is denoted \( x = m \).

**Example 3.23 Salary distributions**

Distributions of salaries are typically skewed to the right (positively skewed) since relatively few people earn large salaries. As a consequence, the median is in many ways a better summary statistic for salary surveys than the mean. The median tells us the midpoint of the salary distribution within a population, in the sense that half of the people earn less and half earn more.

**Exercise 3.17**

(a) In a simulation experiment, a computer is programmed to generate pseudo-random numbers from the continuous uniform distribution on \( U(a, b) \). What is the median of this distribution?

(b) Theory suggests that a particular random variable \( X \) has p.d.f.

\[
f(x) = 3x^2, \quad 0 < x < 1.
\]

(i) Sketch the density function of \( X \).

(ii) The mean of \( X \) is \( \mu = \frac{3}{4} \). Show this on your sketch of the density of \( X \).

(iii) The c.d.f. of \( X \) is given by

\[
F(x) = P(X \leq x) = \int_0^x f(v) \, dv = \int_0^x 3v^2 \, dv = \left[ v^3 \right]_0^x = x^3, \quad 0 \leq x \leq 1.
\]

Calculate the median \( m \) of the random variable \( X \) and show \( m \) on your sketch.

**Quartiles**

The general idea in the definition of sample quartiles in *Chapter 1*, Section 1.3 was to choose values which split the data into proportions of one-quarter and three-quarters. There are, of course, two ways to do the latter, and consequently there are two quartiles: a lower quartile designed to have (approximately) a quarter of the data with values smaller than it and three-quarters with larger values, and an upper quartile for which three-quarters of the data have values below it and a quarter above it. Their difference gives the sample interquartile range. To define population quartiles for continuous distributions, 'proportions' are replaced by 'probabilities', as follows. For
example, to obtain the lower population quartile, denoted by $q_L$, the defining requirement is that $P(X \leq q_L) = \frac{1}{4}$.

For a continuous random variable, the lower quartile is the value $x$ which is the solution to the equation

$$F(x) = \frac{1}{4};$$

the solution is denoted $x = q_L$.

Similarly, the upper quartile is the value $x$ which is the solution to the equation

$$F(x) = \frac{3}{4},$$

and this solution is denoted $x = q_U$.

The interquartile range is $q_U - q_L$.

Just as the sample interquartile range is a measure of the spread of samples, so the population interquartile range, like the standard deviation, is a measure of the dispersion of population models.

**Example 3.24 Childhood growth**

In studies of growth of children, it is often of interest to show the position of a particular child relative to the overall distribution of heights of children of that age. The upper and lower quartiles for the distribution of heights at each age provide useful information. The lower quartile is the height such that 25% of children of that age are shorter; and the upper quartile is the height such that 75% are shorter (only 25% are taller). The median is, of course, the height such that 50% are shorter and 50% are taller.

The interquartile range of the heights of the children for any particular age covers the range of values taken by the 'middle 50%' of children of that age. It is the difference between the heights of the tallest and shortest children left after excluding the shortest 25% and the tallest 25%.

**Exercise 3.18**

Find the interquartile range for the distribution with p.d.f.

$$f(x) = 3x^2, \quad 0 < x < 1.$$
For a continuous random variable with c.d.f. $F(\cdot)$, the $\alpha$-quantile is the value $x$ which is the solution to the equation

$$F(x) = \alpha, \quad 0 < \alpha < 1;$$

this value is denoted $q_\alpha$.

So, in particular, the median is $m = q_{0.5}$, the lower quartile is $q_L = q_{0.25}$ and the upper quartile is $q_U = q_{0.75}$. The terms percentile and percentage point are synonymous with ‘quantile’, and are often used when $\alpha$ is expressed as a percentage. Some other special cases of quantiles also have more specialized names. For example, if $\alpha$ is an integer multiple of 0.1, the corresponding quantiles are sometimes called deciles. There is good reason for generalizing to population $\alpha$-quantiles and that is because some values of $\alpha$ other than $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ will be made much use of as the course develops. From Chapter 4 onwards, you will often find that interest centres on the ‘extremes’ of distributions, and that quantiles associated with values of $\alpha$ like 0.9, 0.95, 0.99 and, at the other extreme, 0.1, 0.05, and so on, may be important.

A diagram should help clarify the idea of a population quantile. Figure 3.15 shows a graph of a typical continuous c.d.f. $F(x)$. For a given value $\alpha$, say, on the horizontal axis, you could evaluate (or read off the graph) the corresponding point $F(\alpha)$ (a probability, $0 < F(\alpha) < 1$) on the vertical axis.

![Figure 3.15](image-url)

**Figure 3.15** Calculating $F(\alpha)$ for a given $\alpha$

In Figure 3.16 the starting point is a number $\alpha$ ($0 < \alpha < 1$) on the vertical axis. The corresponding point on the horizontal axis is the $\alpha$-quantile of $X$, the value $q_\alpha$.

![Figure 3.16](image-url)

**Figure 3.16** Calculating $q_\alpha$ for a given $\alpha$

(Expressed mathematically, the $\alpha$-quantile is given by $q_\alpha = F^{-1}(\alpha)$, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$.)
Quantiles for discrete probability distributions

Quantiles for continuous distributions have been defined readily and unambiguously. The equation \( F(x) = \alpha \) is solved for \( x \). In all of our examples, and in all the cases you will encounter in this course, this equation has a unique solution.

Unfortunately, when quantiles for discrete distributions are considered, this simplicity and lack of ambiguity disappears. Quantiles for discrete distributions are not as important as those for continuous ones, so in this course we shall not dwell on them, but shall simply illustrate some of the problems and how they might be overcome.

Let us begin with the example of the uniform distribution on the integers 1, 2, ..., 6: the model adopted for the score \( X \) when a fair die is rolled. The c.d.f. of \( X \) is given by \( F(x) = x/6 \), tabulated in Table 3.15.

Using the above method for finding the median of \( X \), we need to solve the equation \( F(m) = \frac{1}{2} \) for \( m \). This yields \( m = 3 \).

If the median is regarded as an indication of the 'centre' of the distribution then this is unsatisfactory. If we reversed the distribution, starting at the score of 6, and applied the same process to find the median, then we would obtain a value of 4. Given the symmetry of the distribution, this is not very appealing. An attractive property of a symmetric distribution would be that it has the same 'centre' from whichever end you look at it!

An extra problem is illustrated by the probability mass function of a random variable \( Y \) which is uniform on the integers 1, 2, ..., 5. The tabulated cumulative distribution function is given in Table 3.16.

Again, to find the median we need to solve the equation \( F(y) = \frac{1}{2} \). Unfortunately, this equation is not satisfied for any \( y = 1, 2, \ldots, 5 \). At the very least this requires the definition of the median to be modified. One modification is to redefine the population median as the minimum value \( m \) such that \( F(m) \geq \frac{1}{2} \). Here, this would yield the solution \( m = 3 \), which is at least in the middle of the range.

Similar problems arise with other quantiles. The definition we shall use in this course follows.

For a discrete random variable \( X \) with c.d.f. \( F(x) \), the \( \alpha \)-quantile \( q_\alpha \) is defined to be the minimum value of \( x \) in the range of \( X \) satisfying \( F(x) \geq \alpha \).

Example 3.25 Quartiles of the binomial distribution

If \( X \) is binomial \( B(6, 0.6) \), then \( X \) has the probability mass function \( p(x) \) and cumulative distribution function \( F(x) \) given in Table 3.17.

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x) )</td>
<td>0.004 096</td>
<td>0.036 864</td>
<td>0.138 240</td>
<td>0.276 480</td>
<td>0.311 040</td>
<td>0.186 624</td>
<td>0.046 656</td>
</tr>
<tr>
<td>( F(x) )</td>
<td>0.004 096</td>
<td>0.040 960</td>
<td>0.179 200</td>
<td>0.455 680</td>
<td>0.766 720</td>
<td>0.953 344</td>
<td>1</td>
</tr>
</tbody>
</table>
To find the median of $X$ we need to find the minimum value of $x$ such that $F(x) \geq 0.5$. Since $F(3) = 0.455680$ (less than 0.5) and $F(4) = 0.766720$ (greater than 0.5), $x = 4$ is the least value satisfying $F(x) \geq 0.5$. Hence the median $m$ is 4. Similarly, since $F(2) < 0.25 \leq F(3)$, the lower quartile of $X$ is $q_L = 3$. Finally, since $F(3) < 0.75 \leq F(4)$, the upper quartile of $X$ is $q_U = 4$. Notice that in this example the median and the upper quartile are the same.

**Exercise 3.19**

(a) Find the median of the binomial distribution $B(10, 0.5)$.

(b) Find the median of the binomial distribution $B(17, 0.7)$.

(c) Find the upper quartile of the binomial distribution $B(2, 0.5)$.

(d) Find the interquartile range of the binomial distribution $B(19, 0.25)$.

(e) Find the 0.85-quantile of the binomial distribution $B(15, 0.4)$.

### 3.5.3 Population modes

For discrete probability models, the mode, if there is just one, is the value that has the highest probability of occurring (much as the sample mode is the value that occurs the highest proportion of times in a sample). For continuous population models, interest simply shifts to maxima: the maximum, if there is only one, of the p.d.f. rather than the p.m.f. Also—and this notion is especially useful for continuous distributions—distributions can have one or more maxima and thus may be multimodal. You need not take away much more from this short subsection than the idea that it makes sense to talk of, for example, bimodal probability distributions. An example of a bimodal probability density function is shown in Figure 3.17.

![Figure 3.17 A bimodal density function](image)

The Old Faithful geyser at Yellowstone National Park, Wyoming, USA, was observed from 1–15 August 1985. During that time, data were collected on the duration of eruptions and the waiting time between the starts of successive eruptions. There are 299 waiting times (in minutes), and these are listed in Table 3.18.
**Table 3.18** Waiting times (minutes) between eruptions, Old Faithful geyser

<table>
<thead>
<tr>
<th>80</th>
<th>71</th>
<th>57</th>
<th>80</th>
<th>75</th>
<th>77</th>
<th>60</th>
<th>86</th>
<th>77</th>
<th>56</th>
<th>81</th>
<th>50</th>
<th>89</th>
<th>54</th>
<th>90</th>
<th>73</th>
<th>60</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>82</td>
<td>84</td>
<td>54</td>
<td>85</td>
<td>58</td>
<td>79</td>
<td>57</td>
<td>88</td>
<td>68</td>
<td>76</td>
<td>78</td>
<td>74</td>
<td>85</td>
<td>75</td>
<td>65</td>
<td>76</td>
<td>58</td>
</tr>
<tr>
<td>91</td>
<td>50</td>
<td>87</td>
<td>48</td>
<td>93</td>
<td>54</td>
<td>86</td>
<td>53</td>
<td>78</td>
<td>52</td>
<td>83</td>
<td>60</td>
<td>87</td>
<td>49</td>
<td>80</td>
<td>60</td>
<td>92</td>
<td>43</td>
</tr>
<tr>
<td>89</td>
<td>60</td>
<td>84</td>
<td>69</td>
<td>74</td>
<td>71</td>
<td>108</td>
<td>50</td>
<td>77</td>
<td>57</td>
<td>80</td>
<td>61</td>
<td>82</td>
<td>48</td>
<td>81</td>
<td>73</td>
<td>62</td>
<td>79</td>
</tr>
<tr>
<td>54</td>
<td>80</td>
<td>73</td>
<td>81</td>
<td>62</td>
<td>81</td>
<td>71</td>
<td>79</td>
<td>81</td>
<td>74</td>
<td>59</td>
<td>81</td>
<td>66</td>
<td>87</td>
<td>53</td>
<td>80</td>
<td>50</td>
<td>87</td>
</tr>
<tr>
<td>51</td>
<td>82</td>
<td>58</td>
<td>81</td>
<td>49</td>
<td>92</td>
<td>50</td>
<td>88</td>
<td>62</td>
<td>93</td>
<td>56</td>
<td>89</td>
<td>51</td>
<td>79</td>
<td>58</td>
<td>82</td>
<td>52</td>
<td>88</td>
</tr>
<tr>
<td>52</td>
<td>78</td>
<td>69</td>
<td>75</td>
<td>77</td>
<td>53</td>
<td>80</td>
<td>55</td>
<td>87</td>
<td>53</td>
<td>85</td>
<td>61</td>
<td>93</td>
<td>54</td>
<td>76</td>
<td>80</td>
<td>81</td>
<td>59</td>
</tr>
<tr>
<td>86</td>
<td>78</td>
<td>71</td>
<td>77</td>
<td>76</td>
<td>94</td>
<td>75</td>
<td>50</td>
<td>83</td>
<td>82</td>
<td>72</td>
<td>77</td>
<td>75</td>
<td>65</td>
<td>79</td>
<td>72</td>
<td>78</td>
<td>77</td>
</tr>
<tr>
<td>79</td>
<td>75</td>
<td>78</td>
<td>64</td>
<td>80</td>
<td>49</td>
<td>88</td>
<td>54</td>
<td>85</td>
<td>51</td>
<td>96</td>
<td>50</td>
<td>80</td>
<td>78</td>
<td>81</td>
<td>72</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td>87</td>
<td>69</td>
<td>55</td>
<td>83</td>
<td>49</td>
<td>82</td>
<td>57</td>
<td>84</td>
<td>57</td>
<td>84</td>
<td>73</td>
<td>78</td>
<td>57</td>
<td>79</td>
<td>57</td>
<td>90</td>
<td>62</td>
<td>87</td>
</tr>
<tr>
<td>78</td>
<td>52</td>
<td>98</td>
<td>48</td>
<td>78</td>
<td>79</td>
<td>65</td>
<td>84</td>
<td>50</td>
<td>83</td>
<td>60</td>
<td>80</td>
<td>50</td>
<td>88</td>
<td>50</td>
<td>84</td>
<td>74</td>
<td>76</td>
</tr>
<tr>
<td>65</td>
<td>89</td>
<td>49</td>
<td>88</td>
<td>51</td>
<td>78</td>
<td>85</td>
<td>65</td>
<td>75</td>
<td>77</td>
<td>69</td>
<td>92</td>
<td>68</td>
<td>87</td>
<td>61</td>
<td>81</td>
<td>55</td>
<td>93</td>
</tr>
<tr>
<td>53</td>
<td>84</td>
<td>70</td>
<td>73</td>
<td>93</td>
<td>50</td>
<td>87</td>
<td>77</td>
<td>74</td>
<td>72</td>
<td>82</td>
<td>74</td>
<td>80</td>
<td>49</td>
<td>91</td>
<td>53</td>
<td>86</td>
<td>49</td>
</tr>
<tr>
<td>79</td>
<td>89</td>
<td>87</td>
<td>76</td>
<td>59</td>
<td>80</td>
<td>89</td>
<td>45</td>
<td>93</td>
<td>72</td>
<td>71</td>
<td>54</td>
<td>79</td>
<td>74</td>
<td>65</td>
<td>78</td>
<td>57</td>
<td>87</td>
</tr>
<tr>
<td>72</td>
<td>84</td>
<td>47</td>
<td>84</td>
<td>57</td>
<td>87</td>
<td>68</td>
<td>86</td>
<td>75</td>
<td>73</td>
<td>53</td>
<td>82</td>
<td>93</td>
<td>77</td>
<td>54</td>
<td>96</td>
<td>48</td>
<td>89</td>
</tr>
<tr>
<td>63</td>
<td>84</td>
<td>76</td>
<td>62</td>
<td>83</td>
<td>50</td>
<td>85</td>
<td>78</td>
<td>78</td>
<td>81</td>
<td>78</td>
<td>76</td>
<td>74</td>
<td>81</td>
<td>66</td>
<td>84</td>
<td>48</td>
<td>93</td>
</tr>
<tr>
<td>47</td>
<td>87</td>
<td>51</td>
<td>78</td>
<td>54</td>
<td>87</td>
<td>52</td>
<td>85</td>
<td>58</td>
<td>88</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.18 shows a histogram of the geyser waiting time data. Notice that there are two very pronounced modes. These data were referred to in Chapter 1, Figure 1.19.

**Figure 3.18** Waiting times between eruptions, Old Faithful geyser

The histogram suggests that a good model for the variation in waiting times should also be bimodal. This is important: the bimodality may indicate that waiting times are essentially of two characters. ‘Short’ waiting times last a little under an hour (with some variation); ‘long’ waiting times last around 75–80 minutes (with some variation). Research can then begin into the causes of the interesting phenomenon observed.

**Summary**

1. The mean and variance of a discrete integer-valued random variable \( X \), with probability mass function \( p(x) \), are given by

\[
E(X) = \mu = \sum xp(x),
\]

\[
V(X) = \sigma^2 = E((X - \mu)^2) = \sum (x - \mu)^2 p(x),
\]

where the summations are taken over all \( x \) in the range of \( X \).
2. The mean and variance of a continuous random variable $X$, with probability density function $f(x)$, are given by

$$ E(X) = \mu = \int x f(x) \, dx, $$

$$ V(X) = \sigma^2 = E((X - \mu)^2) = \int (x - \mu)^2 f(x) \, dx, $$

where the integrations extend over the range of $X$.

3. Two discrete random variables $X$ and $Y$ are said to be independent if and only if

$$ P(X = x, Y = y) = P(X = x)P(Y = y), $$

for all $x$ in the range of $X$ and all $y$ in the range of $Y$.

Two continuous random variables $X$ and $Y$ are said to be independent if, for example,

$$ P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), $$

for all $x$ in the range of $X$ and all $y$ in the range of $Y$.

4. In a sequence of independent Bernoulli trials indexed by the parameter $p$, $0 < p < 1$, the number $N$ of trials from success to next success is a random variable following a geometric distribution with parameter $p$, written $N \sim G(p)$.

The random variable $N$ has probability function

$$ p_N(n) = q^{n-1}p, \quad n = 1, 2, 3, \ldots, $$

where $q = 1 - p$.

The cumulative distribution function is

$$ F_N(n) = 1 - q^n, \quad n = 1, 2, 3, \ldots. $$

The first two moments of $N$ are $E(N) = 1/p$ and $V(N) = q/p^2$.

5. The random variable $X$ following a discrete uniform probability distribution with parameter $n$ has probability mass function

$$ p(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n. $$

The mean is $\frac{n + 1}{2}$ and the variance is $\frac{n^2 - 1}{12}$.

6. When a continuous random variable $X$ is constrained to take values between stated limits $a$ and $b$ ($a < b$), but within those bounds no value of $x$ is any more likely than any other, then the random variable $X$ is said to follow a continuous uniform distribution on the interval $a \leq x \leq b$, with probability density function

$$ f(x) = \frac{1}{b - a}, \quad a \leq x \leq b. $$

This is written $X \sim U(a, b)$. 
The mean of $X$ is $\frac{a + b}{2}$ and the variance of $X$ is $\frac{(b - a)^2}{12}$.

7. If $X$ is uniformly distributed on the interval $0 \leq x \leq 1$, then this is written $X \sim U(0, 1)$ and $X$ is said to follow the standard uniform distribution.

The random variable $X$ has probability density function

$$f(x) = 1, \quad 0 \leq x \leq 1.$$ 

The cumulative distribution function is

$$F(x) = x, \quad 0 \leq x \leq 1.$$ 

8. The $\alpha$-quantile of a continuous random variable $X$ with cumulative distribution function $F(x)$ is defined to be the solution of the equation

$$\alpha = F(x),$$

written $x = q_\alpha$. In particular, the first (lower) population quartile $q_L$, the median $m$ and the third (upper) quartile $q_U$ are, respectively, the quantiles $q_{0.25}$, $q_{0.50}$ and $q_{0.75}$.

9. The $\alpha$-quantile for a discrete integer-valued random variable $X$ with cumulative distribution function $F(x)$ is defined to be the minimum value of $x$ in the range of $X$ such that $F(x) \geq \alpha$. 

140