18

The bootstrap

In the forthcoming chapters we will develop statistical methods to infer knowledge about the model distribution and encounter several sample statistics to do this. In the previous chapter we have seen examples of sample statistics that can be used to estimate different model features, for instance, the empirical distribution function to estimate the model distribution function F, and the sample mean to estimate the expectation μ corresponding to F. One of the things we would like to know is how close a sample statistic is to the model feature it is supposed to estimate. For instance, what is the probability that the sample mean and μ differ more than a given tolerance ε? For this we need to know the distribution of $\bar{X}_n - \mu$. More generally, it is important to know how a sample statistic is distributed in relation to the corresponding model feature. For the distribution of the sample mean we saw a normal *limit* approximation in Chapter 14. In this chapter we discuss a simulation procedure that approximates the distribution of the sample mean for *finite* sample size. Moreover, the method is more generally applicable to sample statistics other than the sample mean.

18.1 The bootstrap principle

Consider the Old Faithful data introduced in Chapter 15, which we modeled as the realization of a random sample of size $n = 272$ from some distribution function F. The sample mean $\bar{x}_n$ of the observed durations equals 209.3. What does this say about the expectation μ of F? As we saw in Chapter 17, the value 209.3 is a natural estimate for μ, but to conclude that μ is *equal* to 209.3 is unwise. The reason is that, if we would observe a new dataset of durations, we will obtain a different sample mean as an estimate for μ. This should not come as a surprise. Since the dataset $x_1, x_2, \ldots, x_n$ is just one possible realization of the random sample $X_1, X_2, \ldots, X_n$, the observed sample mean is just one possible realization of the random variable

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

A new dataset is another realization of the random sample, and the corresponding sample mean is another realization of the random variable $\bar{X}_n$. Hence, to infer something about μ, one should take into account how realizations of $\bar{X}_n$ vary. This variation is described by the probability distribution of $\bar{X}_n$.

In principle[1] it is possible to determine the distribution function of $\bar{X}_n$ from the distribution function F of the random sample $X_1, X_2, \ldots, X_n$. However, F is *unknown*. Nevertheless, in Chapter 17 we saw that the observed dataset reflects most features of the "true" probability distribution. Hence the natural thing to do is to compute an *estimate* $\hat{F}$ for the distribution function F and then to consider a random sample from $\hat{F}$ and the corresponding sample mean as substitutes for the random sample $X_1, X_2, \ldots, X_n$ from F and the random variable $\bar{X}_n$. A random sample from $\hat{F}$ is called a *bootstrap random sample*, or briefly bootstrap sample, and is denoted by

$$X_1^*, X_2^*, \ldots, X_n^*$$

to distinguish it from the random sample $X_1, X_2, \ldots, X_n$ from the "true" F. The corresponding average is called the *bootstrapped sample mean*, and this random variable is denoted by

$$\bar{X}_n^* = \frac{X_1^* + X_2^* + \cdots + X_n^*}{n}$$

to distinguish it from the random variable $\bar{X}_n$. The idea is now to use the distribution of $\bar{X}_n^*$ to approximate the distribution of $\bar{X}_n$.

The preceding procedure is called the *bootstrap principle* for the sample mean. Clearly, it can be applied to *any* sample statistic $h(X_1, X_2, \ldots, X_n)$ by approximating its probability distribution by that of the corresponding bootstrapped sample statistic $h(X_1^*, X_2^*, \ldots, X_n^*)$.

> Bootstrap principle. Use the dataset $x_1, x_2, \ldots, x_n$ to compute an estimate $\hat{F}$ for the "true" distribution function F. Replace the random sample $X_1, X_2, \ldots, X_n$ from F by a random sample $X_1^*, X_2^*, \ldots, X_n^*$ from $\hat{F}$, and approximate the probability distribution of $h(X_1, X_2, \ldots, X_n)$ by that of $h(X_1^*, X_2^*, \ldots, X_n^*)$.

Returning to the sample mean, the first question that comes to mind is, of course, how well does the distribution of $\bar{X}_n^*$ approximate the distribution

[1] In Section 11.1 we saw how the distribution of the sum of independent random variables can be computed. Together with the change-of-units rule (see page 106), the distribution of $\bar{X}_n$ can be determined. See also Section 13.1, where this is done for independent $Gam(2,1)$ variables.

of $\bar{X}_n$? Or more generally, how well does the distribution of a bootstrapped sample statistic $h(X_1^*, X_2^*, \ldots, X_n^*)$ approximate the distribution of the sample statistic of interest $h(X_1, X_2, \ldots, X_n)$? Applied in such a straightforward manner, the bootstrap approximation for the distribution of $\bar{X}_n$ by that of $\bar{X}_n^*$ may not be so good (see Remark 18.1). The bootstrap approximation will improve if we approximate the distribution of the *centered* sample mean:

$$\bar{X}_n - \mu,$$

where μ is the expectation corresponding to F. The bootstrapped version would be the random variable

$$\bar{X}_n^* - \mu^*,$$

where μ^* is the expectation corresponding to $\hat{F}$. Often the bootstrap approximation of the distribution of a sample statistic will improve if we somehow normalize the sample statistic by relating it to a corresponding feature of the "true" distribution. An example is the centered sample median

$$\mathrm{Med}(X_1, X_2, \ldots, X_n) - F^{\mathrm{inv}}(0.5),$$

where we subtract the median $F^{\mathrm{inv}}(0.5)$ of F. Another example is the normalized sample variance

$$\frac{S_n^2}{\sigma^2},$$

where we divide by the variance σ^2 of F.

Quick exercise 18.1 Describe how the bootstrap principle should be applied to approximate the distribution of $\mathrm{Med}(X_1, X_2, \ldots, X_n) - F^{\mathrm{inv}}(0.5)$.

Remark 18.1 (The bootstrap for the sample mean). To see why the bootstrap approximation for $\bar{X}_n$ may be bad, consider a dataset $x_1, x_2, \ldots, x_n$ that is a realization of a random sample $X_1, X_2, \ldots, X_n$ from an $N(\mu, 1)$ distribution. In that case the corresponding sample mean $\bar{X}_n$ has an $N(\mu, 1/n)$ distribution. We estimate μ by $\bar{x}_n$ and replace the random sample from an $N(\mu, 1)$ distribution by a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ from an $N(\bar{x}_n, 1)$ distribution. The corresponding bootstrapped sample mean $\bar{X}_n^*$ has an $N(\bar{x}_n, 1/n)$ distribution. Therefore the distribution functions G_n and G_n^* of the random variables $\bar{X}_n$ and $\bar{X}_n^*$ can be determined:

$$G_n(a) = \Phi(\sqrt{n}(a - \mu)) \quad \text{and} \quad G_n^*(a) = \Phi(\sqrt{n}(a - \bar{x}_n)).$$

In this case it turns out that the maximum distance between the two distribution functions is equal to

$$2\Phi\left(\tfrac{1}{2}\sqrt{n}|\bar{x}_n - \mu|\right) - 1.$$

Since $\bar{X}_n$ has an $N(\mu, 1/n)$ distribution, this value is approximately equal to $2\Phi(|z|/2) - 1$, where z is a realization of an $N(0,1)$ random variable Z. This only equals zero for $z = 0$, so that the distance between the distribution functions of $\bar{X}_n$ and $\bar{X}_n^*$ will almost always be strictly positive, even for large n.

The question that remains is what to take as an estimate $\hat{F}$ for F. This will depend on how well F can be specified. For the Old Faithful data we cannot say anything about the type of distribution. However, for the software data it seems reasonable to model the dataset as a realization of a random sample from an $Exp(\lambda)$ distribution and then we only have to estimate the parameter λ. Different assumptions about F give rise to different bootstrap procedures. We will discuss two of them in the next sections.

18.2 The empirical bootstrap

Suppose we consider our dataset $x_1, x_2, \ldots, x_n$ as a realization of a random sample from a distribution function F. When we cannot make any assumptions about the type of F, we can always estimate F by the empirical distribution function of the dataset:

$$\hat{F}(a) = F_n(a) = \frac{\text{number of } x_i \text{ less than or equal to } a}{n}.$$

Since we estimate F by the empirical distribution function, the corresponding bootstrap principle is called the *empirical bootstrap.* Applying this principle to the centered sample mean, the random sample $X_1, X_2, \ldots, X_n$ from F is replaced by a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ from F_n, and the distribution of $\bar{X}_n - \mu$ is approximated by that of $\bar{X}_n^* - \mu^*$, where μ^* denotes the expectation corresponding to F_n. The question is, of course, how good this approximation is. A mathematical theorem tells us that the empirical bootstrap works for the centered sample mean, i.e., the distribution of $\bar{X}_n - \mu$ is well approximated by that of $\bar{X}_n^* - \mu^*$ (see Remark 18.2). On the other hand, there are (normalized) sample statistics for which the empirical bootstrap fails, such as

$$1 - \frac{\text{maximum of } X_1, X_2, \ldots, X_n}{\theta},$$

based on a random sample $X_1, X_2, \ldots, X_n$ from a $U(0, \theta)$ distribution (see Exercise 18.12).

Remark 18.2 (The empirical bootstrap for $\bar{X}_n - \mu$). For the centered sample mean the bootstrap approximation works, even if we estimate F by the empirical distribution function F_n. If G_n denotes the distribution function of $\bar{X}_n - \mu$ and G_n^* the distribution function of its bootstrapped version $\bar{X}_n^* - \mu^*$, then the maximum distance between G_n^* and G_n goes to zero with probability one:

$$\mathrm{P}\left(\lim_{n\to\infty}\sup_{t\in\mathbb{R}}|G_n^*(t)-G_n(t)|=0\right)=1$$

(see, for instance, Singh [32]). In fact, the empirical bootstrap approximation can be improved by approximating the distribution of the standardized average $\sqrt{n}(\bar{X}_n-\mu)/\sigma$ by its bootstrapped version $\sqrt{n}(\bar{X}_n^*-\mu^*)/\sigma^*$, where σ and σ^* denote the standard deviations of F and F_n. This approximation is even better than the normal approximation by the central limit theorem! See, for instance, Hall [14].

Let us continue with approximating the distribution of $\bar{X}_n-\mu$ by that of $\bar{X}_n^*-\mu^*$. First note that the empirical distribution function F_n of the original dataset is the distribution function of a discrete random variable that attains the values $x_1, x_2, \ldots, x_n$, each with probability $1/n$. This means that each of the bootstrap random variables X_i^* has expectation

$$\mu^*=\mathrm{E}\left[X_i^*\right]=x_1\cdot\frac{1}{n}+x_2\cdot\frac{1}{n}+\cdots+x_n\cdot\frac{1}{n}=\bar{x}_n.$$

Therefore, applying the empirical bootstrap to $\bar{X}_n-\mu$ means approximating its distribution by that of $\bar{X}_n^*-\bar{x}_n$. In principle it would be possible to determine the probability distribution of $\bar{X}_n^*-\bar{x}_n$. Indeed, the random variable $\bar{X}_n^*$ is based on the random variables X_i^*, whose distribution we know precisely: it takes values $x_1, x_2, \ldots, x_n$ with equal probability $1/n$. Hence we could determine the possible values of $\bar{X}_n^*-\bar{x}_n$ and the corresponding probabilities. For small n this can be done (see Exercise 18.5), but for large n this becomes cumbersome. Therefore we invoke a second approximation.

Recall the jury example in Section 6.3, where we investigated the variation of two different rules that a jury might use to assign grades. In terms of the present chapter, the jury example deals with a random sample from a $U(-0.5, 0.5)$ distribution and two different sample statistics T and M, corresponding to the two rules. To investigate the distribution of T and M, a simulation was carried out with one thousand runs, where in every run we generated a realization of a random sample from the $U(-0.5, 0.5)$ distribution and computed the corresponding realization of T and M. The one thousand realizations give a good impression of how T and M vary around the deserved score (see Figure 6.4).

Returning to the distribution of $\bar{X}_n^*-\bar{x}_n$, the analogue would be to repeatedly generate a realization of the bootstrap random sample from F_n and every time compute the corresponding realization of $\bar{X}_n^*-\bar{x}_n$. The resulting realizations would give a good impression about the distribution of $\bar{X}_n^*-\bar{x}_n$. A realization of the bootstrap random sample is called a *bootstrap dataset* and is denoted by

$$x_1^*, x_2^*, \ldots, x_n^*$$

to distinguish it from the original dataset $x_1, x_2, \ldots, x_n$. For the centered sample mean the simulation procedure is as follows.

Empirical bootstrap simulation (for $\bar{X}_n - \mu$). Given a dataset $x_1, x_2, \ldots, x_n$, determine its empirical distribution function F_n as an estimate of F, and compute the expectation

$$\mu^* = \bar{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

corresponding to F_n.

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

$$\bar{x}_n^* - \bar{x}_n,$$

where

$$\bar{x}_n^* = \frac{x_1^* + x_2^* + \cdots + x_n^*}{n}.$$

Repeat steps 1 and 2 many times.

Note that generating a value x_i^* from F_n is equivalent to choosing one of the elements $x_1, x_2, \ldots, x_n$ of the original dataset with equal probability $1/n$.

The empirical bootstrap simulation is described for the centered sample mean, but clearly a similar simulation procedure can be formulated for any (normalized) sample statistic.

Remark 18.3 (Some history). Although Efron [7] in 1979 drew attention to diverse applications of the empirical bootstrap simulation, it already existed before that time, but not as a unified widely applicable technique. See Hall [14] for references to earlier ideas along similar lines and to further development of the bootstrap. One of Efron's contributions was to point out how to combine the bootstrap with modern computational power. In this way, the interest in this procedure is a typical consequence of the influence of computers on the development of statistics in the past decades. Efron also coined the term "bootstrap," which is inspired by the American version of one of the tall stories of the Baron von Münchhausen, who claimed to have lifted himself out of a swamp by pulling the strap on his boot (in the European version he lifted himself by pulling his hair).

Quick exercise 18.2 Describe the empirical bootstrap simulation for the centered sample median $\mathrm{Med}(X_1, X_2, \ldots, X_n) - F^{\mathrm{inv}}(0.5)$.

For the Old Faithful data we carried out the empirical bootstrap simulation for the centered sample mean with one thousand repetitions. In Figure 18.1 a histogram (left) and kernel density estimate (right) are displayed of one thousand centered bootstrap sample means

$$\bar{x}_{n,1}^* - \bar{x}_n \quad \bar{x}_{n,2}^* - \bar{x}_n \quad \cdots \quad \bar{x}_{n,1000}^* - \bar{x}_n.$$

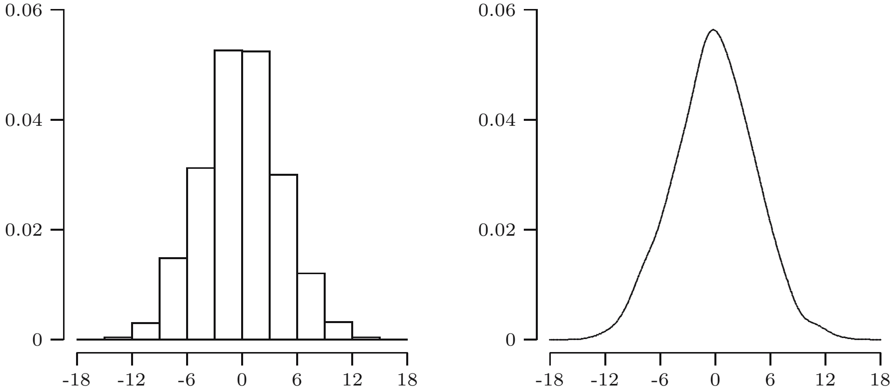

Fig. 18.1. Histogram and kernel density estimate of centered bootstrap sample means.

Since these are realizations of the random variable $\bar{X}_n^* - \bar{x}_n$, we know from Section 17.2 that they reflect the distribution of $\bar{X}_n^* - \bar{x}_n$. Hence, as the distribution of $\bar{X}_n^* - \bar{x}_n$ approximates that of $\bar{X}_n - \mu$, the centered bootstrap sample means also reflect the distribution of $\bar{X}_n - \mu$. This leads to the following application.

An application of the empirical bootstrap

Let us return to our example about the Old Faithful data, which are modeled as a realization of a random sample from some F. Suppose we estimate the expectation μ corresponding to F by $\bar{x}_n = 209.3$. Can we say how far away 209.3 is from the "true" expectation μ? To be honest, the answer is *no*... (oops). In a situation like this, the measurements and their corresponding average are subject to randomness, so that we cannot say anything with absolute certainty about how far away the average will be from μ. One of the things we can say is how *likely* it is that the average is within a given distance from μ.

To get an impression of how close the average of a dataset of $n = 272$ observed durations of the Old Faithful geyser is to μ, we want to compute the probability that the sample mean deviates more than 5 from μ:

$$\mathrm{P}\big(|\bar{X}_n - \mu| > 5\big)\,.$$

Direct computation of this probability is impossible, because we do not know the distribution of the random variable $\bar{X}_n - \mu$. However, since the distribution of $\bar{X}_n^* - \bar{x}_n$ approximates the distribution of $\bar{X}_n - \mu$, we can approximate the probability as follows

$$\mathrm{P}\big(|\bar{X}_n - \mu| > 5\big) \approx \mathrm{P}\big(|\bar{X}_n^* - \bar{x}_n| > 5\big) = \mathrm{P}\big(|\bar{X}_n^* - 209.3| > 5\big)\,,$$

where we have also used that for the Old Faithful data, $\bar{x}_n = 209.3$. As we mentioned before, in principle it is possible to compute the last probability exactly. Since this is too cumbersome, we approximate $\mathrm{P}\big(|\bar{X}_n^* - 209.3| > 5\big)$ by means of the one thousand centered bootstrap sample means obtained from the empirical bootstrap simulation:

$$\bar{x}_{n,1}^* - 209.3 \quad \bar{x}_{n,2}^* - 209.3 \quad \cdots \quad \bar{x}_{n,1000}^* - 209.3.$$

In view of Table 17.2, a natural estimate for $\mathrm{P}\big(|\bar{X}_n^* - 209.3| > 5\big)$ is the relative frequency of centered bootstrap sample means that are greater than 5 in absolute value:

$$\frac{\text{number of } i \text{ with } |\bar{x}_{n,i}^* - 209.3| \text{ greater than } 5}{1000}.$$

For the centered bootstrap sample means of Figure 18.1, this relative frequency is 0.227. Hence, we obtain the following bootstrap approximation

$$\mathrm{P}\big(|\bar{X}_n - \mu| > 5\big) \approx \mathrm{P}\big(|\bar{X}_n^* - 209.3| > 5\big) \approx 0.227.$$

It should be emphasized that the second approximation can be made arbitrarily accurate by increasing the number of repetitions in the bootstrap procedure.

18.3 The parametric bootstrap

Suppose we consider our dataset as a realization of a random sample from a distribution of a specific parametric type. In that case the distribution function is completely determined by a parameter or vector of parameters θ: $F = F_\theta$. Then we do *not* have to estimate the whole distribution function F, but it suffices to estimate the parameter(vector) θ by $\hat{\theta}$ and estimate F by

$$\hat{F} = F_{\hat{\theta}}.$$

The corresponding bootstrap principle is called the *parametric bootstrap*.

Let us investigate what this would mean for the centered sample mean. First we should realize that the expectation of F_θ is also determined by θ: $\mu = \mu_\theta$. The parametric bootstrap for the centered sample mean now amounts to the following. The random sample $X_1, X_2, \ldots, X_n$ from the "true" distribution function F_θ is replaced by a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ from $F_{\hat{\theta}}$, and the probability distribution of $\bar{X}_n - \mu_\theta$ is approximated by that of $\bar{X}_n^* - \mu^*$, where

$$\mu^* = \mu_{\hat{\theta}}$$

denotes the expectation corresponding to $F_{\hat{\theta}}$.

Often the parametric bootstrap approximation is better than the empirical bootstrap approximation, as illustrated in the next quick exercise.

QUICK EXERCISE 18.3 Suppose the dataset $x_1, x_2, \ldots, x_n$ is a realization of a random sample $X_1, X_2, \ldots, X_n$ from an $N(\mu, 1)$ distribution. Estimate μ by $\bar{x}_n$ and consider a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ from an $N(\bar{x}_n, 1)$ distribution. Check that the probability distributions of $\bar{X}_n - \mu$ and $\bar{X}_n^* - \bar{x}_n$ are the *same*: an $N(0, 1/n)$ distribution.

Once more, in principle it is possible to determine the distribution of $\bar{X}_n^* - \mu_{\hat{\theta}}$ exactly. However, in contrast with the situation considered in the previous quick exercise, in some cases this is still cumbersome. Again a simulation procedure may help us out. For the centered sample mean the procedure is as follows.

PARAMETRIC BOOTSTRAP SIMULATION (FOR $\bar{X}_n - \mu$). Given a dataset $x_1, x_2, \ldots, x_n$, compute an estimate $\hat{\theta}$ for θ. Determine $F_{\hat{\theta}}$ as an estimate for F_θ, and compute the expectation $\mu^* = \mu_{\hat{\theta}}$ corresponding to $F_{\hat{\theta}}$.

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from $F_{\hat{\theta}}$.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - \mu_{\hat{\theta}},$$

where

$$\bar{x}_n^* = \frac{x_1^* + x_2^* + \cdots + x_n^*}{n}.$$

Repeat steps 1 and 2 many times.

As an application we will use the parametric bootstrap simulation to investigate whether the exponential distribution is a reasonable model for the software data.

Are the software data exponential?

Consider fitting an exponential distribution to the software data, as discussed in Section 17.3. At first sight, Figure 17.6 shows a reasonable fit with the exponential distribution. One way to quantify the difference between the dataset and the exponential model is to compute the maximum distance between the empirical distribution function F_n of the dataset and the exponential distribution function $F_{\hat{\lambda}}$ estimated from the dataset:

$$t_{\mathrm{ks}} = \sup_{a \in \mathbb{R}} |F_n(a) - F_{\hat{\lambda}}(a)|.$$

Here $F_{\hat{\lambda}}(a) = 0$ for $a < 0$ and

$$F_{\hat{\lambda}}(a) = 1 - \mathrm{e}^{-\hat{\lambda} a} \quad \text{for } a \geq 0,$$

where $\hat{\lambda} = 1/\bar{x}_n$ is estimated from the dataset. The quantity t_{ks} is called the *Kolmogorov-Smirnov distance* between F_n and $F_{\hat{\lambda}}$.

The idea behind the use of this distance is the following. If F denotes the "true" distribution function, then according to Section 17.2 the empirical distribution function F_n will resemble F whether F equals the distribution function F_λ of some $Exp(\lambda)$ distribution or not. On the other hand, if the "true" distribution function *is* F_λ, then the estimated exponential distribution function $F_{\hat{\lambda}}$ will resemble F_λ, because $\hat{\lambda} = 1/\bar{x}_n$ is close to the "true" λ. Therefore, if $F = F_\lambda$, then both F_n and $F_{\hat{\lambda}}$ will be close to the same distribution function, so that t_{ks} is small; if F is different from F_λ, then F_n and $F_{\hat{\lambda}}$ are close to two different distribution functions, so that t_{ks} is large. The value t_{ks} is always between 0 and 1, and the further away this value is from 0, the more it is an indication that the exponential model is inappropriate. For the software dataset we find $\hat{\lambda} = 1/\bar{x}_n = 0.0015$ and $t_{\mathrm{ks}} = 0.176$. Does this speak against the believed exponential model?

One way to investigate this is to find out whether, in the case when the data are truly a realization of an exponential random sample from F_λ, the value 0.176 is unusually large. To answer this question we consider the sample statistic that corresponds to t_{ks}. The estimate $\hat{\lambda} = 1/\bar{x}_n$ is replaced by the random variable $\hat{\Lambda} = 1/\bar{X}_n$, and the empirical distribution function of the dataset is replaced by the empirical distribution function of the random sample $X_1, X_2, \ldots, X_n$ (again denoted by F_n):

$$F_n(a) = \frac{\text{number of } X_i \text{ less than or equal to } a}{n}.$$

In this way, t_{ks} is a realization of the sample statistic

$$T_{\mathrm{ks}} = \sup_{a \in \mathbb{R}} |F_n(a) - F_{\hat{\Lambda}}(a)|.$$

To find out whether 0.176 is an exceptionally large value for the random variable T_{ks}, we must determine the probability distribution of T_{ks}. However, this is impossible because the parameter λ of the $Exp(\lambda)$ distribution is unknown. We will approximate the distribution of T_{ks} by a parametric bootstrap. We use the dataset to estimate λ by $\hat{\lambda} = 1/\bar{x}_n = 0.0015$ and replace the random sample $X_1, X_2, \ldots, X_n$ from F_λ by a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ from $F_{\hat{\lambda}}$. Next we approximate the distribution of T_{ks} by that of its bootstrapped version

$$T_{\mathrm{ks}}^* = \sup_{a \in \mathbb{R}} |F_n^*(a) - F_{\hat{\Lambda}^*}(a)|,$$

where F_n^* is the empirical distribution function of the bootstrap random sample:

$$F_n^*(a) = \frac{\text{number of } X_i^* \text{ less than or equal to } a}{n},$$

and $\hat{\Lambda}^* = 1/\bar{X}_n^*$, with $\bar{X}_n^*$ being the average of the bootstrap random sample. The bootstrapped sample statistic T_{ks}^* is too complicated to determine its probability distribution, and hence we perform a parametric bootstrap simulation:

1. We generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_{135}^*$ from an exponential distribution with parameter $\hat{\lambda} = 0.0015$.
2. We compute the bootstrapped KS distance

$$t_{\mathrm{ks}}^* = \sup_{a \in \mathbb{R}} |F_n^*(a) - F_{\hat{\lambda}^*}(a)|,$$

where F_n^* denotes the empirical distribution function of the bootstrap dataset and $F_{\hat{\lambda}^*}$ denotes the estimated exponential distribution function, where $\hat{\lambda}^* = 1/\bar{x}_n^*$ is computed from the bootstrap dataset.

We repeat steps 1 and 2 one thousand times, which results in one thousand values of the bootstrapped KS distance. In Figure 18.2 we have displayed a histogram and kernel density estimate of the one thousand bootstrapped KS distances. It is clear that if the software data would come from an exponential distribution, the value 0.176 of the KS distance would be very unlikely! This strongly suggests that the exponential distribution is not the right model for the software data. The reason for this is that the Poisson process is the wrong model for the series of failures. A closer inspection shows that the rate at which failures occur over time is not constant, as was assumed in Chapter 17, but decreases.

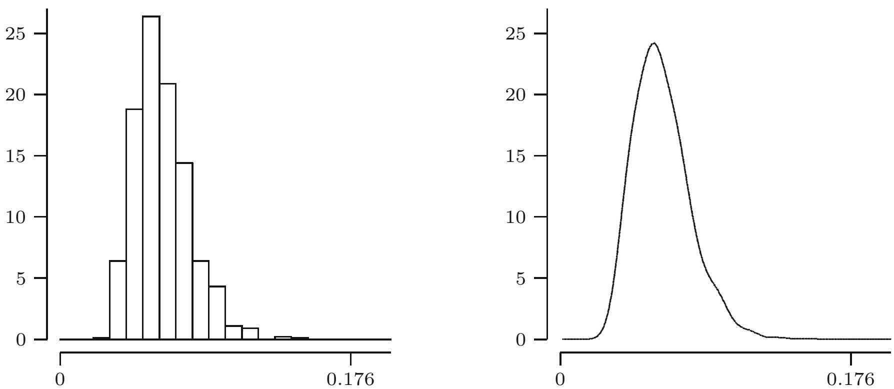

Fig. 18.2. One thousand bootstrapped KS distances.

18.4 Solutions to the quick exercises

18.1 You could have written something like the following: *"Use the dataset* $x_1, x_2, \ldots, x_n$ *to compute an estimate* $\hat{F}$ *for* F*. Replace the random sample* $X_1, X_2, \ldots, X_n$ *from* F *by a random sample* $X_1^*, X_2^*, \ldots, X_n^*$ *from* $\hat{F}$*, and approximate the probability distribution of*

$$Med(X_1, X_2, \ldots, X_n) - F^{\text{inv}}(0.5)$$

by that of $Med(X_1^*, X_2^*, \ldots, X_n^*) - \hat{F}^{\text{inv}}(0.5)$*, where* $\hat{F}^{\text{inv}}(0.5)$ *is the median of* $\hat{F}$*."*

18.2 You could have written something like the following: *"Given a dataset* $x_1, x_2, \ldots, x_n$*, determine its empirical distribution function* F_n *as an estimate of* F*, and the median* $F^{\text{inv}}(0.5)$ *of* F_n*.*

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

$$Med_n^* - F^{\text{inv}}(0.5),$$

where $Med_n^* =$ *sample median of* $x_1^*, x_2^*, \ldots, x_n^*$*.*

Repeat steps 1 and 2 many times."

Note that if n is odd, then $F^{\text{inv}}(0.5)$ equals the sample median of the original dataset, but this is not necessarily so for n even.

18.3 According to Remark 11.2 about the sum of independent normal random variables, the sum of n independent $N(\mu, 1)$ distributed random variables has an $N(n\mu, n)$ distribution. Hence by the change-of-units rule for the normal distribution (see page 106), it follows that $\bar{X}_n$ has an $N(\mu, 1/n)$ distribution, and that $\bar{X}_n - \mu$ has an $N(0, 1/n)$ distribution. Similarly, the average $\bar{X}_n^*$ of n independent $N(\bar{x}_n, 1)$ distributed bootstrap random variables has a normal distribution $N(\bar{x}_n, 1/n)$ distribution, and therefore $\bar{X}_n^* - \bar{x}_n$ again has an $N(0, 1/n)$ distribution.

18.5 Exercises

18.1 ⊡ We generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_6^*$ from the empirical distribution function of the dataset

$$2 \quad 1 \quad 1 \quad 4 \quad 6 \quad 3,$$

i.e., we draw (with replacement) six values from these numbers with equal probability 1/6. How many different bootstrap datasets are possible? Are they all equally likely to occur?

18.2 We generate a bootstrap dataset $x_1^*, x_2^*, x_3^*, x_4^*$ from the empirical distribution function of the dataset

$$1 \quad 3 \quad 4 \quad 6.$$

a. Compute the probability that the bootstrap sample mean is equal to 1.

b. Compute the probability that the maximum of the bootstrap dataset is equal to 6.

c. Compute the probability that exactly two elements in the bootstrap sample are less than 2.

18.3 ⊞ We generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_{10}^*$ from the empirical distribution function of the dataset

$$\begin{array}{ccccc} 0.39 & 0.41 & 0.38 & 0.44 & 0.40 \\ 0.36 & 0.34 & 0.46 & 0.35 & 0.37. \end{array}$$

a. Compute the probability that the bootstrap dataset has exactly three elements equal to 0.35.

b. Compute the probability that the bootstrap dataset has at most two elements less than or equal to 0.38.

c. Compute the probability that the bootstrap dataset has exactly two elements less than or equal to 0.38 and all other elements greater than 0.42.

18.4 ⊡ Consider the dataset from Exercise 18.3, with maximum 0.46.

a. We generate a bootstrap random sample $X_1^*, X_2^*, \ldots, X_{10}^*$ from the empirical distribution function of the dataset. Compute $\mathrm{P}(M_{10}^* < 0.46)$, where $M_{10}^* = \max\{X_1^*, X_2^*, \ldots, X_{10}^*\}$.

b. The same question as in **a**, but now for a dataset with distinct elements $x_1, x_2, \ldots, x_n$ and maximum m_n. Compute $\mathrm{P}(M_n^* < m_n)$, where M_n^* is the maximum of a bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ generated from the empirical distribution function of the dataset.

18.5 ⊡ Suppose we have a dataset

$$0 \quad 3 \quad 6,$$

which is the realization of a random sample from a distribution function F. If we estimate F by the empirical distribution function, then according to the bootstrap principle applied to the centered sample mean $\bar{X}_3 - \mu$, we must replace this random variable by its bootstrapped version $\bar{X}_3^* - \bar{x}_3$. Determine the possible values for the bootstrap random variable $\bar{X}_3^* - \bar{x}_3$ and the corresponding probabilities.

18.6 Suppose that the dataset $x_1, x_2, \ldots, x_n$ is a realization of a random sample from an $Exp(\lambda)$ distribution with distribution function F_λ, and that $\bar{x}_n = 5$.

a. Check that the median of the $Exp(\lambda)$ distribution is $m_\lambda = (\ln 2)/\lambda$ (see also Exercise 5.11).

b. Suppose we estimate λ by $1/\bar{x}_n$. Describe the parametric bootstrap simulation for $\mathrm{Med}(X_1, X_2, \ldots, X_n) - m_\lambda$.

18.7 ⊞ To give an example in which the bootstrapped centered sample mean in the parametric and empirical bootstrap simulations may be *different*, consider the following situation. Suppose that the dataset $x_1, x_2, \ldots, x_n$ is a realization of a random sample from a $U(0, \theta)$ distribution with expectation $\mu = \theta/2$. We estimate θ by

$$\hat{\theta} = \frac{n+1}{n} m_n,$$

where $m_n = \max\{x_1, x_2, \ldots, x_n\}$. Describe the parametric bootstrap simulation for the centered sample mean $\bar{X}_n - \mu$.

18.8 ⊞ Here is an example in which the bootstrapped centered sample mean in the parametric and empirical bootstrap simulations are the *same*. Consider the software data with average $\bar{x}_n = 656.8815$ and median $m_n = 290$, modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from a distribution function F with expectation μ. By means of bootstrap simulation we like to get an impression of the distribution of $\bar{X}_n - \mu$.

a. Suppose that we assume nothing about the distribution of the interfailure times. Describe the appropriate bootstrap simulation procedure with one thousand repetitions.

b. Suppose we assume that F is the distribution function of an $Exp(\lambda)$ distribution, where λ is estimated by $1/\bar{x}_n = 0.0015$. Describe the appropriate bootstrap simulation procedure with one thousand repetitions.

c. Suppose we assume that F is the distribution function of an $Exp(\lambda)$ distribution, and that (as suggested by Exercise 18.6 **a**) the parameter λ is estimated by $(\ln 2)/m_n = 0.0024$. Describe the appropriate bootstrap simulation procedure with one thousand repetitions.

18.9 ⊡ Consider the dataset from Exercises 15.1 and 17.6 consisting of measured chest circumferences of Scottish soldiers with average $\bar{x}_n = 39.85$ and sample standard deviation $s_n = 2.09$. The histogram in Figure 17.11 suggests modeling the data as the realization of a random sample $X_1, X_2, \ldots, X_n$ from an $N(\mu, \sigma^2)$ distribution. We estimate μ by the sample mean and we are interested in the probability that the sample mean deviates more than 1 from μ: $\mathrm{P}\big(|\bar{X}_n - \mu| > 1\big)$. Describe how one can use the bootstrap principle to approximate this probability, i.e., describe the distribution of the bootstrap random sample $X_1^*, X_2^*, \ldots, X_n^*$ and compute $\mathrm{P}\big(|\bar{X}_n^* - \mu^*| > 1\big)$. Note that one does not need a simulation to approximate this latter probability.

18.10 Consider the software data, with average $\bar{x}_n = 656.8815$, modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from a distribution function F. We estimate the expectation μ of F by the sample mean and we are interested in the probability that the sample mean deviates more than ten from μ: $\mathrm{P}\big(|\bar{X}_n - \mu| > 10\big)$.

a. Suppose we assume nothing about the distribution of the interfailure times. Describe how one can obtain a bootstrap approximation for the probability, i.e., describe the appropriate bootstrap simulation procedure with one thousand repetitions and how the results of this simulation can be used to approximate the probability.

b. Suppose we assume that F is the distribution function of an $Exp(\lambda)$ distribution. Describe how one can obtain a bootstrap approximation for the probability.

18.11 Consider the dataset of measured chest circumferences of 5732 Scottish soldiers (see Exercises 15.1, 17.6, and 18.9). The Kolmogorov-Smirnov distance between the empirical distribution function and the distribution function $F_{\bar{x}_n, s_n}$ of the normal distribution with estimated parameters $\hat{\mu} = \bar{x}_n = 39.85$ and $\hat{\sigma} = s_n = 2.09$ is equal to

$$t_{\text{ks}} = \sup_{a \in \mathbb{R}} |F_n(a) - F_{\bar{x}_n, s_n}(a)| = 0.0987,$$

where $\bar{x}_n$ and s_n denote sample mean and sample standard deviation of the dataset. Suppose we want to perform a bootstrap simulation with one thousand repetitions for the KS distance to investigate to which degree the value 0.0987 agrees with the assumed normality of the dataset. Describe the appropriate bootstrap simulation that must be carried out.

18.12 To give an example where the empirical bootstrap fails, consider the following situation. Suppose our dataset $x_1, x_2, \ldots, x_n$ is a realization of a random sample $X_1, X_2, \ldots, X_n$ from a $U(0, \theta)$ distribution. Consider the normalized sample statistic

$$T_n = 1 - \frac{M_n}{\theta},$$

where M_n is the maximum of $X_1, X_2, \ldots, X_n$. Let $X_1^*, X_2^*, \ldots, X_n^*$ be a bootstrap random sample from the empirical distribution function of our dataset, and let M_n^* be the corresponding bootstrap maximum. We are going to compare the distribution functions of T_n and its bootstrap counterpart

$$T_n^* = 1 - \frac{M_n^*}{m_n},$$

where m_n is the maximum of $x_1, x_2, \ldots, x_n$.

a. Check that $\mathrm{P}(T_n \leq 0) = 0$ and show that

$$\mathrm{P}(T_n^* \leq 0) = 1 - \left(1 - \frac{1}{n}\right)^n.$$

Hint: first argue that $\mathrm{P}(T_n^* \leq 0) = \mathrm{P}(M_n^* = m_n)$, and then use the result of Exercise 18.4.

b. Let $G_n(t) = \mathrm{P}(T_n \leq t)$ be the distribution function of T_n, and similarly let $G_n^*(t) = \mathrm{P}(T_n^* \leq t)$ be the distribution function of the bootstrap statistic T_n^*. Conclude from part **a** that the maximum distance between G_n^* and G_n can be bounded from below as follows:

$$\sup_{t\in\mathbb{R}} |G_n^*(t) - G_n(t)| \geq 1 - \left(1 - \frac{1}{n}\right)^n .$$

c. Use part **b** to argue that for all n, the maximum distance between G_n^* and G_n is greater than 0.632:

$$\sup_{t\in\mathbb{R}} |G_n^*(t) - G_n(t)| \geq 1 - \mathrm{e}^{-1} = 0.632.$$

Hint: you may use that $\mathrm{e}^{-x} \geq 1 - x$ for all x.
We conclude that even for very large sample sizes the maximum distance between the distribution functions of T_n and its bootstrap counterpart T_n^* is at least 0.632.

18.13 (Exercise 18.12 continued). In contrast to the empirical bootstrap, the parametric bootstrap for T_n *does* work. Suppose we estimate the parameter θ of the $U(0,\theta)$ distribution by

$$\hat{\theta} = \frac{n+1}{n} m_n, \quad \text{where } m_n = \text{maximum of } x_1, x_2, \ldots, x_n.$$

Let now $X_1^*, X_2^*, \ldots, X_n^*$ be a bootstrap random sample from a $U(0, \hat{\theta})$ distribution, and let M_n^* be the corresponding bootstrap maximum. Again, we are going to compare the distribution function G_n of $T_n = 1 - M_n/\theta$ with the distribution function G_n^* of its bootstrap counterpart $T_n^* = 1 - M_n^*/\hat{\theta}$.

a. Check that the distribution function F_θ of a $U(0,\theta)$ distribution is given by

$$F_\theta(a) = \frac{a}{\theta} \quad \text{for } 0 \leq a \leq \theta.$$

b. Check that the distribution function of T_n is

$$G_n(t) = \mathrm{P}(T_n \leq t) = 1 - (1-t)^n \quad \text{for } 0 \leq t \leq 1.$$

Hint: rewrite $\mathrm{P}(T_n \leq t)$ as $1 - \mathrm{P}(M_n \leq \theta(1-t))$ and use the rule on page 109 about the distribution function of the maximum.

c. Show that T_n^* has the same distribution function:

$$G_n^*(t) = \mathrm{P}(T_n^* \leq t) = 1 - (1-t)^n \quad \text{for } 0 \leq t \leq 1.$$

This means that, in contrast to the empirical bootstrap (see Exercise 18.12), the parametric bootstrap works perfectly in this situation.