23

Confidence intervals for the mean

Sometimes, a *range* of plausible values for an unknown parameter is preferred to a single estimate. We shall discuss how to turn data into what are called *confidence intervals* and show that this can be done in such a manner that definite statements can be made about *how* confident we are that the true parameter value is in the reported interval. This level of confidence is something you can choose. We start this chapter with the general principle of confidence intervals. We continue with confidence intervals for the mean, the common way to refer to confidence intervals made for the expected value of the model distribution. Depending on the situation, one of the four methods presented will apply.

23.1 General principle

In previous chapters we have encountered sample statistics as estimators for distribution features. This started somewhat informally in Chapter 17, where it was claimed, for example, that the sample mean and the sample variance are usually close to μ and σ^2 of the underlying distribution. Bias and MSE of estimators, discussed in Chapters 19 and 20, are used to judge the quality of estimators. If we have at our disposal an estimator T for an unknown parameter θ, we use its realization t as our estimate for θ. For example, when collecting data on the speed of light, as Michelson did (see Section 13.1), the unknown speed of light would be the parameter θ, our estimator T could be the sample mean, and Michelson's data then yield an estimate t for θ of 299 852.4 km/sec. We call this number a *point estimate*: if we are required to select *one* number, this is it. Had the measurements started a day earlier, however, the whole experiment would in essence be the same, but the results might have been different. Hence, we cannot say that the estimate *equals* the speed of light but rather that it is *close to* the true speed of light. For example, we could say something like: "we have great confidence that the true speed of light is somewhere between ... and ... ." In addition to providing an interval of plausible values for θ we would want to add a specific statement about *how* confident we are that the true θ is among them.

In this chapter we shall present methods to make *confidence statements* about unknown parameters, based on knowledge of the sampling distributions of corresponding estimators. To illustrate the main idea, suppose the estimator T is unbiased for the speed of light θ. For the moment, also suppose that T has standard deviation $\sigma_T = 100$ km/sec (we shall drop this unrealistic assumption shortly). Then, applying formula (13.1), which was derived from Chebyshev's inequality (see Section 13.2), we find

$$\mathrm{P}(|T - \theta| < 2\sigma_T) \geq \tfrac{3}{4}. \tag{23.1}$$

In words this reads: with probability at least 75%, the estimator T is within $2\sigma_T = 200$ of the true speed of light θ. We could rephrase this as

$$T \in (\theta - 200, \theta + 200) \quad \text{with probability at least 75\%.}$$

However, if I am near the city of Paris, then the city of Paris is near me: the statement "T is within 200 of θ" is the same as "θ is within 200 of T," and we could equally well rephrase (23.1) as

$$\theta \in (T - 200, T + 200) \quad \text{with probability at least 75\%.}$$

Note that of the last two equations the first is a statement about a *random variable* T being in a *fixed interval*, whereas in the second equation the *interval is random* and the statement is about the probability that the random interval covers the *fixed* but unknown θ. The interval $(T - 200, T + 200)$ is sometimes called an *interval estimator*, and its realization is an *interval estimate*.

Evaluating T for the Michelson data we find as its realization $t = 299\,852.4$, and this yields the statement

$$\theta \in (299\,652.4, 300\,052.4). \tag{23.2}$$

Because we substituted the realization for the random variable, we cannot claim that (23.2) holds with probability at least 75%: either the true speed of light θ belongs to the interval or it does not; the statement we make is either true or false, we just do not know which. However, because the procedure guarantees a probability of at least 75% of getting a "right" statement, we say:

$$\theta \in (299\,652.4, 300\,052.4) \quad \text{with confidence at least 75\%.} \tag{23.3}$$

The construction of this *confidence interval* only involved an unbiased estimator and knowledge of its standard deviation. When more information on the sampling distribution of the estimator is available, more refined statements can be made, as we shall see shortly.

QUICK EXERCISE 23.1 Repeat the preceding derivation, starting from the statement $\mathrm{P}(|T-\theta| < 3\sigma_T) \geq 8/9$ (check that this follows from Chebyshev's inequality). What is the resulting confidence interval for the speed of light, and what is the corresponding confidence?

A general definition

Many confidence intervals are of the form[1]

$$(t - c \cdot \sigma_T,\ t + c \cdot \sigma_T)$$

we just encountered, where c is a number near 2 or 3. The corresponding confidence is often much higher than in the preceding example. Because there are many other ways confidence intervals can (or have to) be constructed, the general definition looks a bit different.

CONFIDENCE INTERVALS. Suppose a dataset $x_1, \ldots, x_n$ is given, modeled as realization of random variables $X_1, \ldots, X_n$. Let θ be the parameter of interest, and γ a number between 0 and 1. If there exist sample statistics $L_n = g(X_1, \ldots, X_n)$ and $U_n = h(X_1, \ldots, X_n)$ such that

$$\mathrm{P}(L_n < \theta < U_n) = \gamma$$

for every value of θ, then

$$(l_n, u_n),$$

where $l_n = g(x_1, \ldots, x_n)$ and $u_n = h(x_1, \ldots, x_n)$, is called a $100\gamma\%$ *confidence interval for* θ. The number γ is called the *confidence level.*

Sometimes sample statistics L_n and U_n as required in the definition do not exist, but one *can* find L_n and U_n that satisfy

$$\mathrm{P}(L_n < \theta < U_n) \geq \gamma.$$

The resulting confidence interval (l_n, u_n) is called a *conservative* $100\gamma\%$ confidence interval for θ: the actual confidence level might be higher. For example, the interval in (23.2) is a conservative 75% confidence interval.

QUICK EXERCISE 23.2 Why is the interval in (23.2) a *conservative* 75% confidence interval?

There is no way of knowing whether an individual confidence interval is correct, in the sense that it indeed *does* cover θ. The procedure guarantees that each time we make a confidence interval we have probability γ of covering θ. What this means in practice can easily be illustrated with an example, using simulation:

[1] Another form is, for example, $(c_1 t,\ c_2 t)$.

Generate $x_1, \ldots, x_{20}$ from an $N(0,1)$ distribution. Next, pretend that it is known that the data are from a normal distribution but that both μ and σ are unknown. Construct the 90% confidence interval for the expectation μ using the method described in the next section, which says to use (l_n, u_n) with

$$l_n = \bar{x}_{20} - 1.729\frac{s_{20}}{\sqrt{20}} \qquad u_n = \bar{x}_{20} + 1.729\frac{s_{20}}{\sqrt{20}},$$

where $\bar{x}_{20}$ and s_{20} are the sample mean and standard deviation. Finally, check whether the "true μ," in this case 0, is in the confidence interval.

We repeated the whole procedure 50 times, making 50 confidence intervals for μ. Each confidence interval is based on a fresh independently generated set of data. The 50 intervals are plotted in Figure 23.1 as horizontal line

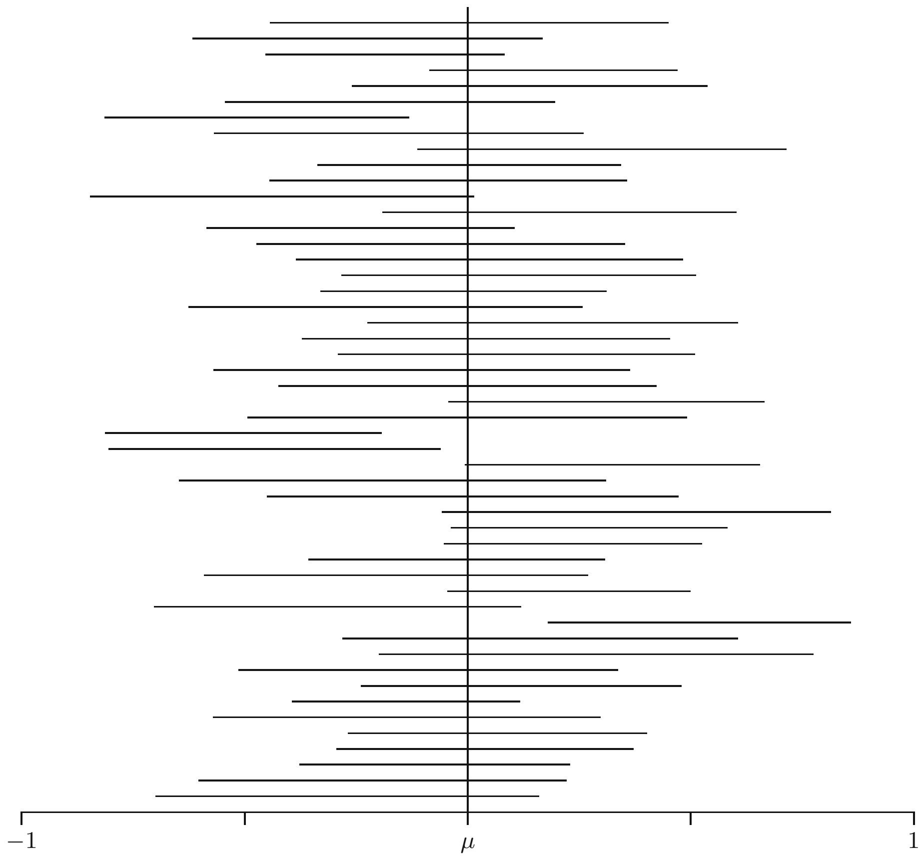

Fig. 23.1. Fifty 90% confidence intervals for $\mu = 0$.

segments, and at μ (0!) a vertical line is drawn. We count 46 "hits": only four intervals do not contain the true μ.

QUICK EXERCISE 23.3 Suppose you were to make 40 confidence intervals with confidence level 95%. About how many of them should you expect to be "wrong"? Should you be surprised if 10 of them are wrong?

In the remainder of this chapter we consider *confidence intervals for the mean*: confidence intervals for the unknown expectation μ of the distribution from which the sample originates. We start with the situation where it is known that the data originate from a normal distribution, first with known variance, then with unknown variance. Then we drop the normal assumption, first use the bootstrap, and finally show how, for very large samples, confidence intervals based on the central limit theorem are made.

23.2 Normal data

Suppose the data can be seen as the realization of a sample $X_1, \ldots, X_n$ from an $N(\mu, \sigma^2)$ distribution and μ is the (unknown) parameter of interest. If the variance σ^2 is known, confidence intervals are easily derived. Before we do this, some preparation has to be done.

Critical values

We shall need so-called critical values for the standard normal distribution. The *critical value* z_p of an $N(0,1)$ distribution is the number that has right tail probability p. It is defined by

$$\mathrm{P}(Z \geq z_p) = p,$$

where Z is an $N(0,1)$ random variable. For example, from Table B.1 we read $\mathrm{P}(Z \geq 1.96) = 0.025$, so $z_{0.025} = 1.96$. In fact, z_p is the $(1-p)$th quantile of the standard normal distribution:

$$\Phi(z_p) = \mathrm{P}(Z \leq z_p) = 1 - p.$$

By the symmetry of the standard normal density, $\mathrm{P}(Z \leq -z_p) = \mathrm{P}(Z \geq z_p) = p$, so $\mathrm{P}(Z \geq -z_p) = 1 - p$ and therefore

$$z_{1-p} = -z_p.$$

For example, $z_{0.975} = -z_{0.025} = -1.96$. All this is illustrated in Figure 23.2.

QUICK EXERCISE 23.4 Determine $z_{0.01}$ and $z_{0.95}$ from Table B.1.

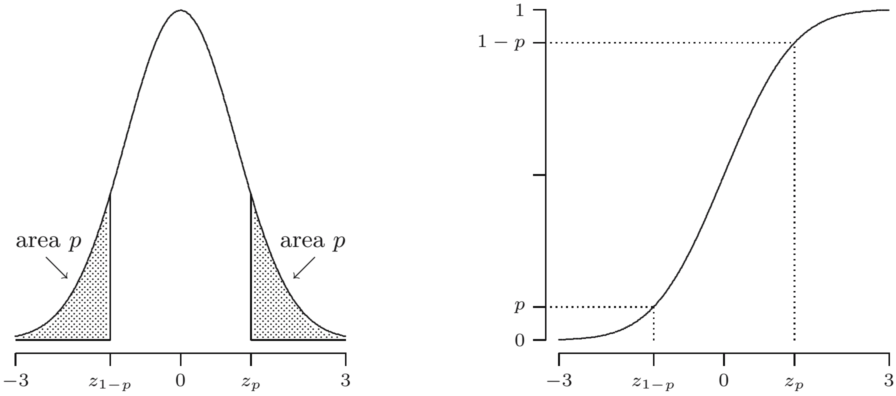

Fig. 23.2. Critical values of the standard normal distribution.

Variance known

If $X_1, \ldots, X_n$ is a random sample from an $N(\mu, \sigma^2)$ distribution, then $\bar{X}_n$ has an $N(\mu, \sigma^2/n)$ distribution, and from the properties of the normal distribution (see page 106), we know that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad \text{has an } N(0,1) \text{ distribution.}$$

If c_l and c_u are chosen such that $\mathrm{P}(c_l < Z < c_u) = \gamma$ for an $N(0,1)$ distributed random variable Z, then

$$\begin{aligned} \gamma &= \mathrm{P}\left(c_l < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < c_u \right) \\ &= \mathrm{P}\left(c_l \frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < c_u \frac{\sigma}{\sqrt{n}} \right) \\ &= \mathrm{P}\left(\bar{X}_n - c_u \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n - c_l \frac{\sigma}{\sqrt{n}} \right). \end{aligned}$$

We have found that

$$L_n = \bar{X}_n - c_u \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad U_n = \bar{X}_n - c_l \frac{\sigma}{\sqrt{n}}$$

satisfy the confidence interval definition: the interval (L_n, U_n) covers μ with probability γ. Therefore

$$\left(\bar{x}_n - c_u \frac{\sigma}{\sqrt{n}},\ \bar{x}_n - c_l \frac{\sigma}{\sqrt{n}} \right)$$

is a $100\gamma\%$ confidence interval for μ. A common choice is to divide $\alpha = 1 - \gamma$ evenly between the tails,[2] that is, solve c_l and c_u from

[2] Here this choice could be motivated by the fact that it leads to the shortest confidence interval; in other examples the shortest interval requires an *asymmetric*

$$\mathrm{P}(Z \geq c_u) = \alpha/2 \quad \text{and} \quad \mathrm{P}(Z \leq c_l) = \alpha/2,$$

so that $c_u = z_{\alpha/2}$ and $c_l = z_{1-\alpha/2} = -z_{\alpha/2}$. Summarizing, the $100(1-\alpha)\%$ confidence interval for μ is:

$$\left(\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

For example, if $\alpha = 0.05$, we use $z_{0.025} = 1.96$ and the 95% confidence interval is

$$\left(\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}},\, \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right).$$

Example: gross calorific content of coal

When a shipment of coal is traded, a number of its properties should be known accurately, because the value of the shipment is determined by them. An important example is the so-called gross calorific value, which characterizes the heat content and is a numerical value in megajoules per kilogram (MJ/kg). The International Organization of Standardization (ISO) issues standard procedures for the determination of these properties. For the gross calorific value, there is a method known as ISO 1928. When the procedure is carried out properly, resulting measurement errors are known to be approximately normal, with a standard deviation of about 0.1 MJ/kg. Laboratories that operate according to standard procedures receive ISO certificates. In Table 23.1, a number of such ISO 1928 measurements is given for a shipment of Osterfeld coal coded 262DE27.

Table 23.1. Gross calorific value measurements for Osterfeld 262DE27.

23.870	23.730	23.712	23.760	23.640	23.850	23.840	23.860
23.940	23.830	23.877	23.700	23.796	23.727	23.778	23.740
23.890	23.780	23.678	23.771	23.860	23.690	23.800	

Source: A.M.H. van der Veen and A.J.M. Broos. Interlaboratory study programme "ILS coal characterization"—reported data. Technical report, NMi Van Swinden Laboratorium B.V., The Netherlands, 1996.

We want to combine these values into a confidence statement about the "true" gross calorific content of Osterfeld 262DE27. From the data, we compute $\bar{x}_n = 23.788$. Using the given $\sigma = 0.1$ and $\alpha = 0.05$, we find the 95% confidence interval

$$\left(23.788 - 1.96\frac{0.1}{\sqrt{23}},\, 23.788 + 1.96\frac{0.1}{\sqrt{23}}\right) = (23.747, 23.829) \quad \text{MJ/kg}.$$

division of α. If you are only concerned with the left or right boundary of the confidence interval, see the next chapter.

Variance unknown

When σ is unknown, the fact that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

has a standard normal distribution has become useless, as it involves this unknown σ, which would subsequently appear in the confidence interval. However, if we substitute the estimator S_n for σ, the resulting random variable

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

has a distribution that only depends on n and *not* on μ or σ. Moreover, its density can be given explicitly.

DEFINITION. A continuous random variable has a *t-distribution with parameter m*, where $m \geq 1$ is an integer, if its probability density is given by

$$f(x) = k_m \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}} \qquad \text{for } -\infty < x < \infty,$$

where $k_m = \Gamma\left(\frac{m+1}{2}\right) / \left(\Gamma\left(\frac{m}{2}\right)\sqrt{m\pi}\right)$. This distribution is denoted by $t(m)$ and is referred to as the t-distribution with m *degrees of freedom.*

The normalizing constant k_m is given in terms of the gamma function, which was defined on page 157. For $m = 1$, it evaluates to $k_1 = 1/\pi$, and the resulting density is that of the standard Cauchy distribution (see page 161). If X has a $t(m)$ distribution, then $\mathrm{E}[X] = 0$ for $m \geq 2$ and $\mathrm{Var}(X) = m/(m-2)$ for $m \geq 3$. Densities of t-distributions look like that of the standard normal distribution: they are also symmetric around 0 and bell-shaped. As m goes to infinity the limit of the $t(m)$ density is the standard normal density. The distinguishing feature is that densities of t-distributions have heavier tails: $f(x)$ goes to zero as x goes to $+\infty$ or $-\infty$, but more slowly than the density $\phi(x)$ of the standard normal distribution. These properties are illustrated in Figure 23.3, which shows the densities and distribution functions of the $t(1)$, $t(2)$, and $t(5)$ distribution as well as those of the standard normal.

We will also need critical values for the $t(m)$ distribution: the critical value $t_{m,p}$ is the number satisfying

$$\mathrm{P}(T \geq t_{m,p}) = p,$$

where T is a $t(m)$ distributed random variable. Because the t-distribution is symmetric around zero, using the same reasoning as for the critical values of the standard normal distribution, we find:

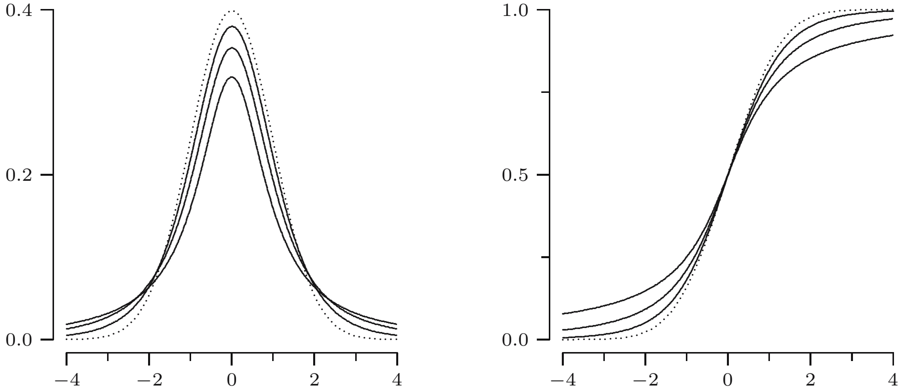

Fig. 23.3. Three t-distributions and the standard normal distribution. The dotted line corresponds to the standard normal. The other distributions depicted are the $t(1)$, $t(2)$, and $t(5)$, which in that order resemble the standard normal more and more.

$$t_{m,1-p} = -t_{m,p}.$$

For example, in Table B.2 we read $t_{10,0.01} = 2.764$, and from this we deduce that $t_{10,0.99} = -2.764$.

Quick exercise 23.5 Determine $t_{3,0.01}$ and $t_{35,0.9975}$ from Table B.2.

We now return to the distribution of

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

and construct a confidence interval for μ.

The studentized mean of a normal random sample. For a random sample $X_1, \ldots, X_n$ from an $N(\mu, \sigma^2)$ distribution, the *studentized mean*

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

has a $t(n-1)$ distribution, regardless of the values of μ and σ.

From this fact and using critical values of the t-distribution, we derive that

$$\mathrm{P}\left(-t_{n-1,\alpha/2} < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < t_{n-1,\alpha/2}\right) = 1 - \alpha, \tag{23.4}$$

and in the same way as when σ is known it now follows that a $100(1-\alpha)\%$ confidence interval for μ is given by:

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right).$$

Returning to the coal example, there was another shipment, of Daw Mill 258GB41 coal, where there were actually some doubts whether the stated accuracy of the ISO 1928 method was attained. We therefore prefer to consider σ unknown and estimate it from the data, which are given in Table 23.2.

Table 23.2. Gross calorific value measurements for Daw Mill 258GB41.

30.990	31.030	31.060	30.921	30.920	30.990	31.024	30.929
31.050	30.991	31.208	30.830	31.330	30.810	31.060	30.800
31.091	31.170	31.026	31.020	30.880	31.125		

Source: A.M.H. van der Veen and A.J.M. Broos. Interlaboratory study programme "ILS coal characterization"—reported data. Technical report, NMi Van Swinden Laboratorium B.V., The Netherlands, 1996.

Doing this, we find $\bar{x}_n = 31.012$ and $s_n = 0.1294$. Because $n = 22$, for a 95% confidence interval we use $t_{21,0.025} = 2.080$ and obtain

$$\left(31.012 - 2.080\frac{0.1294}{\sqrt{22}},\ 31.012 + 2.080\frac{0.1294}{\sqrt{22}}\right) = (30.954, 31.069).$$

Note that this confidence interval is (50%!) wider than the one we made for the Osterfeld coal, with almost the same sample size. There are two reasons for this; one is that $\sigma = 0.1$ is replaced by the (larger) estimate $s_n = 0.1294$, and the second is that the critical value $z_{0.025} = 1.96$ is replaced by the larger $t_{21,0.025} = 2.080$. The differences in the method and the ingredients seem minor, but they matter, especially for small samples.

23.3 Bootstrap confidence intervals

It is not uncommon that the methods of the previous section are used even when the normal distribution is *not* a good model for the data. In some cases this is not a big problem: with small deviations from normality the actual confidence level of a constructed confidence interval may deviate only a few percent from the intended confidence level. For large datasets the central limit theorem in fact ensures that this method provides confidence intervals with approximately correct confidence levels, as we shall see in the next section.

If we doubt the normality of the data and we do *not* have a large sample, usually the best thing to do is to bootstrap. Suppose we have a dataset $x_1, \ldots, x_n$, modeled as a realization of a random sample from some distribution F, and we want to construct a confidence interval for its (unknown) expectation μ.

In the previous section we saw that it suffices to find numbers c_l and c_u such that

$$\mathrm{P}\left(c_l < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < c_u\right) = 1 - \alpha.$$

The $100(1-\alpha)\%$ confidence interval would then be

$$\left(\bar{x}_n - c_u\frac{s_n}{\sqrt{n}}, \bar{x}_n - c_l\frac{s_n}{\sqrt{n}}\right),$$

where, of course, $\bar{x}_n$ and s_n are the sample mean and the sample standard deviation. To find c_l and c_u we need to know the distribution of the studentized mean

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}.$$

We apply the bootstrap principle. From the data $x_1, \ldots, x_n$ we determine an estimate $\hat{F}$ of F. Let $X_1^*, \ldots, X_n^*$ be a random sample from $\hat{F}$, with $\mu^* = \mathrm{E}\left[X_i^*\right]$, and consider

$$T^* = \frac{\bar{X}_n^* - \mu^*}{S_n^*/\sqrt{n}}.$$

The distribution of T^* is now used as an approximation to the distribution of T. If we use $\hat{F} = F_n$, we get the following.

EMPIRICAL BOOTSTRAP SIMULATION FOR THE STUDENTIZED MEAN. Given a dataset $x_1, x_2, \ldots, x_n$, determine its empirical distribution function F_n as an estimate of F. The expectation corresponding to F_n is $\mu^* = \bar{x}_n$.

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from F_n.
2. Compute the studentized mean for the bootstrap dataset:

$$t^* = \frac{\bar{x}_n^* - \bar{x}_n}{s_n^*/\sqrt{n}},$$

where $\bar{x}_n^*$ and s_n^* are the sample mean and sample standard deviation of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 many times.

From the bootstrap experiment we can determine c_l^* and c_u^* such that

$$\mathrm{P}\left(c_l^* < \frac{\bar{X}_n^* - \mu^*}{S_n^*/\sqrt{n}} < c_u^*\right) \approx 1 - \alpha.$$

By the bootstrap principle we may transfer this statement about the distribution of T^* to the distribution of T. That is, we may use these estimated critical values as bootstrap approximations to c_l and c_u:

$$c_l \approx c_l^* \quad \text{and} \quad c_u \approx c_u^*,$$

Therefore, we call

$$\left(\bar{x}_n - c_u^* \frac{s_n}{\sqrt{n}},\ \bar{x}_n - c_l^* \frac{s_n}{\sqrt{n}}\right)$$

a $100(1-\alpha)\%$ *bootstrap confidence interval for* μ.

Example: the software data

Recall the software data, a dataset of interfailure times (see Section 17.3). From the nature of the data—failure times are positive numbers—and the histogram (Figure 17.5), we know that they should not be modeled as a realization of a random sample from a normal distribution. From the data we know $\bar{x}_n = 656.88$, $s_n = 1037.3$, and $n = 135$. We generate one thousand bootstrap datasets, and for each dataset we compute t^* as in step 2 of the procedure. The histogram and empirical distribution function made from these one thousand values are estimates of the density and the distribution function, respectively, of the bootstrap sample statistic T^*; see Figure 23.4.

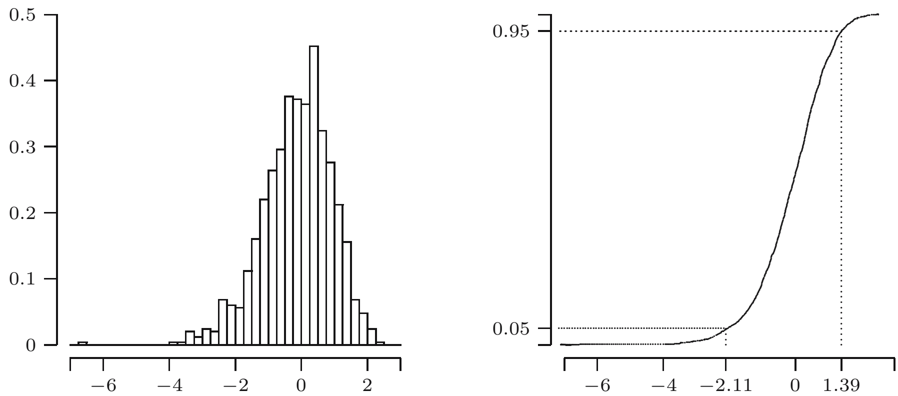

Fig. 23.4. Histogram and empirical distribution function of the studentized bootstrap simulation results for the software data.

We want to make a 90% bootstrap confidence interval, so we need c_l^* and c_u^*, or the 0.05th and 0.95th quantile from the empirical distribution function in Figure 23.4. The 50th order statistic of the one thousand t^* values is -2.107. This means that 50 out of the one thousand values, or 5%, are smaller than or equal to this value, and so $c_l^* = -2.107$. Similarly, from the 951st order statistic, 1.389, we obtain[3] $c_u^* = 1.389$. Inserting these values, we find the following 90% bootstrap confidence interval for μ:

[3] These results deviate slightly from the definition of empirical quantiles as given in Section 16.3. That method is a little more accurate.

$$\left(656.88 - 1.389\frac{1037.3}{\sqrt{135}},\ 656.88 - (-2.107)\frac{1037.3}{\sqrt{135}}\right) = (532.9, 845.0).$$

QUICK EXERCISE 23.6 The 25th and 976th order statistic from the preceding bootstrap results are -2.443 and 1.713, respectively. Use these numbers to construct a confidence interval for μ. What is the corresponding confidence level?

Why the bootstrap may be better

The reason to use the bootstrap is that it should lead to a more accurate approximation of the distribution of the studentized mean than the $t(n-1)$ distribution that follows from *assuming* normality. If, in the previous example, we would think we had normal data, we would use critical values from the $t(134)$ distribution: $t_{134,0.05} = 1.656$. The result would be

$$\left(656.88 - 1.656\frac{1037.3}{\sqrt{135}},\ 656.88 + 1.656\frac{1037.3}{\sqrt{135}}\right) = (509.0, 804.7).$$

Comparing the intervals, we see that here the bootstrap interval is a little larger and, as opposed to the t-interval, not centered around the sample mean but *skewed* to the right side. This is one of the features of the bootstrap: if the distribution from which the data originate is *skewed*, this is reflected in the confidence interval. Looking at the histogram of the software data (Figure 17.5), we see that is it skewed to the right: it has a long tail on the right, but not on the left, so the same most likely holds for the distribution from which these data originate. The skewness is reflected in the confidence interval, which extends more to the right of $\bar{x}_n$ than to the left. In some sense, the bootstrap adapts to the shape of the distribution, and in this way it leads to more accurate confidence statements than using the method for normal data. What we mean by this is that, for example, with the normal method only 90% of the 95% confidence statements would actually cover the true value, whereas for the bootstrap intervals this percentage would be close(r) to 95%.

23.4 Large samples

A variant of the central limit theorem states that as n goes to infinity, the distribution of the studentized mean

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

approaches the standard normal distribution. This fact is the basis for so-called *large sample confidence intervals*. Suppose $X_1, \ldots, X_n$ is a random sample from some distribution F with expectation μ. If n is large enough, we may use

$$\mathrm{P}\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha. \tag{23.5}$$

This implies that if $x_1, \ldots, x_n$ can be seen as a realization of a random sample from some unknown distribution with expectation μ and if n is large enough, then

$$\left(\bar{x}_n - z_{\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{s_n}{\sqrt{n}}\right)$$

is an approximate $100(1-\alpha)\%$ confidence interval for μ.

Just as earlier with the central limit theorem, a key question is "how big should n be?" Again, there is no easy answer. To give you some idea, we have listed in Table 23.3 the results of a small simulation experiment. For each of the distributions, sample sizes, and confidence levels listed, we constructed 10 000 confidence intervals with the large sample method; the numbers listed in the table are the confidence levels as estimated from the simulation, the *coverage probabilities*. The chosen Pareto distribution is very skewed, and this shows; the coverage probabilities for the exponential are just a few percent off.

Table 23.3. Estimated coverage probabilities for large sample confidence intervals for non-normal data.

		γ	
Distribution	n	0.900	0.950
$Exp(1)$	20	0.851	0.899
$Exp(1)$	100	0.890	0.938
$Par(2.1)$	20	0.727	0.774
$Par(2.1)$	100	0.798	0.849

In the case of simulation one can often quite easily generate a very large number of independent repetitions, and then this question poses no problem. In other cases there may be nothing better to do than hope that the dataset is large enough. We give an example where (we believe!) this is definitely the case.

In an article published in 1910 ([28]), Rutherford and Geiger reported their observations on the radioactive decay of the element polonium. Using a small disk coated with polonium they counted the number of emitted alpha-particles during 2608 intervals of 7.5 seconds each. The dataset consists of the counted number of alpha-particles for each of the 2608 intervals and can be summarized as in Table 23.4.

Table 23.4. Alpha-particle counts for 2608 intervals of 7.5 seconds.

Count	0	1	2	3	4
Frequency	57	203	383	525	532
Count	5	6	7	8	9
Frequency	408	273	139	45	27
Count	10	11	12	13	14
Frequency	10	4	0	1	1

Source: E. Rutherford and H. Geiger (with a note by H. Bateman), The probability variations in the distribution of α particles, *Phil.Mag.*, 6: 698–704, 1910; the table on page 701.

The total number of counted alpha-particles is 10 097, the average number per interval is therefore 3.8715. The sample standard deviation can also be computed from the table; it is 1.9225. So we know of the actual data $x_1, x_2, \ldots, x_{2608}$ (where the counts x_i are between 0 and 14) that $\bar{x}_n = 3.8715$ and $s_n = 1.9225$. We construct a 98% confidence interval for the expected number of particles per interval. As $z_{0.01} = 2.33$ this results in

$$\left(3.8715 - 2.33\frac{1.9225}{\sqrt{2608}},\ 3.8715 + 2.33\frac{1.9225}{\sqrt{2608}}\right) = (3.784,\ 3.959).$$

23.5 Solutions to the quick exercises

23.1 From the probability statement, we derive, using $\sigma_T = 100$ and $8/9 = 0.889$:

$$\theta \in (T - 300,\ T + 300) \quad \text{with probability at least 88\%.}$$

With $t = 299\,852.4$, this becomes

$$\theta \in (299\,552.4,\ 300\,152.4) \quad \text{with confidence at least 88\%.}$$

23.2 Chebyshev's inequality only gives an upper bound. The actual value of $\mathrm{P}(|T - \theta| < 2\sigma_T)$ could be higher than 3/4, depending on the distribution of T. For example, in Quick exercise 13.2 we saw that in case of an exponential distribution this probability is 0.865. For other distributions, even higher values are attained; see Exercise 13.1.

23.3 For each of the confidence intervals we have a 5% probability that it is wrong. Therefore, the number of wrong confidence intervals has a $Bin(40, 0.05)$ distribution, and we would expect about $40 \cdot 0.05 = 2$ to be wrong. The standard deviation of this distribution is $\sqrt{40 \cdot 0.05 \cdot 0.95} = 1.38$. The outcome "10 confidence intervals wrong" is $(10 - 2)/1.38 = 5.8$ standard deviations from the expectation and would be a surprising outcome indeed. (The probability of 10 or more wrong is 0.00002.)

23.4 We need to solve $P(Z \geq a) = 0.01$. In Table B.1 we find $P(Z \geq 2.33) = 0.0099 \approx 0.01$, so $z_{0.01} \approx 2.33$. For $z_{0.95}$ we need to solve $P(Z \geq a) = 0.95$, and because this is in the left tail of the distribution, we use $z_{0.95} = -z_{0.05}$. In the table we read $P(Z \geq 1.64) = 0.0505$ and $P(Z \geq 1.65) = 0.0495$, from which we conclude $z_{0.05} \approx (1.64 + 1.65)/2 = 1.645$ and $z_{0.95} \approx -1.645$.

23.5 In Table B.1 we find $P(T_3 \geq 4.541) = 0.01$, so $t_{3,0.01} = 4.541$. For $t_{35,0.9975}$, we need to use $t_{35,0.9975} = -t_{35,0.0025}$. In the table we find $t_{30,0.0025} = 3.030$ and $t_{40,0.0025} = 2.971$, and by interpolation $t_{35,0.0025} \approx (3.030 + 2.971)/2 = 3.0005$. Hence, $t_{35,0.9975} \approx -3.000$.

23.6 The order statistics are estimates for $c^*_{0.025}$ and $c^*_{0.975}$, respectively. So the corresponding α is 0.05, and the 95% bootstrap confidence interval for μ is:

$$\left(656.88 - 1.713\frac{1037.3}{\sqrt{135}}, 656.88 - (-2.443)\frac{1037.3}{\sqrt{135}}\right) = (504.0, 875.0).$$

23.6 Exercises

23.1 ⊡ A bottling machine is known to fill wine bottles with amounts that follow an $N(\mu, \sigma^2)$ distribution, with $\sigma = 5$ (ml). In a sample of 16 bottles, $\bar{x} = 743$ (ml) was found. Construct a 95% confidence interval for μ.

23.2 ⊡ You are given a dataset that may be considered a realization of a normal random sample. The size of the dataset is 34, the average is 3.54, and the sample standard deviation is 0.13. Construct a 98% confidence interval for the unknown expectation μ.

23.3 You have ordered 10 bags of cement, which are supposed to weigh 94 kg each. The average weight of the 10 bags is 93.5 kg. Assuming that the 10 weights can be viewed as a realization of a random sample from a normal distribution with unknown parameters, construct a 95% confidence interval for the expected weight of a bag. The sample standard deviation of the 10 weights is 0.75.

23.4 A new type of car tire is launched by a tire manufacturer. The automobile association performs a durability test on a random sample of 18 of these tires. For each tire the durability is expressed as a percentage: a score of 100 (%) means that the tire lasted exactly as long as the average standard tire, an accepted comparison standard. From the multitude of factors that influence the durability of individual tires the assumption is warranted that the durability of an arbitrary tire follows an $N(\mu, \sigma^2)$ distribution. The parameters μ and σ^2 characterize the tire *type*, and μ could be called the durability index for this type of tire. The automobile association found for the tested tires: $\bar{x}_{18} = 195.3$ and $s_{18} = 16.7$. Construct a 95% confidence interval for μ.

23.5 ⊞ During the 2002 Winter Olympic Games in Salt Lake City a newspaper article mentioned the alleged advantage speed-skaters have in the 1500 m race if they start in the outer lane. In the men's 1500 m, there were 24 races, but in race 13 (really!) someone fell and did not finish. The results in seconds of the remaining 23 races are listed in Table 23.5. You should know that who races against whom, in which race, and who starts in the outer lane are all determined by a fair lottery.

Table 23.5. Speed-skating results in seconds, men's 1500 m (except race 13), 2002 Winter Olympic Games.

Race number	Inner lane	Outer lane	Difference
1	107.04	105.98	1.06
2	109.24	108.20	1.04
3	111.02	108.40	2.62
4	108.02	108.58	−0.56
5	107.83	105.51	2.32
6	109.50	112.01	−2.51
7	111.81	112.87	−1.06
8	111.02	106.40	4.62
9	106.04	104.57	1.47
10	110.15	110.70	−0.55
11	109.42	109.45	−0.03
12	108.13	109.57	−1.44
14	105.86	105.97	−0.11
15	108.27	105.63	2.64
16	107.63	105.41	2.22
17	107.72	110.26	−2.54
18	106.38	105.82	0.56
19	107.78	106.29	1.49
20	108.57	107.26	1.31
21	106.99	103.95	3.04
22	107.21	106.00	1.21
23	105.34	105.26	0.08
24	108.76	106.75	2.01
Mean	108.25	107.43	0.82
St.dev.	1.70	2.42	1.78

a. As a consequence of the lottery and the fact that many different factors contribute to the actual time difference "inner lane minus outer lane" the assumption of a normal distribution for the difference is warranted. The numbers in the last column can be seen as realizations from an $N(\delta, \sigma^2)$

distribution, where δ is the expected outer lane advantage. Construct a 95% confidence interval for δ. N.B. $n = 23$, not 24!

b. You decide to make a bootstrap confidence interval instead. Describe the appropriate bootstrap experiment.

c. The bootstrap experiment was performed with one thousand repetitions. Part of the bootstrap outcomes are listed in the following table. From the *ordered* list of results, numbers 21 to 60 and 941 to 980 are given. Use these to construct a 95% bootstrap confidence interval for δ.

21–25	−2.202	−2.164	−2.111	−2.109	−2.101
26–30	−2.099	−2.006	−1.985	−1.967	−1.929
31–35	−1.917	−1.898	−1.864	−1.830	−1.808
36–40	−1.800	−1.799	−1.774	−1.773	−1.756
41–45	−1.736	−1.732	−1.731	−1.717	−1.716
46–50	−1.699	−1.692	−1.691	−1.683	−1.666
51–55	−1.661	−1.644	−1.638	−1.637	−1.620
56–60	−1.611	−1.611	−1.601	−1.600	−1.593
941–945	1.648	1.667	1.669	1.689	1.696
946–950	1.708	1.722	1.726	1.735	1.814
951–955	1.816	1.825	1.856	1.862	1.864
956–960	1.875	1.877	1.897	1.905	1.917
961–965	1.923	1.948	1.961	1.987	2.001
966–970	2.015	2.015	2.017	2.018	2.034
971–975	2.035	2.037	2.039	2.053	2.060
976–980	2.088	2.092	2.101	2.129	2.143

23.6 ⊞ A dataset $x_1, x_2, \ldots, x_n$ is given, modeled as realization of a sample $X_1, X_2, \ldots, X_n$ from an $N(\mu, 1)$ distribution. Suppose there are sample statistics $L_n = g(X_1, \ldots, X_n)$ and $U_n = h(X_1, \ldots, X_n)$ such that

$$\mathrm{P}(L_n < \mu < U_n) = 0.95$$

for every value of μ. Suppose that the corresponding 95% confidence interval derived from the data is $(l_n, u_n) = (-2, 5)$.

a. Suppose $\theta = 3\mu + 7$. Let $\tilde{L}_n = 3L_n + 7$ and $\tilde{U}_n = 3U_n + 7$. Show that $\mathrm{P}\left(\tilde{L}_n < \theta < \tilde{U}_n\right) = 0.95$.

b. Write the 95% confidence interval for θ in terms of l_n and u_n.

c. Suppose $\theta = 1 - \mu$. Again, find $\tilde{L}_n$ and $\tilde{U}_n$, as well as the confidence interval for θ.

d. Suppose $\theta = \mu^2$. Can you construct a confidence interval for θ?

23.7 ⊡ A 95% confidence interval for the parameter μ of a $Pois(\mu)$ distribution is given: $(2, 3)$. Let X be a random variable with this distribution. Construct a 95% confidence interval for $\mathrm{P}(X = 0) = \mathrm{e}^{-\mu}$.

23.8 Suppose that in Exercise 23.1 the content of the bottles has to be determined by weighing. It is known that the wine bottles involved weigh on average 250 grams, with a standard deviation of 15 grams, and the weights follow a normal distribution. For a sample of 16 bottles, an average weight of 998 grams was found. You may assume that 1 ml of wine weighs 1 gram, and that the filling amount is independent of the bottle weight. Construct a 95% confidence interval for the expected amount of wine per bottle, μ.

23.9 Consider the alpha-particle counts discussed in Section 23.4; the data are given in Table 23.4. We want to bootstrap in order to make a bootstrap confidence interval for the expected number of particles in a 7.5-second interval.

a. Describe in detail how you would perform the bootstrap simulation.

b. The bootstrap experiment was performed with one thousand repetitions. Part of the (ordered) bootstrap t^*'s are given in the following table. Construct the 95% bootstrap confidence interval for the expected number of particles in a 7.5-second interval.

1–5	−2.996	−2.942	−2.831	−2.663	−2.570
6–10	−2.537	−2.505	−2.290	−2.273	−2.228
11–15	−2.193	−2.112	−2.092	−2.086	−2.045
16–20	−1.983	−1.980	−1.978	−1.950	−1.931
21–25	−1.920	−1.910	−1.893	−1.889	−1.888
26–30	−1.865	−1.864	−1.832	−1.817	−1.815
31–35	−1.755	−1.751	−1.749	−1.746	−1.744
36–40	−1.734	−1.723	−1.710	−1.708	−1.705
41–45	−1.703	−1.700	−1.696	−1.692	−1.691
46–50	−1.691	−1.675	−1.660	−1.656	−1.650
951–955	1.635	1.638	1.643	1.648	1.661
956–960	1.666	1.668	1.678	1.681	1.686
961–965	1.692	1.719	1.721	1.753	1.772
966–970	1.773	1.777	1.806	1.814	1.821
971–975	1.824	1.826	1.837	1.838	1.845
976–980	1.862	1.877	1.881	1.883	1.956
981–985	1.971	1.992	2.060	2.063	2.083
986–990	2.089	2.177	2.181	2.186	2.224
991–995	2.234	2.264	2.273	2.310	2.348
996–1000	2.483	2.556	2.870	2.890	3.546

c. Answer this without doing any calculations: if we made the 98% bootstrap confidence interval, would it be smaller or larger than the interval constructed in Section 23.4?

23.10 In a report you encounter a 95% confidence interval $(1.6, 7.8)$ for the parameter μ of an $N(\mu, \sigma^2)$ distribution. The interval is based on 16 observations, constructed according to the studentized mean procedure.

a. What is the mean of the (unknown) dataset?

b. You prefer to have a 99% confidence interval for μ. Construct it.

23.11 ⊞ A 95% confidence interval for the unknown expectation of some distribution contains the number 0.

a. We construct the corresponding 98% confidence interval, using the same data. Will it contain the number 0?

b. The confidence interval in fact is a bootstrap confidence interval. We repeat the bootstrap experiment (using the same data) and construct a new 95% confidence interval based on the results. Will it contain the number 0?

c. We collect new data, resulting in a dataset of the same size. With this data, we construct a 95% confidence interval for the unknown expectation. Will the interval contain 0?

23.12 Let $Z_1, \ldots, Z_n$ be a random sample from an $N(0, 1)$ distribution. Define $X_i = \mu + \sigma Z_i$ for $i = 1, \ldots, n$ and $\sigma > 0$. Let $\bar{Z}$, $\bar{X}$ denote the sample averages and S_Z and S_X the sample standard deviations, of the Z_i and X_i, respectively.

a. Show that $X_1, \ldots, X_n$ is a random sample from an $N(\mu, \sigma^2)$ distribution.

b. Express $\bar{X}$ and S_X in terms of $\bar{Z}$, S_Z, μ, and σ.

c. Verify that

$$\frac{\bar{X} - \mu}{S_X/\sqrt{n}} = \frac{\bar{Z}}{S_Z/\sqrt{n}},$$

and explain why this shows that the distribution of the studentized mean does not depend on μ and σ.