37. Bootstrap, Markov Chain and Estimating Function

In this chapter, we first review bootstrap methods for constructing confidence intervals (regions). We then discuss the following three important properties of these methods: (i) invariance under reparameterization; (ii) automatic computation; and (iii) higher order accuracy. The greatest potential value of the bootstrap lies in complex situations, such as nonlinear regression or high-dimensional parameters for example. It is important to have bootstrap procedures that can be applied to these complex situations, but still have the three desired properties. The main purpose of this chapter is to introduce two recently developed bootstrap methods: the estimation function bootstrap, and the Markov chain marginal bootstrap method. The estimating function bootstrap has all three desired properties and it can also be applied to complex situations. The Markov chain marginal bootstrap is designed for high-dimensional parameters.

37.1 Overview

For a statistical model involving an unknown parameter θ, the two main statistical inference issues are usually: (i) point estimation $\hat{\theta}$ (how to estimate the unknown parameter θ); and (ii) how to assess the accuracy of this estimator $\hat{\theta}$ (in terms of the standard deviation or confidence interval of the unknown θ). Statisticians usually try to find the exact distribution or asymptotic distribution of the estimator $\hat{\theta}$. However, it is difficult to obtain the exact distribution or asymptotic distribution in a lot of situations. Sometimes the asymptotic distribution can be obtained, but the distribution of $\hat{\theta}$ is not well approximated. Bootstrap provides a general methodology for constructing confidence intervals for unknown parameters.

In this chapter, we first discuss bootstrap methods in terms of the following three important properties: (1) invariance under reparameterization, (2) automatic computation, and (3) higher order accuracy. To illustrate these three properties, let's consider the following simple model. Assume that $Y_1, \ldots, Y_n$ is a random sample from some unknown distribution F. Let $y = (y_1, \ldots, y_n)$ be the realization of Y. Suppose that $\theta = \theta(F)$ is the unknown parameter of interest. This θ could be the mean, or variance, or some other function of the distribution F. Let $\hat{\theta} = \hat{\theta}(y_1, \ldots, y_n)$ be the estimator of θ based on the observation y.

37.1.1 Invariance under Reparameterization

Suppose θ is a scale parameter and that $h(\theta)$ is a strictly monotonic function in the parameter space of θ. Then the new reparameterized statistical model is based on $(Y_1, \ldots, Y_n)$ and the parameter $\xi = h(\theta)$. Suppose that an estimation procedure gives $\hat{\theta}$ as the estimator of θ based on $(Y_1, \ldots, Y_n)$ and parameter θ, and $\hat{\xi}$ as the

estimator of ξ based on $(Y_1, \ldots, Y_n)$ and parameter ξ respectively. This procedure is said to be *invariant* under reparameterization if $\hat{\xi} = h(\hat{\theta})$. It is well known that the maximum likelihood procedure is invariant under reparameterization. But the moment estimation procedure is usually not invariant under reparameterization. A confidence interval procedure is *invariant* under reparameterization, if $[\hat{\theta}(\alpha/2), \hat{\theta}(1-\alpha/2)]$ is the $1-\alpha$ level confidence interval of θ based on this procedure, and $[h(\hat{\theta}[\alpha/2]), h(\hat{\theta}[1-\alpha/2])]$ is the $1-\alpha$ level confidence interval of ξ based on this procedure. Here we assume that $h(\theta)$ is a strictly increasing function. When a procedure is not invariant under reparameterization, it is usually very important to select a good transformation and perform statistical inference after transformation. This has been a topic of research in classical statistics.

37.1.2 Automatic Computation

One of the most important advantages of the bootstrap method is its automatic computation; in other words it does not depend on theoretical calculation. A procedure is called an "automatic computation" if it does not depend on any extra analytical inference. In many applications, it is very difficult (sometimes impossible) to perform analytical calculations.

37.1.3 First and Higher Order Accuracy

Suppose that θ is a scale parameter and that $\hat{\theta}[\alpha]$ is the α confidence limit of θ, based on a certain procedure. Then the procedure is said to be *first-order accurate* if $P\big(\theta < \hat{\theta}[\alpha]\big) = \alpha + O(n^{-1/2})$. It is *second-order-accurate* if $P\big(\theta < \hat{\theta}[\alpha]\big) = \alpha + O(n^{-1})$. Higher order accuracy is defined in the same way. It is well known that the standard confidence interval,

$$\left[\hat{\theta} - \hat{\sigma} z_{(\alpha/2)}\,,\ \hat{\theta} + \hat{\sigma} z_{(\alpha/2)}\right],$$

is only first-order-accurate under some conditions. Here $z_{(\alpha/2)}$ is defined by $P(Z \geq z_{(\alpha)}) = \alpha$ for a standard normal random variable Z, while $\hat{\sigma}$ is an estimator of the standard deviation of $\hat{\theta}$. One advantage of using the bootstrap method is getting confidence intervals that are accurate to the second order.

In Sect. 37.2 we will introduce Efron's bootstrap for iid samples. To construct a second-order-accurate confidence interval, four different bootstrap methods are reviewed and discussed in terms of the three properties described above. We then consider three bootstrap methods for a linear model and discuss their properties.

In some more complex situations, the observations could be heteroscedastic; in other words the variances of Y_i are different. It is important to have a bootstrap procedure that remains consistent under heteroscedasticity. When θ is a high-dimensional vector it is usually more difficult to apply a bootstrap procedure because it is: (i) computational intensive; (ii) difficult to construct a good confidence region. For high dimension problems, it is often important to have reliable computational results. Some new bootstrap methods have been proposed for these complex situations.

The main propose of this chapter is to introduce some recent developments in bootstrap methodology. We consider two bootstrap methods. The first is the estimating function (EF) bootstrap proposed in *Hu* and *Kalbfleisch* [37.1]. Instead of resampling the data itself, the EF bootstrap resamples some functions of the data in order to achieve robustness to heteroscedasticity. This EF bootstrap is often the simplest computationally and it is straightforward to define studentized versions of the EF bootstrap which are invariant under reparameterization and require very little additional calculation. This method can be used to get confidence regions that are accurate to higher orders for multidimensional parameters. When the estimating function is differentiable, it can be easily extended to deal with nuisance parameter problems. Another method is called the Markov chain marginal bootstrap (MCMB), which is useful for constructing confidence intervals or regions for high dimension parameters. The MCMB is different from most bootstrap methods in two aspects: first, it solves only one-dimensional equations for a problem with any number of dimensions; second, it produces a Markov chain rather than a (conditionally) independent sequence.

In Sect. 37.3, we introduce the EF bootstrap and discuss the properties of the EF bootstrap. Some examples are used to illustrate the procedure. The Markov chain marginal bootstrap method is introduced in Sect. 37.4. In Sect. 37.5, we use the simple linear model to illustrate the MCMB algorithm and its properties. We also apply the EF bootstrap and MCMB method to different examples. In Sect. 37.6, we discuss some issues with using bootstrap methods.

37.2 Classical Bootstrap

37.2.1 Efron's Bootstrap

To start, let's consider the simplest case. Assume that $Y_1, \ldots, Y_n$ is a random sample from some unknown distribution F. Let $y = (y_1, \ldots, y_n)$ be the realization of Y. Suppose that $\theta = \theta(F)$ is the unknown parameter (scale) of interest. This θ could be the mean, the variance, or some other function of the distribution F. Let $\hat{\theta} = \hat{\theta}(y_1, \ldots, y_n)$ be the estimator of θ based on the observation y.

The main statistical goal is to find the distribution of $\hat{\theta} - \theta$. If we know this distribution, then we can do all kinds of statistical inference, (including deriving standard deviations and confidence intervals). Before Efron's bootstrap paper [37.2], researchers focused on finding the exact distribution or asymptotical distribution based on different theoretical approaches. Most of these methods depended on certain assumptions for the distribution F.

Efron's basic idea was to use computer simulation to investigate the distribution of $\hat{\theta} - \theta$. If F is a given (known) distribution, then we can use Monte Carlo simulation to get the distribution of $\hat{\theta} - \theta$. When F is unknown, the best nonparametric estimator of F is the empirical distribution, $\hat{F}$, which gives a weight of $1/n$ to each y_i. The bootstrap procedure can be summarized as follows:

- [i.] Draw a bootstrap sample $z_1^*, \ldots, z_n^*$ from distribution $\hat{F}$, which is the same as drawing $z_1^*, \ldots, z_n^*$ from $(y_1, \ldots, y_n)$ with replacement;
- [ii.] Calculate the bootstrap estimator $\hat{\theta}^* = \hat{\theta}(z_1^*, \ldots, z_n^*)$;
- [iii.] Repeat steps (i) and (ii) B (the bootstrap sample size) times to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$. Now we define the empirical distribution:

$$\hat{G}(x) = B^{-1} \sum_{i=1}^{n} I(\hat{\theta}_i^* \le x) .$$

  Where $I(\cdot)$ is the indication function.
- [iv.] Use the empirical distribution of $\hat{\theta}^* - \hat{\theta}$ to approximate the distribution of $\hat{\theta} - \theta$.

In early work, the bootstrap estimators described above were used to estimate the bias of the estimator $\hat{\theta}$ [37.2, 3]. However, the main contribution of the bootstrap method is to provide a new way to assess the accuracy of the estimator $\hat{\theta}$. Here we discuss two main applications: (i) estimating the standard deviation of $\hat{\theta}$, and (ii) estimating the confidence interval of θ.

Standard Deviation

A commonly used measure of the accuracy of $\hat{\theta}$ is the standard deviation of $\hat{\theta}$. Based on the bootstrap estimators, we can easily calculate the bootstrap variance estimator as:

$$\hat{\sigma}^* = \left[(B-1)^{-1} \sum_{j=1}^{B} \left(\hat{\theta}_j^* - \bar{\theta}^* \right)^2 \right]^{1/2} ,$$

where $\bar{\theta}^* = B^{-1} \sum_{j=1}^{B} \hat{\theta}_j^*$. Under very general conditions, this is a consistent estimator of the true standard deviation [37.3]. The advantage of this bootstrap standard deviation, $\hat{\sigma}^*$, is that it does not depend on analytical inference. Instead, we can get it by computer simulation. When $\hat{\sigma}$ is not available, this $\hat{\sigma}^*$ can be used as the estimator of the standard deviation of $\hat{\theta}$.

Based on $\hat{\sigma}^*$, we can construct an approximate confidence interval for the unknown parameter θ as

$$\left[\hat{\theta} - \hat{\sigma}^* z_{(\alpha/2)} , \hat{\theta} + \hat{\sigma}^* z_{(\alpha/2)} \right] .$$

While this confidence interval conforms to "automatic computation", it is not "invariant under reparameterization", and is only first-order-accurate.

Confidence Interval

First we consider the following two distributions: (i) the distribution of the estimator $\hat{\theta}$,

$$H(t) = P\left[n^{-1/2} \left(\hat{\theta} - \theta \right) \le t \right] ;$$

and the corresponding distribution of the bootstrap estimator $\hat{\theta}^*$,

$$\hat{H}(t) = P\left[n^{-1/2} \left(\hat{\theta}^* - \hat{\theta} \right) \le t \right] .$$

Under certain conditions, it can be shown that [37.4]

$$\max_{t \in [-\infty, \infty]} |H(t) - \hat{H}(t)| = O_p\left(n^{1/2} \right) .$$

By using the distribution of $\hat{\theta}^* - \hat{\theta}$ to approximate the distribution of $\hat{\theta} - \theta$, we can construct the level $1 - \alpha$ confidence interval of θ as

$$\left[2\hat{\theta} - \hat{\theta}^*(1 - \alpha/2), 2\hat{\theta} - \hat{\theta}^*(\alpha/2) \right] ,$$

where $\hat{\theta}^*(\alpha)$ is the αth quantile of the bootstrap distribution $\hat{G}$. This confidence interval is also obtained via

"automatic computation", but it is not "invariant under reparameterization", and is only first-order-accurate.

Another way to construct the confidence interval is by using the distribution of $\hat{\theta}^*$ directly. The confidence interval is defined as

$$\left[\hat{\theta}^*(\alpha/2), \hat{\theta}^*(1-\alpha/2)\right] .$$

This interval is obtained via "automatic computation" and is "invariant under reparameterization", but is only first-order-accurate.

37.2.2 Second-Order-Accurate Confidence Intervals

One of the problems that has been studied the most in bootstrap literature is how to construct higher accurate bootstrap confidence intervals; see for example [37.5–8]. Here we review four commonly used methods: (i) studentized bootstrap interval; (ii) bias-corrected and accelerated method (BC_a method); (iii) approximated bootstrap confidence (ABC) interval; and (iv) prepivoting bootstrap interval. The advantages and disadvantages of these four methods are also discussed.

Studentized Bootstrap Interval

Instead of considering $n^{1/2}\left(\hat{\theta}-\theta\right)$ directly, we use the studentized statistic

$$T = \frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}(y_1, \ldots, y_n)} ,$$

where $\hat{\sigma}^2 = \hat{\sigma}^2(y_1, \ldots, y_n)$ is an estimate of the asymptotic variance $\mathrm{Var}\left(n^{1/2}\hat{\theta}\right)$. The corresponding bootstrap studentized statistic is then

$$T^* = \frac{n^{1/2}\left(\hat{\theta}^*-\hat{\theta}\right)}{\hat{\sigma}^*(z_1^*, \ldots, z_n^*)} .$$

A large number (B, say) of independent replications give the following estimated percentiles:

$$\hat{T}^{(\alpha)} = \alpha_{th} \text{ quantile of } (T^*(b), b = 1, \ldots, B) .$$

The $100\alpha_{th}$ bootstrap-t confidence endpoint $\hat{\theta}_T[\alpha]$ is then defined to be

$$\hat{\theta}_T[\alpha] = \hat{\theta} - \hat{\sigma}\hat{T}^{(1-\alpha)} .$$

Based on Edgeworth expansions of the statistics T and T^*, *Hall* [37.9] showed that

$$P(T < v) - P(T^* < v) = O_p(n^{-1}) ,$$

where the second probability is under the bootstrap distribution, so the bootstrap-t intervals are usually second-order-accurate. The advantage of this studentized bootstrap is that it is intuitive and easy to understand.

But this method is not "automatic computation"; it depends on the existence of a reliable estimator of the standard deviation, $\hat{\sigma}(y_1, \ldots, y_n)$. In a lot of applications, this may not be available. Secondly, as pointed in [37.8], even with a reliable estimator of the standard deviation, the studentized bootstrap algorithm can be numerically unstable, resulting in very long confidence intervals. Third, the studentized bootstrap intervals are not "invariant under reparameterization".

BC_a Interval

The distribution of $\hat{\theta}^*$ is usually not symmetric but instead skewed to one side. The BC_a ("bias-corrected and accelerated") intervals were studied in [37.5–8] based on the bootstrap distribution. The BC_a intervals depend on two numerical parameters: a bias-correction parameter z_0 and an acceleration a. The upper endpoint $\hat{\theta}_{BC_a}[\alpha]$ of a one-sided level-α BC_a interval is defined as

$$\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1}\left[\Psi\left(z_0 + \frac{z_0 + z^\alpha}{1 - a(z_0 + z^\alpha)}\right)\right] ,$$

where $\hat{G}$ is the empirical distribution function of the bootstrap samples, and Ψ is the standard normal cdf with $z^\alpha = \Psi^{-1}(\alpha)$.

The bias-correction parameter z_0 is usually estimated from the bootstrap sample as

$$\hat{z}_0 = \Psi^{-1}\left\{B^{-1}\sum_{b=1}^{B} I\left[\hat{\theta}^*(b) < \hat{\theta}\right]\right\} .$$

On the other hand, the acceleration parameter a is more subtle and cannot be estimated as directly from the bootstrap sample. *Di Ciccio* and *Efron* [37.8] presented several ways to estimate the acceleration parameter a. The second-order accuracy of the BC_a intervals is discussed in [37.8].

The BC_a intervals are "invariant under reparameterization". Under some conditions, the BC_a intervals are second-order-accurate. However, the BC_a intervals depend on the acceleration parameter a, which cannot be estimated directly from the bootstrap replications. The need to estimate a makes the BC_a method less intuitive to users. Therefore, the BC_a intervals are not obtained via "automatic computation".

ABC Method

The ABC method (short for "approximate bootstrap confidence" interval) is an analytic version of BC_a applied to smoothly defined parameters in exponential families. Instead of estimating z_0 using a bootstrap distribution as in the BC_a method, the ABC method estimates z_0 and the acceleration parameter a analytically. The ABC method requires one further estimate of a nonlinearity parameter. Based on these estimates, we can then construct ABC intervals. *DiCiccio* and *Efron* [37.8] provide the details of this ABC method, and they also show the second-order accuracy of this method. The ABC intervals are "invariant under reparameterization", but are not obtained via "automatic computation".

Prepivoting Method (Bootstrap Calibration)

Calibration is a bootstrap technique for getting confidence intervals accurate to higher orders. Concepts related to it have been proposed and studied in [37.9–12].

Suppose that $\hat{\theta}[\alpha]$ is the upper endpoint of a one-side level-α approximate confidence bound for parameter θ. Let

$$\gamma(\alpha) = P\left[\theta < \hat{\theta}(\alpha)\right]$$

be the calibration curve. If the approximation is perfect, then $\gamma(\alpha) = \alpha$ for any given α. Otherwise, we can use the calibration curve. For example, if $\gamma(0.03) = 0.025$ and $\gamma(0.96) = 0.975$, then we can use $\left(\hat{\theta}[0.03], \hat{\theta}[0.96]\right)$ instead of $\left(\hat{\theta}[0.025], \hat{\theta}[0.975]\right)$ as our approximate 0.95-level confidence interval.

In applications, we do not usually know the calibration curve $\gamma(\alpha)$. But we can use the bootstrap method to estimate $\gamma(\alpha)$ as follows:

$$\hat{\gamma}(\alpha) = P_*\left(\hat{\theta} < \hat{\theta}[\alpha]^*\right) ,$$

where P_* indicates the bootstrap sample and $\hat{\theta}[\alpha]^*$ is the upper α bound based on the bootstrap sample.

We can use bootstrap calibration asymptotically to obtain a higher order confidence interval from a given system of confidence intervals. Therefore, it can be applied to all of the methods reviewed in this chapter. For example, we can use bootstrap calibration to obtain third-order-accurate confidence intervals from studentized bootstrap intervals. However, bootstrap calibration involves more computation. For example, if we use $B = 1000$ (bootstrap sample size), then the bootstrap calibration will require 1 000 000 recomputations of the original statistic $\hat{\theta}$. In practice, the sample size n is usually not very large, so we can usually use one bootstrap calibration.

37.2.3 Linear Regression

The bootstraps discussed so far are based on iid samples, but in many applications this assumption does not hold. Consider the linear model

$$Y_i = \boldsymbol{x}_i\boldsymbol{\beta} + e_i ,$$

where $\boldsymbol{x}_i$ is a $k\times 1$ vector which may be a random or fixed variable. Here $\boldsymbol{\beta}$ is the $k\times 1$ parameter vector of interest, and $e_1, \ldots, e_n$ are uncorrelated errors with means of zero and variances of $\mathrm{Var}(e_i) = \sigma_i^2$, $i = 1, \ldots, n$, respectively. We assume that e_i and $\boldsymbol{x}_i$ are uncorrelated for all i when $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are random. Let

$$\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}} , \boldsymbol{e} = (e_1, \ldots, e_n)^{\mathrm{T}} ,$$
$$\text{and } \mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^{\mathrm{T}}$$

The least square estimator is then

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} .$$

Here $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is assumed to be nonsingular. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ denote the observed $\boldsymbol{Y}$.

When $e_1, \ldots, e_n$ are independent and identically distributed $\left(\sigma_i^2 = \sigma^2 \text{ for all } i\right)$, [37.2] proposed the following bootstrap method based on residuals. Let $r_i = y_i - \boldsymbol{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}$, $i = 1, \ldots, n$. We can treat $r_1, \ldots, r_n$ as observations of $e_1, \ldots, e_n$. We can resample $r_1^*, \ldots, r_n^*$ from $(r_1, \ldots, r_n)$ with replacement. Now define the bootstrap sample as

$$y_i^* = \boldsymbol{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} + r_i^* , i = 1, \ldots, n .$$

Let $\boldsymbol{Y}^* = (y_1^*, \ldots, y_n^*)^{\mathrm{T}}$. The corresponding bootstrap estimator is

$$\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}^* .$$

Based on these bootstrap estimators, we can then apply the techniques in Sect. 37.2.1 and Sect. 37.2.2 to estimate the standard deviation of $\hat{\boldsymbol{\beta}}$ and the confidence intervals for $\boldsymbol{\beta}$. However, when $\boldsymbol{x}_i$ are random and the σ_i^2 values are not the same, Efron's bootstrap, which is based on resampling the residuals, does not provide a consistent result.

To deal with this heteroscedasticity, *Freedman* [37.13] suggests the following "Pair" bootstrap: resample $\left(\boldsymbol{x}_1^*, y_1^*\right), \ldots, \left(\boldsymbol{x}_n^*, y_n^*\right)$ from $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ with replacement and compute the bootstrap least squares estimate

$$\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{*\mathrm{T}}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\mathrm{T}}\boldsymbol{Y}^* ,$$

where $\mathbf{X}^* = \left(\boldsymbol{x}_1^*, \ldots \boldsymbol{x}_n^*\right)^{\mathrm{T}}$ and $\boldsymbol{Y}^* = \left(y_1^*, \ldots, y_n^*\right)^{\mathrm{T}}$. This method is consistent for heteroscedastic errors.

Hu and *Zidek* [37.14] propose another bootstrap method based on the observation that the estimator $\hat{\boldsymbol{\beta}}$ can be rewritten as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i y_i = \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i e_i \,.$$

If we treat $z_i = \boldsymbol{x}_i r_i$ $(i = 1, \ldots, n)$ as an estimate of $\boldsymbol{x}_i e_i$, it is natural to suggest that the bootstrap estimator is:

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\sum_{i=1}^{n} z_i^* \,,$$

where $z_1^*, \ldots, z_n^*$ is the bootstrap sample, which is drawn from $(z_1, \ldots, z_n)$ with replacement. This bootstrap method is also consistent for heteroscedastic errors.

As pointed out in [37.14], the numerical result of Freedman's "pair" bootstrap can be unstable. This is because the design matrix $\mathbf{X}^*$ changes for each bootstrap sample. For the bootstrap method proposed by *Hu* and *Zidek* [37.14], the design matrix $\mathbf{X}$ maintains the sample for each bootstrap sample. This is very important for cases with a small sample size n. When we applied the studentized bootstrap to both methods with heteroscedastic errors, *Hu* and *Zidek's* bootstrap was found to be easy to extend and has substantial numerical advantages over the "pair" bootstrap [37.14].

37.2.4 Some Remarks

We have discussed four second-order-accurate bootstrap methods. These methods are mainly useful for simple situations. However, resample methods are often needed in complex situations, such as nonlinear estimators and models with high-dimensional parameters. In these situations, there are clearly several difficulties that are encountered when using the traditional bootstraps: (i) it is difficult to derive an estimate for the acceleration parameter for the $\mathrm{BC_a}$ and ABC methods; (ii) for models with high-dimensional parameters, it is difficult to apply the studentized bootstrap and the prepivoting method; (iii) models with high-dimensional parameters are computationally intensive; (iv) the bootstrap sample may be quite different from the original sample which may produce unstable results.

In the following two sections, we will describe two recent proposals intended for complex models. The estimating function (EF) bootstrap is designed for estimates obtained from estimating equations. We show that the studentized estimating function bootstrap has the three desired properties. The Markov chain marginal bootstrap (MCMB) is mainly used to reduce computation in models with high-dimensional parameters.

37.3 Bootstrap Based on Estimating Equations

The traditional bootstrap methods based involve resampling the original data over and over again. Typically, the estimator is obtained from some estimating equation (*Godambe* and *Kale* [37.15]). The estimating function (EF) bootstrap proposed by [37.1, 14, 16] emphasizes the estimating function and the equation from which the estimator is obtained.

Following the same notations used by *Hu* and *Kalbfleisch* [37.1], let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be a sequence of independent random vectors of dimension q, and $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$ be an unknown parameter vector. For specified functions $\{\boldsymbol{g}_i\} : R^q \to R^p$, suppose that $E[\boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta})] = 0$ for all $i = 1, \ldots, n$ and $\boldsymbol{\theta} \in \Omega$. We suppose that $\hat{\theta}$ is the solution of the following *unbiased linear estimating equation*

$$\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta}) = n^{-1/2}\sum \boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta}) = 0 \,. \tag{37.1}$$

Here the normalizing constant $\left(n^{-1/2}\right)$ is chosen for the convenience of expressing asymptotic results. For simplicity, we also assume that $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$ is a $1:1$ function of $\boldsymbol{\theta}$ and our main consideration will be the construction of confidence regions for the whole parameter vector $\boldsymbol{\theta}$, or for components or some functions of $\boldsymbol{\theta}$.

When the random vector $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$ is exactly pivotal [37.17, 18], we can use exact methods to obtain confidence intervals or regions. However, in most cases, $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$ is only approximately pivotal and we rely on asymptotic normality and χ^2 approximations to obtain the confidence intervals or regions of $\boldsymbol{\theta}$. Here we propose to use resampling methods to approximate the distribution of $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$.

37.3.1 EF Bootstrap and Studentized EF Bootstrap

The EF Bootstrap

Let $\boldsymbol{z}_i = \boldsymbol{g}_i\left(\boldsymbol{y}_i, \hat{\theta}\right)$.

1. Draw a bootstrap sample $(z_1^*, \ldots, z_n^*)$ from $(z_1, \ldots, z_n)$ with replacement.
2. Compute $S^* = n^{-1/2} \sum z_i^*$. The bootstrap distribution of S^* can be used to approximate the distribution of $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$.
3. Compute θ^* by solving $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta}) = S^*$.

The EF bootstrap generates a bootstrap sequence θ_j^* $(j = 1, \ldots, B$ where B is the bootstrap sample size) by repeating the above process B times. Based on θ_j^* $(j = 1, \ldots, B)$, we can then construct confidence regions of the parameter of interest (some functions of $\boldsymbol{\theta}$). Like Efron's bootstrap, this usually produces confidence intervals that are accurate only to the first order. *Hu* and *Kalbfleisch* [37.16] proposed one type of studentization. This studentization gives an approximation to second-order accuracy, but it is not invariant under reparameterization.

Here we introduce the studentized EF bootstrap proposed in [37.1]. We define

$$\mathbf{V}(\boldsymbol{y}, \boldsymbol{\theta}) = n^{-1} \sum [\boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta}) - \bar{g}][\boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta}) - \bar{g}]^{\mathrm{T}} , \tag{37.2}$$

where $\bar{g} = n^{-1} \sum \boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta})$. In practice, we use the variance estimate

$$\hat{\mathbf{V}} = \mathbf{V}(\boldsymbol{y}, \hat{\theta}) . \tag{37.3}$$

Instead of approximating the distribution of $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$, we use a bootstrap method to approximate the distribution of

$$\boldsymbol{S}_t(\boldsymbol{y}, \boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{y}, \boldsymbol{\theta})^{-1/2} \, \boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta}) . \tag{37.4}$$

In most cases, $\boldsymbol{S}_t(\boldsymbol{y}, \boldsymbol{\theta})$ is a better approximated pivotal.

Studentized EF Bootstrap

First we obtain $(z_1^*, \ldots, z_n^*)$ as in the EF bootstrap; compute

$$\boldsymbol{S}_t^* = \mathbf{V}^{*-1/2} \boldsymbol{S}^* ,$$

where $\mathbf{V}^* = n^{-1} \sum \left(z_i^* - \bar{z}^*\right)\left(z_i^* - \bar{z}^*\right)^{\mathrm{T}}$ and $\bar{z}^* = n^{-1} \sum z_i^*$, and finally solve

$$\boldsymbol{S}_t(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{S}_t^* .$$

Under fairly general conditions [37.1], the studentized EF bootstrap is second-order-accurate and also invariant under reparameterization. The simplicity of its computation is discussed in the following two subsections.

37.3.2 The Case of a Single Parameter

When the parameter θ is a scalar and $S(\boldsymbol{y}, \theta)$ is a monotonic function of θ, confidence intervals for θ based on the EF bootstrap are obtained as follows. For any specified α, we can find $S_{(\alpha)}^*$, the αth quantile of the bootstrap distribution of S^*. The two-sided interval $\left(\theta_{(\alpha/2)}^*, \theta_{(1-\alpha/2)}^*\right)$ obtained from

$$S\left(\boldsymbol{y}, \theta_{(\alpha/2)}^*\right) = S_{(\alpha/2)}^*$$
$$\text{and} \quad S\left(\boldsymbol{y}, \theta_{(1-\alpha/2)}^*\right) = S_{(1-\alpha/2)}^* ,$$

is the $100(1-\alpha)\%$ EF bootstrap confidence interval for θ. To obtain this interval, the equation $S(\boldsymbol{y}, \theta) = S^*$ needs to be solved at only two points.

We can obtain higher order accuracy by using the studentized version based on (37.4). To do this, let $S_{t(\alpha)}^*$ be the αth quantile of the distribution of S_t^*. If $S_t(\boldsymbol{y}, \theta)$ is monotonic in θ, then the equation $S_t(\boldsymbol{y}, \theta) = S_{t(\alpha)}^*$ yields an endpoint for the interval. The confidence intervals obtained using the studentized EF bootstrap are usually second-order-accurate, and their performances are comparable to those of the $\mathrm{BC_a}$ and ABC methods [37.1].

From this simple model, we can see that the EF bootstrap has several advantages over Efron's bootstrap: (i) it is often computationally simpler, because we just have to solve the equation at two points; (ii) the studentized EF bootstrap is straightforward, while the classical studentized bootstrap requires a stable estimate of the variance; (iii) the studentized statistic $S_t(\boldsymbol{y}, \theta)$ is invariant under reparameterization, as are the confidence intervals or regions based on studentized EF bootstrap. By contrast, the EF bootstrap is not invariant and it is usually first-order-accurate.

37.3.3 The Multiparameter Case

For a p-dimensional vector parameter $\boldsymbol{\theta}$, we use the approximate pivotal

$$\begin{aligned} Q(\boldsymbol{y}, \boldsymbol{\theta}) &= \boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})^{\mathrm{T}} \mathbf{V}(\boldsymbol{y}, \boldsymbol{\theta})^{-1} \boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta}) \\ &= \boldsymbol{S}_t(\boldsymbol{y}, \boldsymbol{\theta})^{\mathrm{T}} \boldsymbol{S}_t(\boldsymbol{y}, \boldsymbol{\theta}) . \end{aligned} \tag{37.5}$$

The distribution of $Q(\boldsymbol{y}, \boldsymbol{\theta})$ can be approximated by the bootstrap distribution of

$$Q^* = \boldsymbol{S}^{*\mathrm{T}} \mathbf{V}^{*-1} \boldsymbol{S}^{*\mathrm{T}} = \boldsymbol{S}_t^{*\mathrm{T}} \boldsymbol{S}_t^*$$

using the calculations described in Sect. 37.3.1.

We define q_α^* to be the αth quantile of Q^*, which is determined by $P^*(Q^* > q_\alpha^*) = \alpha$. An approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is then given by

$$C_{1-\alpha}(y) = \{\boldsymbol{\theta} : Q(y, \theta) \le q_\alpha^*\} . \tag{37.6}$$

This is based on the approximation

$$\begin{aligned} P[\boldsymbol{\theta} \in C_{1-\alpha}(y)] &= P[Q(\boldsymbol{y}, \boldsymbol{\theta}) \le q_\alpha] \\ &\approx P^*(Q^* \le q_\alpha) = 1-\alpha . \end{aligned} \tag{37.7}$$

Hu and *Kalbfleisch* [37.1] show that the confidence region in (37.7) is accurate up to order $O_p\left(n^{-3/2}\right)$. This improves on the usual χ^2 approximation, which is accurate up to order $O_p\left(n^{-1}\right)$. To construct the confidence region for a given confidence coefficient $1-\alpha$, one only needs to solve (37.6) for the relevant contour. This method is invariant under reparameterization.

The above approach does not generally work for inference on components or functions of $\boldsymbol{\theta}$. When $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$ is a differentiable function of $\boldsymbol{\theta}$, *Hu* and *Kalbfleisch* [37.1] proposed a simple method based on some projections. However, the proposed method is usually accurate up to order $O_p\left(n^{-1}\right)$, and it is not invariant under reparameterization. When $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{\theta})$ is not differentiable, one needs to use the Markov chain marginal bootstrap (MCMB) proposed by *He* and *Hu* [37.19], which is introduced in Sect. 37.4.

37.3.4 Some Examples

Example 1. Estimating the population mean. Observations $y_1, \ldots, y_n$ are made on independent and identically distributed random variables, each with an unspecified distribution function, F. Interest focuses on the mean, μ, of F which is estimated with $\hat{\mu} = \bar{y}$. In the usual classical bootstrap (Efron's bootstrap), we (i) draw the bootstrap sample $\{y_1^*, \ldots, y_n^*\}$ from $\{y_1, \ldots, y_n\}$ and (ii) calculate the bootstrap sample mean $\hat{\mu}_C^* = n^{-1} \sum y_i^*$. These steps are repeated and the empirical distribution of the $\left(\hat{\mu}_C^* - \hat{\mu}\right)$ is the bootstrap approximation to the sampling distribution of $\hat{\mu} - \mu$.

In contrast, the EF bootstrap begins with the estimating equation $\sum(y_i - \mu) = 0$, whose solution is $\hat{\mu} = \bar{y}$. The component functions $y_i - \mu$ are estimated with $z_i = y_i - \bar{y}$, $i = 1, \ldots, n$. The method proceeds as follows: (i) draw a bootstrap sample $\{z_1^*, \ldots, z_n^*\}$ from $\{z_1, \ldots, z_n\}$; (ii) calculate $S^* = n^{-1/2} \sum z_i^*$. The bootstrap distribution of S^* approximates the sampling distribution of $S(\boldsymbol{y}, \mu) = \sqrt{n}(\hat{\mu} - \mu)$. Note that if μ^* is the solution to $S(\boldsymbol{y}, \mu) = S^*$, the bootstrap distribution of $\mu^* - \hat{\mu}$ approximates the distribution of $\mu - \hat{\mu}$.

The difference between the methods is evident, even though they give, in the end, identical results. With the classical bootstrap, $\hat{\mu}_C^* - \hat{\mu}$ approximates $\hat{\mu} - \mu$, whereas in the EF procedure, $\mu^* - \hat{\mu}$ approximates $\mu - \hat{\mu}$. As a consequence, μ^* is "bias corrected". The comparison between the studentized versions is similar.

Example 2. Common mean with known and unknown variances. Suppose that $y_1, \ldots, y_n$ are from populations with $Ey_i = \mu$ and $var(y_i) = \sigma_i^2$. When σ_i^2 are known, the estimating equation,

$$\sum \frac{y_i - \mu}{\sigma_i^2} = 0$$

gives rise to the weighted least squares estimator,

$$\hat{\mu} = \left(\sum y_i/\sigma_i^2\right) / \left(\sum 1/\sigma_i^2\right) .$$

The EF and classical bootstraps can be applied to this problem in a straightforward way. (As noted above, the classical bootstrap is equivalent to the classical procedure of resampling (y_i, σ_i), $i = 1, \ldots, n$.) *Hu* and *Kalbfleisch* [37.16] compare the EF bootstrap with the classical bootstrap and the asymptotic normal approximation assuming normal and uniform errors. All methods do reasonably well, though the studentized versions of the EF and classical bootstraps do somewhat better than the other methods with abnormal errors.

Suppose there are k independent strata and in the ith stratum $y_{ij} \sim N\left(\mu, \sigma_i^2\right)$, $j = 1, \ldots, n_i$ independently, where $n_i \ge 3$ and $i = 1, \ldots, k$. The variances σ_i^2 are unknown and interest centers on the estimation of μ. This problem has received much attention in the literature [37.20–24] [Bartlett (1936), Neyman and Scott (1948), Kalbfleisch and Sprott (1970), Barndorff-Nielsen (1983). *Neyman* and *Scott* (1948) showed that the maximum likelihood estimator can be inefficient. They (and many others) proposed the estimating equation

$$\sum_{i=1}^{k} \frac{n_i(n_i-2)(\bar{y}_i - \mu)}{T_i(\mu)} = 0 ,$$

where $T_i(\mu) = \sum_{j=1}^{n_i}(y_{ij} - \mu)^2$ and $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. More generally, we could relax the condition of normal errors and still use the above equation to estimate μ.

When the number of strata k is large and the individual n_i's are small, usual inferential techniques can cause substantial difficulty. This is the case considered here, although other situations are also of interest and will be discussed elsewhere.

Let $y_i = (y_{i1}, \ldots, y_{in_i})$ and $g_i(y_i, \mu) = n_i(n_i - 2)(\bar{y}_i - \mu)/T_i(\mu)$. The estimating equation can therefore be rewritten as

$$\sum_{i=1}^{k} g_i(y_i, \mu) = 0 ,$$

and the EF bootstrap can now be applied in a straightforward manner.

37.4 Markov Chain Marginal Bootstrap

For statistical models with high-dimensional parameters, it is usually difficult to apply the bootstrap method because it is computationally intensive. For example, if one needs one minute to obtain the estimator, then one needs 1000 min to apply the bootstrap method to a sample of size $B = 1000$. To reduce the computational complexity of applying common bootstrap methods to high-dimensional parameters, *He* and *Hu* [37.19] propose the Markov chain marginal bootstrap (MCMB). In this section, we only review the MCMB for M-estimators of a linear model. Please see [37.19] for more general models and estimators.

Consider the linear regression problem $Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$, $(i = 1, \cdots, n)$ with independently and identically distributed errors e_i. An M-estimator, $\hat{\boldsymbol{\beta}}$, solves

$$n^{-1} \sum_{i=1}^{n} \psi\left(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right) \boldsymbol{x}_i = 0 \qquad (37.8)$$

for a score function ψ. In most applications, the function ψ is bounded and continuous. An important exception is the least absolute deviation estimator with $\psi(r) = \mathrm{sgn}(r)$. In this case, the equation (37.8) may not be solved exactly, but minimizing $\sum_{i=1}^{n} |Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|$ over $\boldsymbol{\beta} \in \mathbb{R}^p$ guarantees a solution so that (37.8) holds approximately.

Under some suitable conditions [37.25], the estimator $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal,

$$n^{1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to N\left(0, \left\{E\psi^2(e)/[E\psi'(e)]^2\right\}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right) ,$$

where $\mathbf{X}$ is the design matrix. A direct estimate of the variance does not always produce reliable confidence levels for inference. This is because it is difficult to estimate the constant $\left[E\psi'(e)\right]^2$ in a lot of cases.

For example, consider the minimum L_d-norm estimator that minimizes $\sum_{i=1}^{n} |y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|^d$ $(d = 1.5)$. In this case, $E\psi'(e) = 0.5E|e|^{-0.5}$. One needs to estimate the constant $E|e|^{-0.5}$ to construct a confidence interval based on the asymptotic variance. A natural estimator is the average of n absolute residuals $r_i = y_i - \boldsymbol{x}_i'\hat{\beta}_n$. When $n = 20$ and e has a standard normal distribution and the residuals resemble a random sample drawn from it, then a simple simulation shows that the average of $|r_i|^{-0.5}$ has a mean of 1.71 and standard error of 0.80. When one or a few residuals are very close to 0, the estimate could be very large. Therefore, the confidence intervals constructed from this estimated asymptotic variance would be poor.

To avoid estimating the asymptotic variance directly, one can use the usual bootstrap methods (residual bootstrap or pair-wise bootstrap). In this case, a p-dimensional nonlinear system has to be solved for each bootstrap sample. This can become a computational burden for large p. Also, the pair-wise bootstrap can be numerical unstable, because the design matrix changes for each bootstrap sample. The EF bootstrap or studentized EF bootstrap is often more stable because it uses all of the design points in each resample, but its computational complexity is no less than that of the usual bootstrap methods. When ψ is differentiable, one can solve the computational problem using projection [37.1]. However, ψ is not differentiable in a lot of cases.

MCMB overcomes the computational complexity by breaking up the p-dimensional system into p marginal (one-dimensional) equations. The algorithm proceeds as follows. Let subscript β_j be the jth component of $\boldsymbol{\beta}$ and subscript $\boldsymbol{\beta}^{(k)}$ be the kth iteration of the algorithm. Suppose that $\hat{\boldsymbol{\beta}}$ is the estimate from (37.8) and $r_i = y_i - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$ are the residuals. Let $z_i = \psi(r_i)\boldsymbol{x}_i$ be the scores. The jth component of z_i will be denoted by z_{ij} $(i = 1, \ldots, n$ and $j = 1, \ldots, p)$. For the kth iteration with $k = 0, 1, ...,$ we perform

1. For the jth component, we resample $\left\{z_{ij}^*, i = 1, ..., n\right\}$ from $\left\{z_{ij}, i = 1, \ldots, n\right\}$ without replacement.

2. Let $s_j^{(k)} = \sum_{i=1}^n z_{ij}^*$ and solve $\beta_j^{(k)}$ from

$$\sum_i^n \psi(y_i - \sum_{l=1}^{j-1} x_{il}\beta_l^{(k)} - x_{ij}\beta_j - \sum_{l=j+1}^{n} x_{il}\beta_l^{(k-1)})x_{ij} = s_j^{(k)} . \quad (37.9)$$

These two steps are performed for $j = 1, \ldots, p$.

This algorithm yields a sequence $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\beta}^{(1)}, \ldots \boldsymbol{\beta}^{(k)}, \ldots$. It is clearly a Markov chain. This method is called the Markov chain marginal bootstrap (MCMB), since a resampling process (bootstrap) is used with each marginal equation (37.9). In fact, the MCMB shares two properties with MCMC. That is, both methodologies aim to break up a high-dimensional problem into several one-dimensional ones, and both yield Markov chains as products. However, we must note that MCMB does not use any MCMC algorithm, and it is not derived from the MCMC framework.

Now we explain why the MCMB method reduces the computational complexity of the usual bootstrap method. To generate an additional variate $\beta_j^{(k)}$, one needs to resample and solve a one-dimensional equation, both of which are of the complexity $O(n)$. For $\boldsymbol{\beta}^{(k)}$, the complexity is $O(np)$ for large n and p. However, common bootstrap methods have to solve a p-dimensional system. Even the simplest system (a linear system) requires of the order of $O(np^2)$ computations. Therefore, the MCMB method reduces the computational complexity for large p. Some other studies have been discussed in [37.19].

Like the EF bootstrap, the MCMB method has another advantage; that all of the design points are used in each iteration. This leads to more reliable numerical results, especially when there are leverage points present in the data, as compared to the pairwise bootstrap method that can suffer from poor bootstrap estimates when a leverage point is excluded or duplicated in a resample.

The MCMB method can be used for the maximum likelihood estimators from general parametric models. The asymptotic validity of the MCMB method for general parametric models has been given in [37.19]. The use of MCMB for general M-estimators (or GEE estimators) is explored in [37.26].

The MCMB is usually not invariant under reparameterization. *He* and *Hu* [37.19] also show that the MCMB is first-order-accurate. However, it is unknown whether MCMB is second-order-accurate. Future research is clearly needed to understand the MCMB method.

37.5 Applications

In this section, we will apply the above bootstraps to two examples. The first example is a simple linear model. We use this example to illustrate the MCMB algorithm and show why the MCMB bootstrap works. The second example involves a linear estimating equation from L_q estimation. For more discussions of these examples, please refer to [37.1, 19].

Example 1. Simple linear model. First, we consider a simple regression model with sample size n and $p = 2$. In this special case, we have

$$n^{-1/2} \sum_{i=1}^n \left(Y_i - \boldsymbol{x}_{i1}\hat{\beta}_1^{(k)} - \boldsymbol{x}_{i2}\hat{\beta}_2^{(k-1)}\right)\boldsymbol{x}_{i1} = d_1^{(k)}$$

and

$$n^{-1/2} \sum_{i=1}^n \left(Y_i - \boldsymbol{x}_{i1}\hat{\beta}_1^{(k)} - \boldsymbol{x}_{i2}\hat{\beta}_2^{(k)}\right)\boldsymbol{x}_{i2} = d_2^{(k)} ,$$

where $d_1^{(k)} = n^{-1/2}\sum_{i=1}^n x_{i1}e_{i1}^{*(k)}$ and $d_2^{(k)} = n^{-1/2} \times \sum_{i=1}^n \boldsymbol{x}_{i2}e_{i2}^{*(k)}$, and both $e_{i1}^{*(k)}$ and $e_{i2}^{*(k)}$ are drawn independently with replacement from $r_i = Y_i - \hat{\beta}_1\boldsymbol{x}_{i1} - \hat{\beta}_2\boldsymbol{x}_{i2}$ $(i = 1, \cdots, n)$, the residuals from the parameter estimate $(\hat{\beta}_1, \hat{\beta}_2)$. Now let $s_{11} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{i1}^2$, $s_{12} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{i1}\boldsymbol{x}_{i2}$ and $s_{22} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{i2}^2$. Then the two equations can be written as

$$s_{11}n^{1/2}\left(\hat{\beta}_1 - \hat{\beta}_1^{(k)}\right) = d_1^{(k)} - s_{12}n^{1/2}\left(\hat{\beta}_2 - \hat{\beta}_2^{(k-1)}\right) ,$$

$$s_{22}n^{1/2}\left(\hat{\beta}_2 - \hat{\beta}_2^{(k)}\right) = d_2^{(k)} - s_{12}n^{1/2}\left(\hat{\beta}_1 - \hat{\beta}_1^{(k)}\right) .$$

Note that the right hand sides of the above equations are sums of two independent variables, so by using variance–covariance operation and assuming that the covariance matrix of $n^{1/2}\left(\hat{\beta} - \hat{\beta}^{(k)}\right)$ stabilizes to $V = (v_{ij})_{2\times 2}$ as $k \to \infty$, we have

$$s_{11}^2 v_{11} = s_{11}\sigma^2 + s_{12}^2 v_{22} ,$$
$$s_{22}^2 v_{22} = s_{22}\sigma^2 + s_{12}^2 v_{11} ,$$
$$s_{22}v_{12} = -s_{12}v_{11} .$$

Using some simple calculations, we can show that $\mathbf{V} = \sigma^2[(s_{ij})_{2\times 2}]^{-1}$. That is, the bootstrap variance–

covariance of $n^{1/2}\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{(k)}\right)$ stabilizes to the desired asymptotic covariance matrix for the least squares estimator.

Now we move to a real example about grade point prediction. The director of admissions of a small college administered a newly designed entrance test to eight students selected at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from their entrance test score (x).

The eight pairs of scores were: (5.5, 3.1), (4.8, 2.3), (4.7, 3.0), (5.9, 3.8), (4.1, 2.2), (4.7, 1.5), (4.5, 3.0) and (5.3, 3.6). After we fit the linear regression, we get the estimated regression line

$$\hat{Y} = -1.646 + 0.903x\,.$$

The residuals are $r = (-0.22, -0.39, 0.53, 0.12, 0.14, -1.10, 0.57, 0.45)$. For the above MCMB algorithm, we have $n = 8$, $x_{i1} = 1$ for $i = 1, \ldots, 8$, $(x_{12}, \ldots, x_{82}) = (5.5, \ldots, 5.3)$. Then we can apply the MCMB algorithm to $k = 200$ to get the 95% confidence intervals: β_1: $[-5.526, 2.134]$ and β_2: [0.138,1.668]. In this example, it is very easy to calculate the confidence intervals from other methods, but we are just using it to show how MCMB can be applied. In this simple example, there is no advantage to using the MCMB method. As we mentioned earlier, the main advantage of the MCMB is that it works well for the following two cases: (i) high-dimensional parameters, and (ii) estimating equations that are not differentiable. More complete simulation studies can be found in [37.19].

Example 2. The L_q estimation. Consider a linear estimating equation in which $g_i(y_i, \theta)$ is not differentiable with respect to θ. Such situations are quite common in nonparametric and semiparametric models [37.27, 28] and in robust regression [37.25]. Estimating functions that are not differentiable can give rise to various difficulties. Classical statistical results do not apply in general, and other methods (bootstrap methods) for confidence interval estimation are needed.

We consider the general regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, i = 1, \ldots, n\,, \quad (37.10)$$

and suppose that β is to be estimated by minimizing

$$\sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})|^{1.5}\,.$$

The corresponding estimating equation

$$\sum_{i=1}^{n} \operatorname{sgn}\left(y_i - \boldsymbol{x}_i^{\mathrm{T}}\beta\right)\, \boldsymbol{x}_i\, |y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}|^{1/2} = 0\,, \quad (37.11)$$

where $\boldsymbol{x}_i = (1, x_{1i}, x_{2i})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$. The EF bootstrap procedure for estimating the whole parameter $\boldsymbol{\beta}$ or components of $\boldsymbol{\beta}$ can be applied to this problem in a straightforward manner.

Consider a fixed design where $n = 20$, $\boldsymbol{x}_1 = (1.27, -1.10, 2.19, 0.73, -0.07, 0.42, 0.37, 0.45, -0.78, 0.76, 0.44, 1.32, -0.40, 0.33, -0.40, 0.55, 0.51, -0.11, -1.15, 1.71)$, and $\boldsymbol{x}_2 = (1.60, 1.09, -0.02, -0.83, 3.05, 0.34, -0.87, 0.45, -0.78, 0.76, 0.44, 1.32, -0.40, 0.33, -1.85, 0.69, 0.11, 1.47, 0.87, 0.12)$, and y_i are generated from (37.10) with $\beta_0 = \beta_1 = \beta_2 = 1$.

For the whole parameter vector $\boldsymbol{\beta}$, the studentized estimating function bootstrap method can be used to obtain a highly accurate $O_p\left(n^{-3/2}\right)$ confidence region. Here we just report a result based on 1000 simulations. For each simulation, we can construct a 95% confidence region for $\boldsymbol{\beta}$. Of the 1000 confidence regions, 963 confidence regions cover the true parameter $\boldsymbol{\beta} = (1, 1, 1)$.

For single parameters, the estimating function bootstrap method depends on whether the estimating function is differentiable. In this example, the estimating function is not differentiable at the point 0, so we cannot use the simple method proposed by *Hu* and *Kalbfleisch* [37.1]. In this case, the MCMB method can be used to construct the confidence interval for each component of $\boldsymbol{\beta}$.

The average confidence interval in Table 37.1 is obtained by taking the averages of the two end points of the intervals over 500 cases. We consider three methods here. NORM represents the usual confidence interval

Table 37.1 Minimum L_q distance estimator ($q = 1.5$). Simulated coverage probabilities and average confidence intervals (fixed design)

	β_0		β_1		β_2	
MCMB	90.8	[0.56, 1.43]	89.0	[0.55, 1.46]	87.6	[0.66, 1.33]
NORM	76.8	[0.63, 1.35]	76.4	[0.62, 1.38]	75.6	[0.72, 1.29]
PAIR	88.0	[0.53, 1.43]	86.0	[0.52, 1.50]	86.2	[0.62, 1.38]

based on normal approximation. PAIR represents the paired bootstrap introduced in Sect. 37.2.

For each of the 500 samples there is an estimate of β_0, β_1 and β_2. Based on these estimators, we can calculate the standard deviations, and they are 0.24, 0.28 and 0.22 respectively. The confidence intervals, constructed from these estimators using the standard formula of the average plus or minus 1.64 times the SD, are [0.56,1.42], [0.55,1.48] and [0.66,1.34], respectively. We may use these three intervals as benchmarks for the other methods under consideration. It is clear from Table 37.1 that MCMB performed well.

37.6 Discussion

We have reviewed different bootstrap methods for independent observations. However, for a lot of applications, the observations may depend on each other. For stationary processes, several bootstrap procedures have been proposed, which include the block bootstrap and others. The estimating function bootstrap can also be extended to dependent observations. *Hu* and *Kalbfleisch* [37.29] considered linear and nonlinear autoregressive models.

One important application of bootstrap is in longitudinal data analysis. In this application, a generalized estimating equation (GEE) is usually available. Within each stratum (for each patient), the observations are dependent. But the observations are independent between stratums. The estimating function bootstrap can be applied as in the common mean problem in Sect. 37.3. However, some modifications are necessary to apply the classical bootstrap procedures.

Major problems with using bootstrap for high-dimensional parameters include that it is computationally intensive and can produce unreliable numerical results. The estimating function bootstrap method solves this problem by fixing one side of the estimating equation. When the estimating function is differentiable, we can use EF bootstrap to construct confidence intervals (regions). When the estimating function is not differentiable, we can then use the MCMB to solve this problem by considering a one-dimensional equation at each step.

In applications, it is also important to choose the bootstrap sample size B appropriately. When the original process (to get the estimator, $\hat{\theta}$) does not involve intensive computation, $B = 1000$ or 2000 is recommended. In general, to estimate the variance–covariance matrix, we may only need a bootstrap sample size of 100 to 200. For confidence intervals based on quantiles, it would be better to use $B = 1000$ or 2000.

References

37.1 F. Hu, J. D. Kalbfleisch: The estimating equation bootstrap (with discussions), Can. J. Stat. **28**, 449–499 (2000)

37.2 B. Efron: Bootstrap methods: another look at the jackknife, Ann. Stat. **7**, 1–26 (1979)

37.3 B. Efron: *The jackknife, the bootstrap, and other resampling plans*, Soc. Ind. Appl. Math. CBMS-Natl. Sci. Found. Monogr., Vol. 38 (SIAM, Philadelphia 1982)

37.4 P. Hall: *The Bootstrap and Edgeworth Expansion*. (Springer, Berlin Heidelberg 1992)

37.5 B. Efron: Better bootstrap confidence intervals (with discussion), J. Am. Stat. Assoc. **82**, 171–200 (1987)

37.6 B. Efron, R. J. Tibshirani: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Stat. Sci. **1**, 54–77 (1986)

37.7 T. J. Diciccio, J. P. Romano: A review of bootstrap confidence intervals, J. Roy. Stat. Soc. **50**, 338–354 (1988)

37.8 T. J. Diciccio, B. Efron: Bootstrap confidence intervals (with discussion), Stat. Sci. **11**, 189–228 (1996)

37.9 P. Hall: Theoretical comparison of bootstrap confidence intervals (with discussion), Ann. Stat. **16**, 927–985 (1988)

37.10 W. Y. Loh: Calibrating confidence coefficients, J. Am. Stat. Assoc. **82**, 155–162 (1987)

37.11 R. Beran: Prepivoting to reduce level error of confidence sets, Biometrika **74**, 457–468 (1987)

37.12 P. Hall, M. A. Martin: Comment on paper by DiCiccio and Efron, Stat. Sci. **11**, 189–228 (1996)

37.13 D. A. Freedman: Bootstrapping regression models, Ann. Stat. **9**, 1218–1228 (1981)

37.14 F. Hu, J. V. Zidek: A bootstrap based on the estimating equations of the linear model, Biometrika **82**, 263–275 (1995)

37.15 V. P. Godambe, B. K. Kale: *Estimating Functions*, ed. by V. P. Godambe (Oxford Univ. Press, Oxford 1992) pp. 3–20

37.16 F. Hu, J. D. Kalbfleisch: Estimating equations and the bootstrap. In: *Selected Proceedings of the Symposium on Estimating Equations.*, IMS Lect. Note Monogr. Ser., Vol. 32, ed. by I. V. Basawa, V. P. Godambe, R. L. Taylor (Institute of Mathematical Statistics, Hayward 1997) pp. 405–416
37.17 R. J. Buehler: Fiducial inference. In: *Encyclopedia of Statistical Sciences*, Vol. 3, ed. by Wiley (S. Kotz and N. L. Johnson, New York 1983) pp. 76–79
37.18 M. I. Parzen, L. J. Wei, Z. Ying: A resampling method based on pivotal estimating functions, Biometrika **81**, 341–50 (1994)
37.19 X. He, F. Hu: Markov chain marginal bootstrap, J. Am. Stat. Assoc. **97**, 783–795 (2002)
37.20 D. R. Cox, N. M. Reid: Parameter orthogonality and approximate conditional inference (with discussion), J. R. Stat. Soc. **49**, 1–39 (1987)
37.21 M. S. Bartlett: The information available in small samples, Proc. Camb. Phil. Soc. **34**, 33–40 (1936)
37.22 O. E. Barndorff-Nielsen: On a formula for the distribution of a maximum likelihood estimator, Biometrika **70**, 343–365 (1983)
37.23 J. D. Kalbfleisch, D. A. Sprott: Application of likelihood methods to models involving large numbers of nuisance parameters (with discussion), J. R. Stat. Soc. B **32**, 175–208 (1970)
37.24 Neyman, Scott: Consistent estimates based on partially consistent observations, Econometrica **16**, 1–32 (1948)
37.25 P. J. Huber: *Robust Statistics*. (Wiley, New York 1981)
37.26 M. Kocherginsky: Contributions to Bootstrap-Based Inference in Linear and Nonlinear Models, Ph.D. Dissertation (2002)
37.27 R. Koenker, G. J. Bassett: Regression quantiles, Econometrica **84**, 33–50 (1978)
37.28 T. P. Hettmansperger: *Statistical Inference Based on Ranks*. (Wiley, New York 1984)
37.29 F. Hu and J. D. Kalbfleisch: *An estimating function bootstrap for linear and non-linear autoregressive models*, unpublished