As a convention, notation for random and nonrandom entities are "local" to a section where they appear, i.e., the same symbol may have different meanings in two different sections. Similarly, the numbering of conditions are "local" to a chapter. Unless otherwise mentioned, the symbols for random and nonrandom entities and the condition labels refer to their local definitions. For referring to a condition introduced in another chapter, we add the chapter number as a prefix. For example, an occurrence of Condition $5.D_r$ in Chapter 6 refers to Condition D_r of Chapter 5, etc. We use the abbreviations cdf (cumulative distribution function), CI (confidence interval), iid (independent and identically distributed), and MSE (mean squared error), as convenient. We also use a box □ to denote the end of a proof or of an example.

2
Bootstrap Methods

2.1 Introduction

In this chapter, we describe various commonly used bootstrap methods that have been proposed in the literature. Section 2.2 begins with a brief description of Efron's (1979) bootstrap method based on simple random sampling of the data, which forms the basis for almost all other bootstrap methods. In Section 2.3, we describe the famous example of Singh (1981), which points out the limitation of this resampling scheme for dependent variables. In Section 2.4, we present bootstrap methods for time-series models driven by iid variables, such as the autoregression model. In Sections 2.5, 2.6, and 2.7, we describe various block bootstrap methods. A description of the subsampling method is given in Section 2.8. Bootstrap methods based on the discrete Fourier transform of the data are described in Section 2.9, while those based on the method of sieves are presented in Section 2.10.

2.2 IID Bootstrap

In this book, we refer to the nonparametric resampling scheme of Efron (1979) , introduced in the context of "iid data," as the IID bootstrap. There are a few alternative terms used in the literature for Efron's (1979) bootstrap, such as "naive" bootstrap, "ordinary" bootstrap, etc. These terms may have a different meaning in this book, since (for example) using the IID bootstrap may not be the "naive" thing to do for data with a *dependence* structure.

We begin with the formulation of the *IID bootstrap method* of Efron (1979). For the discussion in this section, assume that $X_1, X_2, \ldots$ is a sequence of iid random variables with common distribution F. Suppose, $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ generate the data at hand and let $T_n = t_n(\mathcal{X}_n; F)$, $n \geq 1$ be a random variable of interest. Note that T_n depends on the data as well as on the underlying unknown distribution F. Typical examples of T_n include the *normalized* sample mean $T_n \equiv n^{1/2}(\bar{X}_n - \mu)/\sigma$ and the *studentized* sample mean $T_n \equiv n^{1/2}(\bar{X}_n - \mu)/s_n$ where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $s_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, $\mu = E(X_1)$, and $\sigma^2 = \text{Var}(X_1)$. Let G_n denote the sampling distribution of T_n. The goal is to find an accurate approximation to the unknown distribution of T_n or to some population characteristics, e.g., the standard error, of T_n. The bootstrap method of Efron (1979) provides an effective way of addressing these problems without any model assumptions on F.

Given $\mathcal{X}_n$, we draw a simple random sample $\mathcal{X}_m^* = \{X_1^*, \ldots, X_m^*\}$ of size m with replacement from $\mathcal{X}_n$. Thus, conditional on $\mathcal{X}_n$, $\{X_1^*, \ldots, X_m^*\}$ are iid random variables with

$$P_*(X_1^* = X_i) = \frac{1}{n}, \quad 1 \leq i \leq n ,$$

where P_* denotes the conditional probability given $\mathcal{X}_n$. Hence, the common distribution of X_i^*'s is given by the empirical distribution

$$F_n = n^{-1}\sum_{i=1}^n \delta_{X_i} ,$$

where δ_y denotes the probability measure putting unit mass at y. Usually, one chooses the resample size $m = n$. However, there are several known examples where a different choice of m is desirable. See, for example, Athreya (1987), Arcones and Giné (1989, 1991), Bickel, Götze and van Zwet (1997), Fukuchi (1994), and the references therein.

Next define the bootstrap version $T_{m,n}^*$ of T_n by replacing $\mathcal{X}_n$ with $\mathcal{X}_m^*$ and F with F_n as

$$T_{m,n}^* = t_m(\mathcal{X}_n^*; F_n) .$$

Also, let $\hat{G}_{m,n}$ denote the conditional distribution of $T_{m,n}^*$, given $\mathcal{X}_n$. Then the bootstrap principle advocates $\hat{G}_{m,n}$ as an estimator of the unknown sampling distribution G_n of T_n. If, instead of G_n, one is interested in estimating only a certain functional $\varphi(G_n)$ of the sampling distribution of T_n, then the corresponding bootstrap estimator is given by *plugging-in* $\hat{G}_{m,n}$ for G_n, i.e., the bootstrap estimator of $\varphi(G_n)$ is given by $\varphi(\hat{G}_{m,n})$. For example, if $\varphi(G_n) = \text{Var}(T_n) = \int x^2 dG_n(x) - (\int x dG_n(x))^2$, the bootstrap estimator of $\text{Var}(T_n)$ is given by $\varphi(\hat{G}_{m,n}) = \text{Var}(T_{m,n}^* \mid \mathcal{X}_n) = \int x^2 d\hat{G}_{m,n}(x) - (\int x d\hat{G}_{m,n}(x))^2$. Once the variables $\mathcal{X}_n$ have been observed, the common distribution F_n of X_i^*'s becomes known, and, hence, it is possible (at least theoretically) to find the conditional distribution $\hat{G}_{m,n}$ and the bootstrap estimator $\varphi(\hat{G}_{m,n})$ from the *knowledge* of the data. In practice, however, finding $\hat{G}_{m,n}$ *exactly* may be a daunting task, even in moderate samples. This is because the number of possible distinct values of $\mathcal{X}_m^*$ grows very rapidly, at the rate $O(n^m)$ as $n \to \infty$, $m \to \infty$ under the IID bootstrap. Consequently, the conditional distribution of $T_{m,n}^*$ is further approximated by Monte-Carlo simulations as described in Chapter 1.

To illustrate the main ideas, again consider the simplest example where $T_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, the centered and scaled sample mean. Here $\mu = EX_1$ is the level-1 parameter we want to infer about. Following the description given above, the bootstrap version $T_{m,n}^*$ of T_n based on a bootstrap sample of size m is given by

$$T_{m,n}^* = \sqrt{m}(\bar{X}_m^* - E_* X_1^*)/(\text{Var}_*(X_1^*))^{1/2}$$

where $\bar{X}_m^* = m^{-1}\sum_{i=1}^m X_i^*$ denotes the bootstrap sample mean based on $X_1^*, \ldots, X_m^*$, and E_* and Var_* respectively denote the conditional expectation and conditional variance, given $\mathcal{X}_n$. It is clear that for any $k \geq 1$,

$$E_*(X_1^*)^k = \int x^k dF_n(x) = n^{-1}\sum_{i=1}^n X_i^k . \tag{2.1}$$

In particular, this implies $E_*(X_1^*) = \bar{X}_n$, and $\text{Var}_*(X_1^*) \equiv s_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Hence, we define $T_{m,n}^*$ by replacing $\bar{X}_n$ with $\bar{X}_n^*$ and μ and σ^2 by $E_*(X_1^*)$ and $\text{Var}_*(X_1^*)$, respectively. Thus, the bootstrap version of T_n is given by

$$T_{m,n}^* = \sqrt{m}(\bar{X}_m^* - \bar{X}_n)/s_n . \tag{2.2}$$

If, for example, we are interested in estimating $\varphi_\alpha(G_n) =$ the αth quantile of T_n for some $\alpha \in (0,1)$, then the bootstrap estimator of $\varphi_\alpha(G_n)$ is $\varphi_\alpha(\hat{G}_{m,n})$, the αth quantile of the conditional distribution of $T_{m,n}^*$.

As mentioned above, determining $\hat{G}_{m,n}$ exactly is not very easy even in this simple case. However, when $EX_1^2 < \infty$, and $m = n$, we have the following result. Recall that we use the abbreviation a.s. for almost sure or almost surely, as appropriate, and we write $\Phi(\cdot)$ to denote the distribution function of the standard normal distribution on $\mathbb{R}$.

Theorem 2.1 *If $X_1, X_2, \ldots$ are iid with $\sigma^2 = Var(X_1) \in (0, \infty)$, then*

$$\sup_x |P_*(T_{n,n}^* \leq x) - \Phi(x/\sigma)| = o(1) \quad as \quad n \to \infty, \quad \text{a.s.} \tag{2.3}$$

Proof: Since $X_1^*, \ldots, X_n^*$ are iid, by the Berry-Esseen Theorem (see Theorem A.6, Appendix A)

$$\sup_x |P_*(T_{n,n}^* \leq x) - \Phi(x)| \leq (2.75)\hat{\Delta}_n , \tag{2.4}$$

where $s_n^2 = E_*(X_1^* - \bar{X}_n)^2$ and $\hat{\Delta}_n = E_*|X_1^* - \bar{X}_n|^3/(s_n^3\sqrt{n})$. Clearly, by the Strong Law of Large Numbers (SLLN) (see Theorem A.3, Appendix A),

$$s_n^2 = n^{-1}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 \to \sigma^2 \quad \text{a.s.}$$

and by the Marcinkiewicz-Zygmund SLLN (see Theorem A.4, Appendix A),

$$n^{-3/2}\sum_{i=1}^{n} |X_i|^3 \to 0 \quad \text{a.s.}$$

Hence, $\hat{\Delta}_n \to 0$ a.s. as $n \to \infty$, and Theorem 2.1 follows. □

Actually Theorem 2.1 holds for any resample size m_n that goes to infinity at a rate faster than loglog n, but the proof requires a different argument. See Arcones and Giné (1989, 1991) for details.

Note that by the Central Limit Theorem (CLT), T_n also converges in distribution to the $N(0,1)$ distribution. Hence, it follows that

$$\begin{aligned}\tilde{\Delta}_n &\equiv \sup_x |\hat{G}_{n,n}(x) - G_n(x)| \\ &= \sup_x |P_*(T_{n,n}^* \le x) - P(T_n \le x)| = o(1) \quad \text{as} \quad n \to \infty, \quad \text{a.s.}\,,\end{aligned} \tag{2.5}$$

i.e., the conditional distribution $\hat{G}_{n,n}$ of $T_{n,n}^*$ generated by the IID bootstrap method provides a valid approximation for the sampling distribution G_n of T_n. Under some additional conditions, Singh (1981) showed that

$$\tilde{\Delta}_n = O(n^{-1}(\log\log n)^{1/2}) \quad \text{as} \quad n \to \infty, \quad \text{a.s.}$$

Therefore, the bootstrap approximation for $P(T_n \le \cdot)$ is far more accurate than the classical normal approximation, which has an error of order $O(n^{-1/2})$. Similar optimality properties of the bootstrap approximation have been established in many important problems. The literature on bootstrap methods for independent data is quite extensive. By now, there exist some excellent sources that give comprehensive accounts of the theory and applications of the bootstrap methods for independent data. We refer the reader to the monographs by Efron (1982), Hall (1992), Mammen (1992), Efron and Tibshirani (1993), Barbe and Bertail (1995), Shao and Tu (1995), Davison and Hinkley (1997), and Chernick (1999) for the bootstrap methodology for independent data. Here, we have described Efron's (1979) bootstrap for iid data mainly as a prelude to the bootstrap methods for dependent data considered in later sections, as the basic principles in both cases are the same. Furthermore, it provides a historical account of the developments that culminated in formulation of the bootstrap methods for dependent data.

2.3 Inadequacy of IID Bootstrap for Dependent Data

The IID bootstrap method of Efron (1979), being very simple and general, has found application to a hoard of statistical problems. However, the general perception that the bootstrap is an "omnibus" method, giving accurate results in all problems automatically, is misleading. A prime example of this appears in the seminal paper by Singh (1981), which in addition to providing the first theoretical confirmation of the superiority of the IID bootstrap, also pointed out its inadequacy for dependent data.

In this section we consider the aforementioned example of Singh (1981). Suppose $X_1, X_2, \ldots$ is a sequence of m-dependent random variables with $EX_1 = \mu$ and $EX_1^2 < \infty$. Recall that $\{X_n\}_{n\ge1}$ is called *m-dependent* for some integer $m \ge 0$ if $\{X_1, \ldots X_k\}$ and $\{X_{k+m+1}, \ldots\}$ are independent for all $k \ge 1$. Thus, an iid sequence of random variables $\{\epsilon_n\}_{n\ge1}$ is 0-dependent and if we define $X_n = \epsilon_n + 0.5\epsilon_{n+1}$, $n \ge 1$, with this iid sequence $\{\epsilon_n\}_{n\ge1}$, then $\{X_n\}_{n\ge1}$ is 1-dependent.

Next, let $\sigma_m^2 = \text{Var}(X_1) + 2\sum_{i=1}^{m-1}\text{Cov}(X_1, X_{1+i})$ and $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. If $\sigma_m^2 \in (0,\infty)$, then by the CLT for m-dependent variables (cf. Theorem A.7, Appendix A),

$$\sqrt{n}(\bar{X}_n - \mu) \longrightarrow^d N(0, \sigma_m^2)\,, \tag{2.6}$$

where $\longrightarrow^d$ denotes convergence in distribution. Now, suppose that we want to estimate the sampling distribution of the random variable $T_n = \sqrt{n}(\bar{X}_n - \mu)$ using the IID bootstrap. For simplicity, assume that the resample size equals the sample size, i.e., from $\mathcal{X}_n = (X_1, \ldots, X_n)$, an equal number of bootstrap variables $X_1^*, \ldots, X_n^*$ are generated. Then, the bootstrap version $T_{n,n}^*$ of T_n is given by

$$T_{n,n}^* = \sqrt{n}(\bar{X}_n^* - \bar{X}_n)\,,$$

where $\bar{X}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*$. The conditional distribution of $T_{n,n}^*$ under the IID bootstrap method still converges to a normal distribution, but with a "wrong" variance, as shown below.

Theorem 2.2 *Suppose $\{X_n\}_{n\ge1}$ is a sequence of stationary m-dependent random variables with $EX_1 = \mu$, and $\sigma^2 = Var(X_1) \in (0,\infty)$. Then*

$$\sup_x |P_*(T_{n,n}^* \le x) - \Phi(x/\sigma)| = o(1) \quad \textit{as} \quad n \to \infty, \quad \text{a.s.} \tag{2.7}$$

Proof: Note that conditional on $\mathcal{X}_n$, $X_1^*, \ldots, X_n^*$ are iid random variables. As in the proof of Theorem 2.1, by the Berry-Esseen Theorem, it is enough to show that

$$s_n^2 \to \sigma^2 \quad \text{as} \quad n \to \infty \quad \text{a.s.}$$

and

$$n^{-3/2} \sum_{i=1}^{n} |X_i|^3 \to 0 \quad \text{as} \quad n \to \infty, \quad \text{a.s.}$$

These follow easily from the following lemma. Hence Theorem 2.2 is proved. □

Lemma 2.1 *Let $\{X_n\}_{n \geq 1}$ be a sequence of stationary m-dependent random variables. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function with $E|f(X_1)|^p < \infty$ for some $p \in (0, \infty)$, and that $Ef(X_1) = 0$ if $p \geq 1$. Then,*

$$n^{-1/p} \sum_{i=1}^{n} f(X_i) \to 0 \quad as \quad n \to \infty, \quad \text{a.s.}$$

Proof: This is most easily proved by splitting the given m-dependent sequence $\{X_n\}_{n \geq 1}$ into $m+1$ iid subsequences $\{Y_{ji}\}_{i \geq 1}$, $j = 1, \ldots, m+1$, defined by $Y_{ji} = X_{j+(i-1)(m+1)}$, and then applying the standard results for iid random variables to $\{Y_{ji}\}_{i \geq 1}$'s (cf. Liu and Singh (1992)). For $1 \leq j \leq m+1$, let $I_j \equiv I_{jn} = \{1 \leq i \leq n : j + (i-1)(m+1) \leq n\}$ and let $N_j \equiv N_{jn}$ denote the size of the set I_j. Note that $N_j/n \to (m+1)^{-1}$ as $n \to \infty$ for all $1 \leq j \leq m+1$. Then, by the Marcinkiewicz-Zygmund SLLN (cf. Theorem A.4, Appendix A) applied to each of the sequence of iid random variables $\{Y_{ji}\}_{i \geq 1}$, $j = 1, \ldots, m+1$, we get

$$n^{-1/p} \sum_{i=1}^{n} f(X_i) = \sum_{j=1}^{m+1} \left[N_j^{-1/p} \sum_{i \in I_j} f(Y_{ji}) \right] \cdot (N_j/n)^{1/p} \to 0 \text{ as } n \to \infty, \quad \text{a.s.}$$

This completes the proof of Lemma 2.1. □.

Corollary 2.1 *Under the conditions of Theorem 2.2, if $\sum_{i=1}^{m} Cov(X_1, X_{1+i}) \neq 0$ and $\sigma_\infty^2 \neq 0$, then for any $x \neq 0$,*

$$\lim_{n \to \infty} [P_*(T_{n,n}^* \leq x) - P(T_n \leq x)] = [\Phi(x/\sigma) - \Phi(x/\sigma_\infty)] \neq 0 \quad \text{a.s.}$$

Proof: Follows from Theorem 2.2 and (2.6). □

Thus, for all $x \neq 0$, the IID bootstrap estimator $P_*(T_{n,n}^* \leq x)$ of the level-2 parameter $P(T_n \leq x)$ has a mean squared error that tends to a *nonzero* number in the limit and the bootstrap estimator of $P(T_n \leq x)$ is not consistent. Therefore, the IID bootstrap method fails drastically for *dependent* data. It follows from the proof of Theorem 2.2 that resampling individual X_i's from the data $\mathcal{X}_n$ ignores the dependence structure of the sequence $\{X_n\}_{n \geq 1}$ completely, and thus, fails to account for the lag-covariance terms (viz., $\text{Cov}(X_1, X_{1+i})$, $1 \leq i \leq m$) in the asymptotic variance.

Following this result, there have been several attempts in the literature to extend the IID bootstrap method to the dependent case. In the next section, we first look at extensions of this method to certain dependent models generated by iid random variables. More general resampling schemes (such as the block bootstrap and the frequency domain bootstrap methods), which are applicable without any parametric model assumptions, have been put forward in the literature much later. These are presented in Sections 2.5–2.10.

2.4 Bootstrap Based on IID Innovations

Suppose $\{X_n\}_{n \geq 1}$ is a sequence of random variables satisfying the equation

$$X_n = h(X_{n-1}, \ldots, X_{n-p}; \beta) + \epsilon_n \ , \tag{2.8}$$

$n > p$, where β is a $q \times 1$ vector of parameters, $h : \mathbb{R}^{p+q} \to \mathbb{R}$ is a known Borel measurable function, and $\{\epsilon_n\}_{n > p}$ is a sequence of iid random variables with common distribution F that are independent of the random variables $X_1, \ldots, X_p$. For identifiability of the model (2.8), assume that $E\epsilon_1 = 0$. A commonly used example of model (2.8) is the autoregressive process of order p (cf. (2.9) below). Noting that the process $\{X_n\}_{n \geq 1}$ is driven by the innovations ϵ_i's that are iid, the IID bootstrap method can be easily extended to the dependent model (2.8).

As before, suppose that $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ denotes the sample and that we want to approximate the sampling distribution of a random variable $T_n = t_n(\mathcal{X}_n; F, \beta)$. Let $\hat{\beta}_n$ be an estimator, e.g., the least squares estimator, of β based on $\mathcal{X}_n$. Define the residuals

$$\hat{\epsilon}_i = X_i - h(X_{i-1}, \ldots, X_{i-p}; \hat{\beta}_n), \quad p < i \leq n.$$

Note that, in general,

$$\bar{\epsilon}_n \equiv (n-p)^{-1} \sum_{i=1}^{n-p} \hat{\epsilon}_{i+p} \neq 0 \ .$$

Hence, we center the "raw" residuals $\hat{\epsilon}_i$'s and define the "centered" residuals

$$\tilde{\epsilon}_i = \hat{\epsilon}_i - \bar{\epsilon}_n, \quad p < i \leq n \ .$$

Without such a centering, the resulting bootstrap approximation often has a random bias that does not vanish in the limit and renders the approximation useless. (See, for example, Freedman (1981), Shorack (1982), and Lahiri (1992b) that treat a similar bias phenomenon in regression problems.)

Next draw a simple random sample $\epsilon_{p+1}^*, \ldots, \epsilon_m^*$ of size $(m-p)$ from $\{\tilde{\epsilon}_i : p < i \leq n\}$ with replacement and define the bootstrap pseudo-observations, using the model structure (2.8), as:

$$X_i^* = X_i \quad \text{for} \quad i = 1, \ldots, p, \quad \text{and}$$

$$X_i^* = h(X_{i-1}^*, \ldots, X_{i-p}^*; \hat{\beta}_n) + \epsilon_i^*, \quad p < i \le m .$$

Note that by construction ϵ_i^*, $p < i \le m$ are iid and $E_* \epsilon_1^* = 0$. The bootstrap version of the random variable $T_n = t_n(\mathcal{X}_n; F, \beta)$ is defined as

$$T_{m,n}^* = t_m(\mathcal{X}_m^*; F_n, \hat{\beta}_n) ,$$

where $\mathcal{X}_m^* = \{X_1^*, \ldots, X_m^*\}$ and F_n denotes the empirical distribution of the centered residuals $\tilde{\epsilon}_i$, $p < i \le n$. The sampling distribution of T_n is approximated by the conditional distribution of $T_{m,n}^*$ given $\mathcal{X}_n$. For certain time-series models satisfying (2.8), different versions of this resampling scheme have been proposed by Freedman (1984), Efron and Tibshirani (1986), Swanepoel and van Wyk (1986), and Kreiss and Franke (1992). The IID-innovation-bootstrap method can be applied with some simple modifications to popular parametric models for spatial data as well (e.g., the spatial autoregression model); see Chapter 7, Cressie (1993).

A special case of model (2.8) is the autoregression model of order p (AR(p)), given by

$$X_n = \beta_1 X_{n-1} + \ldots + \beta_p X_{n-p} + \epsilon_n, \; n > p , \qquad (2.9)$$

where $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of autoregressive parameters, and $\{\epsilon_n\}_{n>p}$ is an iid sequence satisfying the requirements of model (2.8). For AR(p)-models, validity and the rate of approximation of the IID-Innovation bootstrap have been well-studied in the literature. When the sequence $\{X_n\}_{n \ge 1}$ is *stationary*, Bose (1988) shows that under suitable regularity conditions, a version of the IID-innovation bootstrap approximation to the sampling distribution of the standardized least square estimator is more accurate than the normal approximation. For *nonstationary* cases, performance of this method has been studied by Basawa, Mallik, McCormick and Taylor (1989), Basawa, Mallik, McCormick, Reeves and Taylor (1991), Datta (1995, 1996), Datta and Sriram (1997), and Heimann and Kreiss (1996), among others. It follows from their work that the IID-innovation bootstrap method is very sensitive to the values of the autoregression parameter vector β. Indeed, if the value of β is such that the roots of the characteristic equation $z^p + \beta_1 z^{p-1} + \ldots + \beta_p = 0$ lie on the unit circle, then the IID-innovation bootstrap fails. Because of its dependence on the validity of the model (2.9), and drastic change in the performance with a small change in the parameter value, one needs to be particularly careful when applying the IID-innovation bootstrap method. Properties of the IID-innovation bootstrap and related model based bootstrap methods are described in Chapter 8.

2.5 Moving Block Bootstrap

Bootstrap methods described in the previous sections are applicable either under the hypothesis of independence or under specific model assumptions for dependent data. The main idea in the latter case is to use the approximate independence of the residuals, and then apply the resampling scheme of the IID-bootstrap method to get the right approximation. In a problem where the statistician does not have enough prior knowledge to specify such models, these methods are not very useful. In a significant breakthrough, Künsch (1989) and Liu and Singh (1992) independently formulated a substantially new resampling scheme, called the *moving block bootstrap* (MBB), that is applicable to dependent data without any parametric model assumptions. In contrast to resampling a *single* observation at a time, as has been commonly done under the earlier formulations of the bootstrap, the MBB resamples *blocks* of (consecutive) observations at a time. As a result, the dependence structure of the original observations is preserved within each block. Furthermore, the common length of the blocks increases with the sample size. As a result, when the data are generated by a *weakly* dependent process, the MBB reproduces the underlying dependence structure of the process *asymptotically*. Essentially the same principle was put forward by Hall (1985) in the context of bootstrapping spatial data and by Carlstein (1986) for estimating the variance of a statistic based on time series data. A description of Carlstein's method will be given in the next section. We now turn to a description of the MBB.

Let $X_1, X_2, \ldots$ be a sequence of stationary random variables, and let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ denote the observations. We shall define the MBB version of estimators of the form $\hat{\theta}_n = T(F_n)$, where F_n denotes the empirical distribution function of $X_1, \ldots, X_n$, and where $T(\cdot)$ is a (real-valued) functional of F_n. Suppose $\ell \equiv \ell_n \in [1, n]$ is an integer. For dependent data, we typically require that

$$\ell \to \infty \quad \text{and} \quad n^{-1}\ell \to 0 \quad \text{as} \quad n \to \infty .$$

However, a description of the MBB can be given without this restriction. Let $\mathcal{B}_i = (X_i, \ldots, X_{i+\ell-1})$ denote the block of length ℓ starting with X_i, $1 \le i \le N$ where $N = n - \ell + 1$. (See Figure 2.1 below.) To obtain the MBB samples, we randomly select a suitable number of blocks from the collection $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$. Accordingly, let $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$ denote a simple random sample drawn with replacement from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$. Note that each of the selected blocks contains ℓ elements. Denote the elements in $\mathcal{B}_i^*$ by $(X_{(i-1)\ell+1}^*, \ldots, X_{i\ell}^*)$, $i = 1, \ldots, k$. Then, $X_1^*, \ldots, X_m^*$ constitute the MBB sample of size $m \equiv k\ell$. The MBB version $\theta_{m,n}^*$ of $\hat{\theta}_n$ is defined as

$$\theta_{m,n}^* = T(F_{m,n}^*) ,$$

where $F_{m,n}^*$ denotes the empirical distribution of $(X_1^*, \ldots, X_m^*)$.

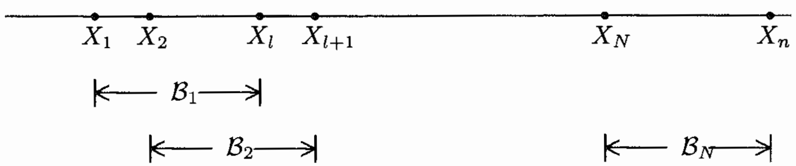

FIGURE 2.1. The collection $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ of overlapping blocks under the MBB.

An alternative formulation of the MBB can be given as follows. Note that selecting the blocks $\mathcal{B}_i^*$'s randomly from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ is equivalent to selecting k indices at random from the set $\{1, \ldots, N\}$. Accordingly, let $I_1, \ldots, I_k$ be iid random variables with the discrete uniform distribution on $\{1, \ldots, N\}$. If we set $\mathcal{B}_i^* = \mathcal{B}_{I_i}$ for $i = 1, \ldots, k$, then $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$ represent a simple random sample drawn with replacement from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$. The bootstrap sample $X_1^*, \ldots, X_m^*$ can be defined using the resampled blocks $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$ as before. Note that conditional on the data $\mathcal{X}_n$, the resampled blocks of observations $(X_1^*, \ldots, X_\ell^*)', (X_{\ell+1}^*, \ldots, X_{2\ell}^*)', \ldots, (X_{(k-1)\ell+1}^*, \ldots, X_{k\ell}^*)'$ are iid ℓ-dimensional random vectors with

$$\begin{aligned} &P_*((X_1^*, \ldots, X_\ell^*)' = (X_j, \ldots, X_{j+\ell-1})') \\ &= P_*(I_1 = j) \\ &= N^{-1}, \quad \text{for} \quad 1 \le j \le N, \end{aligned} \tag{2.10}$$

where P_* denotes the conditional probability given $\mathcal{X}_n$. In the special case, when each block consists of a single element (i.e., $\ell = 1$), then by (2.10), $X_1^*, \ldots, X_m^*$ are iid with the common distribution F_n, and hence, the MBB reduces to the IID bootstrap method of Efron (1979) described in Section 2.2. For $\ell > 1$, the ℓ-dimensional joint distribution of the underlying process $\{X_n\}_{n\ge1}$ is preserved *within* the resampled blocks. Since ℓ tends to infinity with n, any finite-dimensional joint distribution of $\{X_n\}_{n\ge1}$-process at a given number of finite lag distances can be eventually recovered from the resampled values. As a result, the MBB can effectively capture those characteristics of the underlying process $\{X_n\}_{n\ge1}$ that are determined by the dependence structure of the observations at short lags.

As in the case of the IID bootstrap, the MBB sample size is typically chosen to be of the same order as the original sample size. If b_1 denotes the smallest integer such that $b_1\ell \ge n$, then one may select $k = b_1$ blocks to generate the MBB samples, and use only the first n values to define the bootstrap version of T_n. However, there are some inference problems where a smaller sample size works better (cf. Chapter 11).

Though estimators of the form $\hat\theta_n = T(F_n)$ considered above include many commonly used estimators, e.g., the sample mean, M-estimators of location and scale, von Mises functionals, etc., they are not sufficiently rich for applications in the time series context. This is primarily because $\hat\theta_n$ above depends only on the *one-dimensional* marginal empirical distribution F_n, and hence does not cover standard statistics like the sample lag correlations, or the spectral density estimators. We shall now consider a more general version of the MBB that covers such statistics.

Given the observations $\mathcal{X}_n$, let $F_{p,n}$ denote the p-dimensional empirical measure

$$F_{p,n} = (n-p+1)^{-1} \sum_{j=1}^{n-p+1} \delta_{Y_j} \; ,$$

where $Y_j = (X_j, \ldots, X_{j+p-1})$ and where for any $y \in \mathbb{R}^p$, δ_y denotes the probability measure on $\mathbb{R}^p$ putting unit mass on y. The general version of the MBB concerns estimators of the form

$$\hat\theta_n = T(F_{p,n}) \; , \tag{2.11}$$

where $T(\cdot)$ is now a functional defined on a (rich) subset of the set of all probability measures on $\mathbb{R}^p$. Here, $p \ge 1$ may be a fixed integer, or it may tend to infinity with n suitably. Some important examples of (2.11) are given below.

Example 2.1: A version of the sample lag covariance of order $k \ge 0$ is given by

$$\hat\gamma_n(k) = (n-k)^{-1} \sum_{j=1}^{n-k} (X_{j+k} - \bar X_{n,k})(X_j - \bar X_{n,k}) \; ,$$

where $\bar X_{n,k} = (n-k)^{-1} \sum_{j=1}^{n-k} X_j$. Then, $\hat\gamma_n(k)$ is of the form (2.11) with $p = k+1$. □

Example 2.2: Let ψ be a function from $\mathbb{R}^p \times \mathbb{R}^k$ into $\mathbb{R}^k$ such that

$$E\psi(X_1, \ldots, X_p; \theta) = 0.$$

Here, θ is a functional of the p-dimensional joint distribution of $(X_1, \ldots, X_p)$, implicitly defined by the equation above. A *generalized M-estimator* of the parameter $\theta \in \mathbb{R}^k$ is defined (cf. Bustos (1982)) as a solution of the equation

$$\sum_{j=1}^{n-p+1} \psi(X_j, \ldots, X_{j+p-1}; T_n) = 0.$$

The generalized M-estimators can also be expressed in the form (2.11). □

Example 2.3: Let $f(\cdot)$ denote the spectral density of the process $\{X_n\}_{n\geq 1}$. Then, a lag-window estimator of the spectral density (cf., Chapter 6, Priestley (1981)) is given by

$$\hat{f}_n(\lambda) = \sum_{k=-(n-1)}^{(n-1)} w(k/p)\hat{\gamma}_n(k)\cos(k\lambda),\ \ \lambda \in [0, \pi],$$

where $p \equiv p_n$ tends to infinity at a rate slower than n and where w is a weight function such that $w(0) = (2\pi)^{-1}$ and w vanishes outside the interval $(-1, 1)$. For different choices of w, one gets various commonly used estimators of the spectral density, such as the truncated periodogram estimator, the Bartlett estimator, etc. Since $\hat{f}_n$ is a function of $\hat{\gamma}_n(0), \ldots, \hat{\gamma}_n(p)$, from Example 2.1, it follows that we can express it in the form (2.11). Note that in this example, p tends to infinity with n. □

To define the MBB version of $\hat{\theta}_n$ in (2.11), fix a block size ℓ, $1 < \ell < n - p + 1$, and define the blocks in terms of Y_i's as

$$\tilde{\mathcal{B}}_j = (Y_j, \ldots, Y_{j+\ell-1}),\ \ 1 \leq j \leq n - p - \ell + 2 .$$

For $k \geq 1$, select k blocks randomly from the collection $\{\tilde{\mathcal{B}}_i : 1 \leq i \leq n-p-\ell+2\}$ to generate the MBB observations $Y_1^*, \ldots, Y_\ell^*; Y_{\ell+1}^*, \ldots, Y_{2\ell}^*; \cdots, Y_m^*$, where $m = k\ell$. The MBB version of (2.11) is now defined as

$$\theta^*_{m,n} = T(\tilde{F}^*_{m,n}) , \tag{2.12}$$

where $\tilde{F}^*_{m,n} \equiv m^{-1}\sum_{j=1}^m \delta_{Y_j^*}$ denotes the empirical distribution of $Y_1^*, \ldots, Y_m^*$. Thus, for estimators of the form (2.11), the MBB version is defined by resampling from blocks of Y-values *instead* of blocks of X-values themselves. This formulation of the MBB was initially given by Künsch (1989) and was further explored by Politis and Romano (1992a). Clearly, the definition (2.12) applies to both the cases where p is fixed and where p tends to infinity with n. In the latter case, Politis and Romano (1992a) called the modified blocking mechanism as the "blocks of blocks" bootstrap, and gave a more general formulation that allows one to control the amount of overlap between the successive blocks of Y-values. We refer the reader to Politis and Romano (1992a) for the other versions of the "blocks of blocks" bootstrap method.

Note that for the more general class of statistics $\hat{\theta}_n$ given by (2.11) for some $p \geq 2$, there is an alternative way of defining the bootstrap version of $\hat{\theta}_n$. Since the estimator $\hat{\theta}_n$ can always be expressed as a function of the given observations $X_1, \ldots, X_n$, one may define the bootstrap version of $\hat{\theta}_n$ by resampling from $X_1, \ldots, X_n$ *directly*. Specifically, suppose that the

block bootstrap observations $X_1^*, \ldots, X_m^*$ are generated by resampling from the blocks $\mathcal{B}_i = \{X_i, \ldots, X_{i+\ell-1}\}$, $i = 1, \ldots, N$ of X-values. Then, define bootstrap "analogs" of the p-dimensional variable $Y_i \equiv (X_i, \ldots, X_{i+p-1})'$ in terms of $X_1^*, \ldots, X_m^*$ as $Y_i^{**} \equiv (X_i^*, \ldots, X_{i+p-1}^*)'$, $i = 1, \ldots, m - p + 1$. Then, the bootstrap version of $\hat{\theta}_n$ under this alternative approach is defined as

$$\theta^{**}_{m,n} = T(\tilde{F}^{**}_{m,n}) ,$$

where $\tilde{F}^{**}_{m,n} = \sum_{i=1}^{m-p+1} \delta_{Y_i^{**}}$. We call this approach of defining the moving block bootstrap version of $\hat{\theta}_n$ as the "naive" approach, and the other approach leading to $\theta^*_{m,n}$ in (2.12) as the "ordinary" approach of the MBB. We shall also use the terms "naive" and "ordinary" in the context of bootstrapping estimators of the form (2.11) using *other* block bootstrap methods described later in this chapter.

For a comparison of the two approaches, suppose that $\{X_n\}_{n\geq 1}$ is a sequence of stationary random variables. Then, for each i, the random vector $Y_i = (X_i, \ldots, X_{i+p-1})'$ has the same distribution as $(X_1, \ldots, X_p)'$, and hence, the resampled vectors Y_i^* under the "ordinary" approach always retains the dependence structure of $(X_1, \ldots, X_p)'$. However, when the bootstrap blocks are selected by the "naive" approach, the bootstrap observations X_i^*'s, that are at lags less than p and that lie near the boundary of two adjacent resampled blocks $\mathcal{B}_j^*$ and $\mathcal{B}_{j+1}^*$, are *independent.* Thus the components of Y_i^{**} under the "naive" approach do *not* retain the dependence structure of $(X_1, \ldots, X_p)'$. As a result, the naive approach introduces additional bias in the bootstrap version $\theta^{**}_{m,n}$ of $\hat{\theta}_n$. We shall, therefore, always use the "ordinary" form of a block bootstrap method while defining the bootstrap version of estimators $\hat{\theta}_n$ given by (2.11). For a numerical example comparing the naive and the ordinary versions of the MBB and certain other block bootstrap methods, see Section 4.5.

We conclude this section with two remarks. First, it is easy to see that the above description of the MBB and the "blocks of blocks" bootstrap applies almost verbatim if, to begin with, the observations $X_1, \ldots, X_n$ were random *vectors* instead of random variables. Second, performance of a MBB estimator critically depends on the block size ℓ. Since the sampling distribution of a given estimator typically depends on the *joint* distribution of $X_1, \ldots, X_n$, the block size ℓ must *grow to infinity* with the sample size n to capture the dependence structure of the series $\{X_n\}_{n\geq 1}$, eventually. Typical choices of ℓ are of the form $\ell = Cn^\delta$ for some constants $C > 0$, $\delta \in (0, 1/2)$. For more on properties of MBB estimators and effects of block lengths on their performance, see Chapters 3–7.

2.6 Nonoverlapping Block Bootstrap

In this section, we consider the blocking rule due to Carlstein (1986). For simplicity, here we shall consider estimators given by (2.11) with $p = 1$ only. Extension to the case of a general $p \geq 1$ is straightforward. The key feature of Carlstein's blocking rule is to use nonoverlapping segments of the data to define the blocks. The corresponding block bootstrap method will be called the nonoverlapping block bootstrap (NBB). Suppose that $\ell \equiv \ell_n \in [1, n]$ is an integer and $b \geq 1$ is the largest integer satisfying $\ell b \leq n$. Then, define the blocks

$$\mathcal{B}_i^{(2)} = (X_{(i-1)\ell+1}, \ldots, X_{i\ell})', \quad i = 1, \ldots, b\,.$$

(Here we use the index "2" in the superscipt to denote the blocks for the NBB resampling scheme. We reserve the index 1 for the MBB and we shall use the indices 3, 4, etc. for the other block bootstrap methods described later.) Note that while the blocks in the MBB overlap, the blocks $\mathcal{B}_i^{(2)}$'s under the NBB do not. See Figure 2.2. As a result, the collection of blocks from which the bootstrap blocks are selected is smaller than the collection for the MBB.

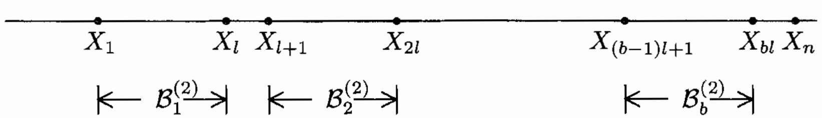

FIGURE 2.2. The collection $\{\mathcal{B}_1^{(2)}, \ldots, \mathcal{B}_b^{(2)}\}$ of nonoverlapping blocks under Carlstein's (1986) rule.

The next step in implementing the NBB is exactly the same as that for the MBB. We select a simple random sample of blocks $\mathcal{B}_1^{*(2)}, \ldots, \mathcal{B}_k^{*(2)}$ with replacement from $\{\mathcal{B}_1^{(2)}, \ldots, \mathcal{B}_b^{(2)}\}$ for some suitable integer $k \geq 1$. With $m = k\ell$, let $F_{m,n}^{*(2)}$ denote the empirical distribution of the bootstrap sample $(X_{2,1}^*, \ldots, X_{2,\ell}^*; \ldots; X_{2,\{(b-1)\ell+1\}}^*, \ldots, X_{2,m}^*)$, obtained by writing the elements of $\mathcal{B}_1^{*(2)}, \ldots, \mathcal{B}_k^{*(2)}$ in a sequence. Then, the bootstrap version of an estimator $\hat{\theta}_n = T(F_n)$ is given by

$$\theta_{m,n}^{*(2)} = T(F_{m,n}^{*(2)})\,. \tag{2.13}$$

Even though the definition of the bootstrapped estimators are very similar for the MBB and for the NBB, the resulting bootstrap versions $\theta_{m,n}^*$ and $\theta_{m,n}^{*(2)}$ have different distributional properties. We illustrate the point with the simplest case, where $\hat{\theta}_n = n^{-1}\sum_{j=1}^n X_j$ is the sample mean. The

bootstrap version of $\hat{\theta}_n$ under the two methods are respectively given by

$$\theta_{m,n}^* = m^{-1}\sum_{j=1}^m X_j^*, \quad \text{and} \quad \theta_{m,n}^{*(2)} = m^{-1}\sum_{j=1}^m X_{2,j}^*\,.$$

From (2.10), we get

$$\begin{aligned} E_*(\theta_{m,n}^*) &= E_*(\ell^{-1}\sum_{i=1}^{\ell} X_i^*) \\ &= N^{-1}\sum_{j=1}^{N}\Big(\ell^{-1}\sum_{i=1}^{\ell} X_{j+i-1}\Big) \\ &= N^{-1}\Big\{n\bar{X}_n - \ell^{-1}\sum_{j=1}^{\ell-1}(\ell-j)(X_j + X_{n-j+1})\Big\}. \end{aligned} \tag{2.14}$$

To obtain a similar expression for $E_*(\theta_{m,n}^{*(2)})$, note that under the NBB, the bootstrap variables $(X_{2,1}^*, \ldots, X_{2,\ell}^*), \ldots, (X_{2,(m-\ell+1)}^*, \ldots, X_{2,m}^*)$ are iid, with common distribution

$$P_*\Big((X_{2,1}^*, \ldots, X_{2,\ell}^*) = (X_{(j-1)\ell+1}, \ldots, X_{j\ell})\Big) = 1/b \tag{2.15}$$

for $j = 1, \ldots, b$. Hence,

$$\begin{aligned} E_*(\theta_{m,n}^{*(2)}) &= E_*(\ell^{-1}\sum_{i=1}^{\ell} X_{2,i}^*) \\ &= b^{-1}\sum_{j=1}^{b}\Big(\ell^{-1}\sum_{i=1}^{\ell} X_{(j-1)\ell+i}\Big) \\ &= (b\ell)^{-1}\Big\{n\bar{X}_n - \sum_{i=b\ell+1}^{n} X_i\Big\}\,, \end{aligned} \tag{2.16}$$

which equals $\bar{X}_n$ if n is a multiple of ℓ. Thus, the bootstrapped estimators have different (conditional) means under the two methods. However, note that if the process $\{X_n\}_{n\geq 1}$ satisfies some standard moment and mixing conditions, then $E\{E_*(\theta_{m,n}^*) - E_*\theta_{m,n}^{*(2)}\}^2 = O(\ell/n^2)$. Hence the difference between the two is negligible for large sample sizes.

2.7 Generalized Block Bootstrap

As follows from its description (cf. Section 2.5), the MBB resampling scheme suffers from an undesirable boundary effect as it assigns lesser weights to the observations toward the beginning and the end of the data set than to the middle part. Indeed, for $\ell \le j \le n-\ell$, the jth observation X_j appears in exactly ℓ of the blocks $\{\mathcal{B}_1,\ldots,\mathcal{B}_N\}$, whereas for $1 \le j \le \ell-1$, X_j and X_{n-j+1} appear only in j blocks. Since there is no observation beyond X_n (or prior to X_1), we cannot define new blocks to get rid of this boundary effect. A similar problem also exists under the NBB with the observations near the end of the data sequence when n is not a multiple of ℓ. Politis and Romano (1992b) suggested a simple way out of this boundary problem. Their idea is to wrap the data around a circle and form additional blocks using the "circularly defined" observations. Politis and Romano (1992b, 1994b) put forward two resampling schemes based on circular blocks, called the "circular block bootstrap" (CBB) and the "stationary bootstrap" (SB). Here we describe a generalization of their idea and formulate the generalized block bootstrap method, which provides a unified framework for describing different block bootstrap methods, including the CBB and the SB.

Given the variables $\mathcal{X}_n = \{X_1,\ldots,X_n\}$, first we define a new time series $Y_{n,i}$, $i \ge 1$ by *periodic extension*. Note that for any $i \ge 1$, there are integers $k_i \ge 0$ and $j_i \in [1,n]$ such that $i = k_i n + j_i$. Then, $i = j_i$ (modulo n). We define the variables $Y_{n,i}$, $i \ge 1$ by the relation $Y_{n,i} = X_{j_i}$. Note that this is equivalent to writing the variables $X_1,\ldots,X_n$ repeatedly on a line and labeling them serially as $Y_{n,i}$, $i \ge 1$. See Figure 2.3.

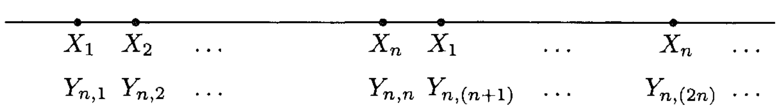

FIGURE 2.3. The periodically extended time series $Y_{n,i}$, $i \ge 1$.

Next define the blocks

$$\mathcal{B}(i,j) = (Y_{n,i},\ldots,Y_{n,(i+j-1)})$$

for $i \ge 1$, $j \ge 1$. Let Γ_n be a transition probability function on the set $\mathbb{R}^n \times \bigotimes_{t=1}^{\infty}(\{1,\ldots,n\}\times\mathbb{N})$, i.e., for each $x \in \mathbb{R}^n$, $\Gamma_n(x;\cdot)$ is a probability measure on $\bigotimes_{t=1}^{\infty}(\{1,\ldots,n\}\times\mathbb{N}) \equiv \Big\{\{i_t,\ell_t\}_{t=1}^{\infty} : 1 \le i_t \le n, 1 \le \ell_t < \infty \text{ for all } t \ge 1\Big\}$ and for any set $A \subset \bigotimes_{t=1}^{\infty}(\{1,\ldots,n\}\times\mathbb{N})$, $\Gamma_n(\cdot;A)$ is a Borel measurable function from $\mathbb{R}^n$ into [0,1]. Then, the generalized block bootstrap (GBB) resamples blocks from the collection $\{\mathcal{B}(i,j) : i \ge 1, j \ge 1\}$ according to the transition probability function Γ_n as follows. Let $(I_1,J_1),(I_2,J_2),\ldots$ be a sequence of random vectors with conditional joint distribution $\Gamma_n(\mathcal{X}_n;\cdot)$, given $\mathcal{X}_n$. Then, the blocks selected by the GBB are given by $\mathcal{B}(I_1,J_1),\mathcal{B}(I_2,J_2),\ldots$ (which may *not* be independent). Let $X^*_{G,1},X^*_{G,2},\ldots$ denote the elements of these resampled blocks. Then, the bootstrap version of an estimator $\hat\theta_n = T(F_n)$ under the GBB is defined as $\theta^{*(G)}_{m,n} = T(F^{*(G)}_{m,n})$ for a suitable choice of $m \ge 1$, where $F^{*(G)}_{m,n}$ denotes the empirical distribution of $X^*_{G,1},\ldots,X^*_{G,m}$.

Almost all block bootstrap methods proposed in the literature can be shown to be special cases of the GBB. For example, for the MBB based on a block length ℓ, $1 \le \ell \le n$, the transition probability function Γ_n is given by

$$\Gamma_n(x;\cdot) = \bigotimes_{i=1}^{\infty}\Big((N^{-1}\sum_{j=1}^{N}\delta_j)\times\delta_\ell\Big),\ x \in \mathbb{R}^n$$

where $N = n-\ell+1$ and δ_y is the probability measure putting mass one at y. In this case, $\Gamma_n(x;\cdot)$ does not depend on $x \in \mathbb{R}^n$, and the random indices $(I_1,J_1),(I_2,J_2),\ldots$ are *conditionally iid* random vectors with conditional distribution

$$P_*(I_1=j, J_1=k) = \begin{cases} N^{-1} & \text{if } 1 \le j \le N \text{ and } k=\ell \\ 0 & \text{otherwise .}\end{cases}$$

As a consequence, the resampled blocks $\mathcal{B}(I_1,J_1),\mathcal{B}(I_2,J_2),\ldots$, come from the subcollection $\{\mathcal{B}(i,j) : 1 \le i \le N, j = \ell\}$, which is the same as the collection of overlapping blocks $\{\mathcal{B}_1,\ldots,\mathcal{B}_N\}$ defined in Section 2.5. Similarly, the NBB method can also be shown to be a special case of the GBB. Here, we consider a few other examples.

2.7.1 Circular Block Bootstrap

The Circular Block Bootstrap (CBB) method, proposed by Politis and Romano (1992b) resamples overlapping and periodically extended blocks of a given length ℓ (say) satisfying $1 \ll \ell \ll n$ from the subcollection $\{\mathcal{B}(i,\ell),\ldots,\mathcal{B}(n,\ell)\}$. The transition function Γ_n for the CBB is given by

$$\Gamma_n(x;\cdot) = \bigotimes_{i=1}^{\infty}\Big((n^{-1}\sum_{j=1}^{n}\delta_j)\times\delta_\ell\Big),\ x \in \mathbb{R}^n. \tag{2.17}$$

Denote the resampling block indices for the CBB (i.e., the variables I_i's in the collection $(I_1,J_1),(I_2,J_2),\ldots$ whose joint distribution is specified by the $\Gamma_n(\cdot,\cdot)$ of (2.17)) by $I_{3,1},I_{3,2},\ldots$. Then, (2.17) implies that the variables $I_{3,1},I_{3,2},\ldots$ are *conditionally iid* with $P_*(I_{3,1}=i) = n^{-1}$ and $P_*(J_i=\ell)=1$ for all $i=1,\ldots,n$. Since each X_i appears exactly ℓ times in the collection of blocks $\{\mathcal{B}(i,\ell),\ldots,\mathcal{B}(n,\ell)\}$, and since the CBB resamples the blocks from this collection with equal probability, each of the original observations $X_1,\ldots,X_n$ receives equal weight under the CBB. This property distinguishes the CBB from its predecessors, viz., the MBB and the

NBB, which suffer from edge effects. This is also evident from the following observation. Let $X^*_{3,1}, X^*_{3,2}, \ldots$ denote the CBB observations obtained by arranging the elements of the resampled blocks $\{\mathcal{B}(I_{3,i}, \ell) : i \geq 1\}$ and let $\bar{X}^{*(3)}_m$ denote the CBB sample mean based on m bootstrap observations, where $m = k\ell$ for some integer $k \geq 1$. Then, by (2.17),

$$\begin{aligned} E_*\bar{X}^{*(3)}_m &= E_*\Big[m^{-1}\sum_{i=1}^{m} X^*_{3,i}\Big] \\ &= \ell^{-1}E_*\Big[\sum_{i=1}^{\ell} X^*_{3,i}\Big] \\ &= \ell^{-1}\Big[n^{-1}\sum_{j=1}^{n}\Big\{\sum_{i=1}^{\ell} Y_{n,(j+i-1)}\Big\}\Big] \\ &= \ell^{-1}\Big[\ell\bar{X}_n\Big] \\ &= \bar{X}_n. \end{aligned} \tag{2.18}$$

Thus, the conditional expectation of the bootstrap sample mean under the CBB equals the sample mean of the data $\mathcal{X}_n$, a property not shared by the MBB or the NBB. As noted by Politis and Romano (1992b), this makes it easier to define the bootstrap version of a pivotal quantity of the form $T_n = t_n(\bar{X}_n; \mu)$, where $\mu = EX_1$. Under the CBB, $T^*_{m,n} = t_m(\bar{X}^{*(3)}_m; \bar{X}_n)$ gives the appropriate bootstrap version of T_n. However, replacing the population parameter μ simply by $\bar{X}_n$ to define the bootstrap version of T_n under the MBB or the NBB introduces some extra bias and hence, it is no longer the right thing to do (cf. Lahiri (1992a)). We shall look at properties of the CBB method in Chapters 3, 4, and 5.

2.7.2 *Stationary Block Bootstrap*

The stationary bootstrap (SB) of Politis and Romano (1994b) differ from the earlier block bootstrap methods (i.e., from MBB, NBB, and CBB) in that it uses blocks of *random* lengths rather than blocks of a fixed length ℓ. Let $p \equiv p_n \in (0,1)$ be such that $p \to 0$ and $np \to \infty$ as $n \to \infty$. Then the SB resamples the blocks $\mathcal{B}(I_{4,1}, J_{4,1}), \mathcal{B}(I_{4,2}, J_{4,2}), \ldots$ where the index vectors $(I_{4,1}, J_{4,1}), (I_{4,2}, J_{4,2}), \ldots$ are *conditionally iid* with $I_{4,1}$ having the discrete uniform distribution on $\{1, \ldots, n\}$, and $J_{4,1}$ having the geometric distribution ν_n with parameter p, i.e.,

$$P_*(J_{4,1} = j) \equiv \nu_n(j) = p(1-p)^{j-1}, \quad j = 1, 2, \ldots . \tag{2.19}$$

Furthermore, $I_{4,1}$ and $J_{4,1}$ are independent. Thus, the SB corresponds to the GBB method with the transition function $\Gamma_n(\cdot;\cdot)$ given by

$$\Gamma_n(x;\cdot) = \bigotimes_{i=1}^{\infty}\Big((n^{-1}\sum_{j=1}^{n}\delta_j) \times \nu_n\Big), \quad x \in \mathbb{R}^n.$$

Note that here also, $\Gamma_n(x;\cdot)$ does not depend on $x \in \mathbb{R}^n$.

The SB method can be described through an alternative formulation, also given by Politis and Romano (1994b). Suppose, $X^*_{4,1}, X^*_{4,2}, \ldots$ denote the SB observations, obtained by arranging the elements of the resampled blocks $\mathcal{B}(I_{4,1}, J_{4,1}), \mathcal{B}(I_{4,2}, J_{4,2}), \ldots$ in a sequence. The sequence $\{X^*_{4,i}\}_{i\in\mathbb{N}}$ may also be generated by the following resampling mechanism. Let $X^*_{4,1}$ be picked at random from $\{X_1, \ldots, X_n\}$, i.e., let $X^*_{4,1} = Y_{n,I_{4,1}}$ where $I_{4,1}$ is as above. To select the next observation $X^*_{4,2}$, we further randomize and perform a binary experiment with probability of "Success" equal to p. If the binary experiment results in a "Success," then we select $X^*_{4,2}$ again *at random* from $\{X_1, \ldots, X_n\}$. Otherwise, we set $X^*_{4,2} = Y_{n,(I_{4,1}+1)}$, the observation next to $X^*_{4,1} \equiv Y_{n,I_{4,1}}$ in the periodically extended series $\{Y_{n,i}\}_{i\geq 1}$. In general, given that $X^*_{4,i}$ has been chosen and is given by Y_{n,i_0} for some $i_0 \geq 1$, the next SB observation $X^*_{4,(i+1)}$ is chosen as $Y_{n,(i_0+1)}$ with probability $(1-p)$ and is drawn at random from the original data set $\mathcal{X}_n$ with probability p.

To see that these two formulations are equivalent, let W_i denote the variable associated with the binary experiment for selecting $X^*_{4,i}$, $i \geq 2$. Then, conditional on $\mathcal{X}_n$, $W_i, i \geq 2$ are iid random variables with $P_*(W_i = 1) = p = 1 - P_*(W_i = 0)$, and $\{W_i : i \geq 2\}$ is independent of $\{I_{4,i} : i \geq 1\}$. Next define the variables $M_j, j \geq 0$, by

$$\begin{aligned} M_0 &\equiv 1, \\ M_j &= \inf\{i \geq M_{j-1} + 1 : W_i = 1\}, \ j \geq 1. \end{aligned}$$

Thus, M_j, $j \in \mathbb{N}$ denotes the *trial number* in the sequence of trials $\{W_i : i \geq 2\}$ at which the binary experiment resulted in the jth "Success" and has a negative binomial distribution with parameters j and p (up to a translation). Note that the corresponding SB observation, viz., X^*_{4,M_j}, is then selected at random from $\{X_1, \ldots, X_n\}$ as $X^*_{4,M_j} = Y_{n,I_{4,(j+1)}}$, $j \geq 1$. On the other end, for any i between $M_{j-1} + 1$ and $M_j - 1$, the binary experiment resulted in a block of "Failures," and the corresponding SB observations are selected by picking $(M_j - M_{j-1} - 1)$ variables following $Y_{n,I_{4,j}}$ in the sequence $\{Y_{n,i}\}_{i\in\mathbb{N}}$. Thus, the binary trials $\{W_i : i = M_{j-1}, \ldots, M_j - 1\}$ lead to the "SB block" of observations $\{X^*_{4,M_{j-1}}, \ldots, X^*_{4,(M_j-1)}\} = \{Y_{n,I_{4,j}}, \ldots, Y_{n,(I_{4,j}+M_j-M_{j-1}-1)}\}, j \geq 1$. Now, defining $J_{4,j} = M_j - M_{j-1}$, $j \geq 1$ and using the properties of the negative binomial distribution (cf. Section XI.2, Feller (1971a)), we may conclude that $J_{4,1}, J_{4,2}, \ldots$ are (conditionally) iid and follow the geometric distribution with parameter p. Hence, the two formulations of the SB are equivalent.

An important property of the SB method is that conditional on $\mathcal{X}_n$, the bootstrap observations $\{X^*_{4,i}\}_{i\in\mathbb{N}}$ are *stationary* (which is why it is called the "stationary" bootstrap). A simple proof of this fact can be derived using the second formulation of the SB as follows. Let $\{Z_i\}_{i\in\mathbb{N}}$ be a Markov chain on $\{1,\ldots,n\}$ such that conditional on $\mathcal{X}_n$, the initial distribution of the chain is $\boldsymbol{\pi} \equiv (n^{-1},\ldots,n^{-1})'$ and the stationary transition probability matrix of $\{Z_i\}_{i\in\mathbb{N}}$ is $Q \equiv ((q_{ij}))$, where

$$q_{ij} = \begin{cases} p+n^{-1}(1-p) & 1\le i<n, j=i+1 \\ n^{-1}(1-p) & 1\le i<n, j\ne i+1 \\ n^{-1}(1-p) & i=n, 2\le j\le n \\ p+n^{-1}(1-p) & i=n, j=1\ . \end{cases} \tag{2.20}$$

Thus, Z_1 takes the values $1,\ldots,n$ with probability n^{-1} each. Also, for any $k\ge 1$, given that $Z_k = i$, $1\le i\le n$, the next index Z_{k+1} takes the value $i+1$ (modulo n) with probability $p+n^{-1}(1-p)$ and it takes each of the remaining $(n-1)$ values with probability $n^{-1}(1-p)$. Thus, from the second formulation of the SB described earlier, it follows that the SB observations $\{X^*_{4,i}\}_{i\in\mathbb{N}}$ may also be generated by the index variables $\{Z_i\}_{i\in\mathbb{N}}$ as

$$X^*_{4,i} = X_{Z_i},\ i\ge 1\ . \tag{2.21}$$

To see that $\{X^*_{4,i}\}_{i\in\mathbb{N}}$ is stationary, note that by definition, the transition matrix Q is doubly stochastic and that it satisfies the relation $\boldsymbol{\pi}'Q = \boldsymbol{\pi}'$. Therefore, $\boldsymbol{\pi}$ is the *stationary* distribution of $\{Z_i\}_{i\in\mathbb{N}}$ and $\{Z_i\}_{i\in\mathbb{N}}$ is a *stationary* Markov chain. Thus, we have proved the following Theorem.

Theorem 2.3 *Let $\mathcal{F}_{in}$ denote the σ-field generated by Z_i and $\mathcal{X}_n$, $i\ge 1$. Then, conditional on $\mathcal{X}_n$, $\{X^*_{4,i},\mathcal{F}_{in}\}_{i\in\mathbb{N}}$ is a stationary Markov chain for each $n\ge 1$, i.e.,*

$$\mathcal{L}(X^*_{4,i}|\mathcal{X}_n) = \mathcal{L}(X^*_{4,1}|\mathcal{X}_n) \quad \textit{for all} \quad i\ge 1$$

and

$$\mathcal{L}(X^*_{4,(i+1)}|Z_1,\ldots,Z_i,\mathcal{X}_n) = \mathcal{L}(X^*_{4,(i+1)}|Z_i,\mathcal{X}_n) \quad \textit{for all} \quad i\ge 1\ .$$

In particular, Theorem 2.3 implies that conditional on $\mathcal{X}_n$, $\{X^*_{4,i}\}_{i\ge 1}$ is stationary. Furthermore, by (2.20) and (2.21), for a given resample size m, the conditional expectation of the SB sample mean $\bar{X}^{*(4)}_m \equiv m^{-1}\sum_{i=1}^m X^*_{4,i}$ is given by

$$E_*(\bar{X}^{*(4)}_m) = E_* X^*_{4,1} = \bar{X}_n\ . \tag{2.22}$$

We shall consider other properties of the SB method in Chapters 3–5.

2.8 Subsampling

Use of different subsets of the data to approximate the bias and variance of an estimator is a common practice, particularly in the context of iid observations. For example, the Jackknife bias and variance estimators are computed using subsets of size $n-1$ from the full sample $\mathcal{X}_n = (X_1,\ldots,X_n)$ (cf. Efron (1982)). However, as noted recently (see Carlstein (1986), Politis and Romano (1994a), Hall and Jing (1996), Bickel et al. (1997), and the references therein), subseries of dependent observations can also be used to produce valid estimators of the bias, the variances, and more generally, of the sampling distribution of a statistic under very weak assumptions.

To describe the subsampling method, suppose that $\hat{\theta}_n = t_n(\mathcal{X}_n)$ is an estimator of a parameter θ, such that for some normalizing constant $a_n > 0$, the probability distribution $Q_n(x) = P(a_n(\hat{\theta}_n - \theta)\le x)$ of the centered and scaled estimator $\hat{\theta}_n$ converges weakly to a limit distribution $Q(x)$, i.e.,

$$Q_n(x) \to Q(x) \quad \text{as} \quad n\to\infty \tag{2.23}$$

for all continuity points x of Q. Furthermore, assume that $a_n \to \infty$ as $n\to\infty$ and that Q is *not degenerate* at zero, i.e., $Q(\{0\}) < 1$. Let $1\le \ell\le n$ be a given integer and let

$$\mathcal{B}_i = (X_i,\ldots,X_{i+\ell-1})'\ ,$$

$1\le i\le N$, denote the overlapping blocks of length ℓ where $N = n-\ell+1$. Note that the blocks $\mathcal{B}_i$'s are the same as those defined in Section 2.4 for the MBB. Then, the subsampling estimator of Q_n, based on the *overlapping* version of the subsampling method, is given by

$$\hat{Q}_n(x) = N^{-1}\sum_{i=1}^N \mathbb{1}(a_\ell(\hat{\theta}_{i,\ell} - \hat{\theta}_n)\le x),\ x\in\mathbb{R}\ , \tag{2.24}$$

where $\hat{\theta}_{i,\ell}$ is a "copy" of the estimator $\hat{\theta}_n$ on the block $\mathcal{B}_i$, defined by $\hat{\theta}_{i,\ell} = t_\ell(\mathcal{B}_i)$, $i = 1,\ldots,N$. Note that we used $t_\ell(\cdot)$ (in place of $t_n(\cdot)$) to define the subsample copy "$\hat{\theta}_{i,\ell}$," as the ith block $\mathcal{B}_i$ contains only ℓ observations. That is also the reason behind using the scaling constant a_ℓ instead of a_n. From the above description, it follows that the overlapping version of the subsampling method is a special case of the MBB where a *single* block is resampled.

The estimator $\hat{Q}_n$ of the distribution function $Q_n(x)$ can be used to obtain subsampling estimators of the bias and the variance of $\hat{\theta}_n$. Note that the bias of $\hat{\theta}_n$ is given by

$$\text{Bias}(\hat{\theta}_n) = E\hat{\theta}_n - \theta = a_n^{-1}\int x\,dQ_n(x)\ .$$

The subsampling estimator of $\text{Bias}(\hat{\theta}_n)$ is then obtained by replacing $Q_n(\cdot)$ by $\hat{Q}_n(\cdot)$, viz.,

$$\widehat{\text{Bias}}(\hat{\theta}_n) = a_n^{-1}\int x d\hat{Q}_n(x) = a_\ell a_n^{-1}\Big(N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell} - \hat{\theta}_n\Big). \tag{2.25}$$

Similarly, the subsampling estimator of the variance of $\hat{\theta}_n$ is given by

$$\widehat{\text{Var}}(\hat{\theta}_n) = a_\ell^2 a_n^{-2}\Bigg[N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell}^2 - \Big(N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell}\Big)^2\Bigg], \tag{2.26}$$

which is the sample variance of $\hat{\theta}_{i,\ell}$'s multiplied by the scaling factor $a_\ell^2 a_n^{-2}$. In (2.25) and (2.26), we need to use the correction factors (a_ℓ/a_n) and $(a_\ell/a_n)^2$ to *scale up* from the level of $\hat{\theta}_{i,\ell}$'s, which are defined using ℓ-observations, to the level of $\hat{\theta}_n$, which is defined using n-observations. In applying a bootstrap method, one typically uses a resample size that is comparable to the original sample size, and therefore, such explicit corrections of the bootstrap bias and variance estimators are usually unnecessary.

In analogy to the bootstrap methods, one may attempt to apply the subsampling method to a centered variable of the form $T_{1n} \equiv (\hat{\theta}_n - \theta)$. However, this may *not* be the right thing to do. Indeed, if instead of using the subsampling method for the scaled random variable $a_n(\hat{\theta}_n - \theta)$, we consider only the centered variable $T_{1n} = (\hat{\theta}_n - \theta)$, then the subsampling estimator of the distribution Q_{1n}, say, of T_{1n} would be given by

$$\hat{Q}_{1n}(x) \equiv N^{-1}\sum_{i=1}^{N}\mathbb{1}((\hat{\theta}_{i,\ell} - \hat{\theta}_n) \le x), \ x \in \mathbb{R} .$$

Since $\text{Bias}(\hat{\theta}_n) = ET_{1n} = \int x d\hat{Q}_{1n}(x)$, using $\hat{Q}_{1n}(x)$, we would get

$$\widehat{\text{Bias}}_{1n}(\hat{\theta}_n) = \int x d\hat{Q}_{1n}(x) = \Big(N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell} - \hat{\theta}_n\Big),$$

as an estimator of $\text{Bias}(\hat{\theta}_n)$, and similarly, we would get

$$\widehat{\text{Var}}_{1n}(\hat{\theta}_n) = N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell}^2 - \Big(N^{-1}\sum_{i=1}^{N}\hat{\theta}_{i,\ell}\Big)^2$$

as an estimator of $\text{Var}(\hat{\theta}_n)$. However, these subsampling estimators of the bias and the variance of $\hat{\theta}_n$, defined using $\hat{Q}_{1n}(x)$, are very "poor" estimators of the corresponding population parameters. To appreciate why, consider the case where $\hat{\theta}_n = \bar{X}_n$ and $\theta = EX_1$ and $n^{1/2}(\hat{\theta}_n - \theta) \longrightarrow^d N(0, \sigma_\infty^2)$ as $n \to \infty$ with $\sigma_\infty^2 = \sum_{i=-\infty}^{\infty} \text{Cov}\,(X_1, X_{i+1}) \neq 0$. Write

$\bar{X}_{i,\ell}$ for the average of the ℓ observations in $\mathcal{B}_i$, $i = 1, \ldots, N$. Then, $\widehat{\text{Var}}_{1n}(\hat{\theta}_n) = N^{-1}\sum_{i=1}^{N}\bar{X}_{i,\ell}^2 - \hat{\mu}_n^2$, where $\hat{\mu}_n \equiv N^{-1}\sum_{i=1}^{N}\bar{X}_{i,\ell}$ is the average of the N block averages. Then, it is not difficult to show that under some standard moment and weak dependence conditions on the process $\{X_i\}_{i\in\mathbb{Z}}$ and under the assumption that $\ell^{-1} + n^{-1}\ell = o(1)$ as $n \to \infty$,

$$\begin{aligned} &\widehat{\text{Var}}_{1n}(\hat{\theta}_n) \\ = \ & \text{Var}(\bar{X}_\ell) + N^{-1}\sum_{i=1}^{N}\Big\{[\bar{X}_{i,\ell} - \theta]^2 - \text{Var}(\bar{X}_\ell)\Big\} - [\hat{\mu}_n - \theta]^2 \\ = \ & \ell^{-1}\sigma_\infty^2 + O(\ell^{-2}) + O_p([n\ell]^{-1/2}) + O_p(n^{-1}) , \end{aligned} \tag{2.27}$$

whereas $\text{Var}(\bar{X}_n) = n^{-1}\sigma_\infty^2 + O(n^{-2})$ as $n \to \infty$. Thus, $\widehat{\text{Var}}_{1n}(\hat{\theta}_n)$ indeed overestimates the variance of $\hat{\theta}_n$ by a scaling factor of n/ℓ, which blows up to infinity with n. It is easy to see that the other estimator, viz., $\widehat{\text{Var}}(\hat{\theta}_n)$ is equal to ℓ/n times $\widehat{\text{Var}}_{1n}(\hat{\theta}_n)$ in this case and thus, provides a sensible estimator of $\text{Var}(\bar{X}_n)$. The reason that the subsampling estimator based on T_{1n} does *not* work in this case is that the limit distribution of T_{1n} is *degenerate at zero*, and does not satisfy the nondegeneracy requirement stated above.

Formulas (2.24), (2.25), and (2.26) illustrate a very desirable property of the subsampling method that holds true generally. Computations of $\hat{Q}_n(\cdot)$ and of estimates of other population quantities based on $\hat{Q}_n$ do not involve any resampling and hence, are less demanding. Typically, a simple, closed-form expression can be written down for a subsampling estimator of a level-2 parameter, and it needs computation of the subsampling version $\hat{\theta}_{i,\ell}$ of the estimator $\hat{\theta}_n$ only N times, as compared to a much larger number of times for the resampling methods like the MBB. However, the price paid is the lack of "automatic" second-order correctness of the subsampling method compared to the MBB and other block bootstrap methods.

We conclude this section with an observation. As noted previously, the subsampling method is a special case of the MBB where the number of resampled blocks is identically equal to 1. Exploiting this fact, we may similarly define other versions of the subsampling method based on nonoverlapping blocks or circular blocks. More generally, it is possible to extend the subsampling method in the spirit of the GBB method. We define the "generalized subsampling" method as the GBB method with a single sample (I_1, J_1) of the indices. Thus, the generalized subsampling estimator of $Q_n(x)$ (cf. (2.23)) is given by

$$\hat{Q}_n^{GS}(x) = E_*\mathbb{1}\Big(a_{J_1}[\hat{\theta}^*_{J_1,n} - \hat{\theta}_n] \le x\Big), \ x \in \mathbb{R} ,$$

where $\hat{\theta}^*_{J_1,n} = t_{J_1}(\mathcal{B}(I_1, J_1))$ is a copy of $\hat{\theta}_n$ based on the GBB samples from a *single* block $\mathcal{B}(I_1, J_1)$ of length J_1.

2.9 Transformation-Based Bootstrap

As described in Chapter 1, the basic idea behind the bootstrap method is to recreate the relation between the population and the sample using the sample itself. For dependent data, the most common approach to this problem is to resample "blocks" of observations instead of single observations, which preserves the dependence structure of the underlying process *within* the resampled blocks and is able to reproduce the effect of dependence at short lags. A quite different approach to the problem was suggested by Hurvich and Zeger (1987). In their seminal work, Hurvich and Zeger (1987) considered the discrete Fourier transform (DFT) of the data and rather than resampling the data values directly, they applied the IID bootstrap method of Efron (1979) to the DFT values. The *transformation based bootstrap* (TBB) described here is a generalization of Hurvich and Zeger's (1987) idea.

To describe it, let $\theta \equiv \theta(P)$ be a parameter of interest, which depends on the underlying probability measure P generating the sequence $\{X_i\}_{i\in\mathbb{Z}}$, and let $T_n \equiv t_n(\mathcal{X}_n)$ be an estimator of θ based on the observations $\mathcal{X}_n = (X_1, \ldots, X_n)$. Our goal is to approximate the sampling distribution of a normalized or studentized statistic $R_n = r_n(\mathcal{X}_n; \theta)$. Let $\mathcal{Y}_n = h_n(\mathcal{X}_n)$ be a (one-to one) transformation of $\mathcal{X}_n$ such that the components of $\mathcal{Y}_n$, say, $\{Y_i : i \in \mathcal{I}_n\}$, are "approximately independent". Also suppose that the variable R_n can be expressed (at least to a close approximation) in terms of $\mathcal{Y}_n$ as $R_n = r_{1n}(\mathcal{Y}_n; \theta)$ for some reasonable function r_{1n}. Then, to approximate the distribution of R_n by the TBB, we resample from a *suitable* subcollection $\{Y_i : i \in \mathcal{J}_n\}$ of $\{Y_i : i \in \mathcal{I}_n\}$ to get the bootstrap observations $\mathcal{Y}_n^* \equiv \{Y_i^* : i \in \mathcal{I}_n\}$ either by selecting a single Y-value at a time as in the IID-bootstrap method of Efron (1979) or by selecting a block of Y-values from $\{Y_i : i \in \mathcal{J}_n\}$ as in the MBB, depending on the dependence structure of $\{Y_i : i \in \mathcal{J}_n\}$. The TBB estimator of the distribution of R_n is then given by the conditional distribution of $R_n^* \equiv r_{1n}(\mathcal{Y}_n^*; \hat{\theta}_n)$ given the data $\mathcal{X}_n$, where $\hat{\theta}_n$ is an estimator of θ based on $\mathcal{X}_n$. Thus, as a principle, the TBB method suggests an *additional* transformation step to reduce the dependence in the data to an iid structure or to a weaker form of dependence.

An important example of the TBB method is the *Frequency Domain Bootstrap* (FDB), which uses the Fourier transform of the data to generate the Y-variables of the TBB. Suppose that $\{X_i\}_{i\in\mathbb{Z}}$ is a sequence of stationary, weakly dependent random variables. The Fourier transform of the observations $\mathcal{X}_n$ is defined as

$$Y_n(w) = n^{-1/2} \sum_{j=1}^{n} X_j \exp(-\iota w j), \ \ w \in (-\pi, \pi] \ ,$$

where recall that $\iota = \sqrt{-1}$. Though the X_i's are dependent, a well known result in time-series states (cf. Brockwell and Davis (1991, Chapter 10); Lahiri (2003a)) that for any set of distinct ordinates $-\pi < \lambda_1, \ldots, \lambda_k \leq \pi$, the Fourier transforms $Y_n(\lambda_1), \ldots, Y_n(\lambda_k)$ are *asymptotically independent*. Furthermore, the original observations $\mathcal{X}_n$ admit a representation in terms of the transformed values $\mathcal{Y}_n = \{Y_n(w_j) : j \in \mathcal{I}_n\}$ as (cf. Brockwell and Davis (1991, Chapter 10)),

$$X_t = n^{-1/2} \sum_{j\in\mathcal{I}_n} Y_n(w_j) \exp(\iota t w_j), \ \ t = 1, \ldots, n \tag{2.28}$$

where $w_j = 2\pi j/n$ and $\mathcal{I}_n = \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$. Thus, using the *inversion formula* (2.28), we can express a given variable $R_n = r_n(\mathcal{X}_n; \theta)$ also in terms of the transformed values $\mathcal{Y}_n$. Since the variables in $\mathcal{Y}_n$ are approximately independent, we may (suitably) resample these Y-values to define the FDB version of R_n. Here, however, some care must be taken since the (asymptotic) variance of the Y-variables are not necessarily identical. A more complete description of the FDB method and its properties are given in Chapter 9.

2.10 Sieve Bootstrap

Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary time series and let $T_n = t_n(X_1, \ldots, X_n)$ be an estimator of a level-1 parameter of interest $\theta = \theta(P)$, where P denotes the (unknown) joint distribution of $\{X_i\}_{i\in\mathbb{Z}}$. Then, the sampling distribution of T_n is given by

$$G_n(B) = P(T_n \in B) = P \circ t_n^{-1}(B) \tag{2.29}$$

for Borel sets B in $\mathbb{R}$, where $P \circ t_n^{-1}$ denotes the probability distribution on $\mathbb{R}$ induced by the transformation $t_n(\cdot)$ under P. As described in Chapter 1, the bootstrap and other resampling methods are general estimation methods for estimating the level-2 parameters like $G_n(B)$, $\text{Var}(T_n)$, etc. When the X_i's are iid with a common distribution F, we may write $P = F^\infty$ and an estimator of $G_n(B)$ in (2.29) may be generated by replacing P with $\hat{P}_n = \hat{F}_n^\infty$ in (2.28), where $\hat{F}_n$ is an estimator of F. However, when the X_i's are dependent, such a factorization of P does not hold. In this case, estimation of the level-2 parameter $G_n(B)$ can be thought of as a two-step procedure where, in the first step, P is approximated by a "simpler" probability distribution $\tilde{P}_n$ and in the next step, $\tilde{P}_n$ is estimated using the data $\{X_1, \ldots, X_n\}$. The idea of the sieve bootstrap is to choose $\{\tilde{P}_n\}_{n\geq 1}$ to be a sieve approximation to P, i.e., $\{\tilde{P}_n\}_{n\geq 1}$ is a sequence of probability measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ such that for each n, $\tilde{P}_{n+1}$ is a finer approximation to P than $\tilde{P}_n$ and $\tilde{P}_n$ converges to P (in some suitable sense) as $n \to \infty$.

For the block bootstrap methods like the NBB or the MBB, the first step approximation $\tilde{P}_n$ is taken to be $P_\ell \otimes P_\ell \otimes \ldots$, where P_ℓ denotes the joint distribution of the block $\{X_1, \ldots, X_\ell\}$ of length ℓ. In the second step, P_ℓ is estimated by the empirical distribution of all overlapping (under MBB) or nonoverlapping (under NBB) blocks of length ℓ contained in the data. For a large class of stationary processes, Bühlmann (1997) presents a sieve bootstrap method based on a sieve of autoregressive processes of increasing order, which we shall briefly describe here. However, other choices of $\{\tilde{P}_n\}_{n\geq 1}$ is possible. See Bühlmann (2002) for another interesting proposal based on variable length Markov chains for finite state space categorical time series. In general, there is a trade-off between the accuracy and the range of validity of a given sieve bootstrap method. Typically, one may choose a sieve to obtain a more accurate bootstrap estimator, but only at the expense of restricting the applicability to a smaller class of processes (cf. Lahiri (2002b)).

Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary process with $EX_1 = \mu$ such that it admits the one-sided moving average representation

$$X_i - \mu = \sum_{j=0}^{\infty} \alpha_j \epsilon_{i-j}\ , \quad i \in \mathbb{Z} \tag{2.30}$$

where $\{\epsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of zero mean uncorrelated random variables and where $\alpha_0 = 1$, $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$. Suppose that $\{X_i\}_{i\in\mathbb{Z}}$ satisfies the standard invertibility conditions for a linear process (cf. Theorem 7.6.9, Anderson (1971)). Then, we can represent $\{X_i - \mu\}_{i\in\mathbb{Z}}$ as a one-sided infinite order autoregressive process

$$(X_i - \mu) = \sum_{j=1}^{\infty} \beta_j (X_{i-j} - \mu) + \epsilon_i\ , \quad i \in \mathbb{Z} \tag{2.31}$$

with $\sum_{j=1}^{\infty} \beta_j^2 < \infty$. The representation (2.31) suggests that autoregressive processes of finite orders p_n, $n \geq 1$, may be used to define a sieve approximation for the joint distribution P of $\{X_i\}_{i\in\mathbb{Z}}$. To describe the sieve bootstrap based on autoregression, let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ denote the observations from the process $\{X_i\}_{i\in\mathbb{Z}}$. Let $\{p_n\}_{n\geq 1}$ be a sequence of positive integers such that $p_n \uparrow \infty$ as $n \to \infty$, but $n^{-1}p_n \to 0$ as $n \to \infty$. The sieve approximation $\tilde{P}_n$ to P is determined by the autoregressive process

$$X_i - \mu = \sum_{j=1}^{p_n} \beta_j (X_{i-j} - \mu) + \epsilon_i\ , \quad i \in \mathbb{Z}\ . \tag{2.32}$$

Next, we fit the AR(p_n) model (2.32) to the data $\mathcal{X}_n$ to obtain estimators of the autoregression parameters $\hat{\beta}_{1n}, \ldots, \hat{\beta}_{p_n n}$ (for example, by the least squares method). This yields the residuals

$$\hat{\epsilon}_{in} = (X_i - \bar{X}) - \sum_{j=1}^{p_n} \hat{\beta}_{jn} (X_{i-j} - \bar{X}_n)\ , \quad p_n + 1 \leq i \leq n$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. As in Section 2.4, we center the residuals at $\bar{\epsilon}_n = (n - p_n)^{-1}\sum_{i=p_n+1}^{n} \hat{\epsilon}_{in}$ and resample from the centered residuals $\{\hat{\epsilon}_{in} - \bar{\epsilon}_n : p_n + 1 \leq i \leq n\}$ to generate the sieve bootstrap error variables ϵ_i^*, $i \geq p_n + 1$. Then, the sieve bootstrap observations are generated by the recursion relation

$$(X_i^* - \bar{X}_n) = \sum_{j=1}^{p_n} \hat{\beta}_{jn} (X_{i-j}^* - \bar{X}_n) + \epsilon_i^*\ , \quad i \geq p_n + 1$$

by setting the initial p_n-variables $X_1^*, \ldots, X_{p_n}^*$ equal to $\bar{X}_n$. The autoregressive sieve bootstrap version of the estimator $T_n = t_n(X_1, \ldots, X_n)$ is now given by

$$T_{m,n}^* = t_m(X_1^*, \ldots, X_m^*)\ , \quad m > p_n\ .$$

Under some regularity conditions on the variables $\{\epsilon_i\}_{i\in\mathbb{Z}}$ of (2.30) and the sieve parameter p_n, Bühlmann (1997) establishes consistency of the autoregressive sieve bootstrap. It follows from his results that the autoregressive sieve bootstrap provides a more accurate variance estimator for the class of estimators given by (2.11) than the MBB and the NBB. However, consistency of the autoregressive sieve bootstrap variance estimators holds for a more restricted class of processes than the block bootstrap methods. See Bühlmann (1997), Choi and Hall (2000), and the references therein for more about the properties of the autoregressive sieve bootstrap.

7

Empirical Choice of the Block Size

7.1 Introduction

As we have seen in the earlier chapters, performance of block bootstrap methods critically depends on the block size. In this chapter, we describe the theoretical optimal block lengths for the estimation of various level-2 parameters and discuss the problem of choosing the optimal block sizes empirically. For definiteness, we restrict attention to the MBB method. Analogs of the block size estimation methods presented here can be defined for other block bootstrap methods. In Section 7.2, we describe the forms of the MSE-optimal block lengths for estimating the variance and the distribution function. In Section 7.3, we present a data-based method for choosing the optimal block length based on the subsampling method. This is based on the work of Hall, Horowitz and Jing (1995). A second method based on the Jackknife-After-Bootstrap (JAB) method is presented in Section 7.4. Numerical results on finite sample performance of these optimal block length selection rules are also given in the respective sections.

7.2 Theoretical Optimal Block Lengths

Let $(X_1, \ldots, X_n) = \mathcal{X}_n$ denote a finite stretch of random variables, observed from a stationary weakly dependent process $\{X_i\}_{i \in \mathbb{Z}}$ in $\mathbb{R}^d$. Let $\hat{\theta}_n$ be an estimator of a level-1 parameter of interest $\theta \in \mathbb{R}$, based on $\mathcal{X}_n$. In this section, we obtain expansions for the MSEs of block bootstrap estimators for various characteristics of the distribution of $\hat{\theta}_n$. Let G_n denote the distribution of the centered estimator $(\hat{\theta}_n - \theta)$, i.e.,

$$G_n(x) = P(\hat{\theta}_n - \theta \le x), \; x \in \mathbb{R} . \tag{7.1}$$

The level-2 parameters of interest here are given by

$$\varphi_{1n} = \text{Bias}(\hat{\theta}_n) = \int x dG_n(x) \tag{7.2}$$

$$\varphi_{2n} = \text{Var}(\hat{\theta}_n) = \int x^2 dG_n(x) - \Big(\int x dG_n(x)\Big)^2 \tag{7.3}$$

$$\varphi_{3n} \equiv \varphi_{3n}(x_0) = P\Big(\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\tau_n} \le x_0\Big) = G_n\Big(\frac{x_0\tau_n}{\sqrt{n}}\Big) \tag{7.4}$$

$$\varphi_{4n} = \varphi_{4n}(y_0) \equiv P\Big(\Big|\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\tau_n}\Big| \le y_0\Big) = G_n\Big(\frac{y_0\tau_n}{\sqrt{n}}\Big) - G_n\Big(\frac{-y_0\tau_n}{\sqrt{n}}\Big) , \tag{7.5}$$

where $x_0 \in \mathbb{R}$ and $y_0 \in (0, \infty)$ are given real numbers and where τ_n^2 is the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta)$. Here, φ_{1n} and φ_{2n} are, respectively, the bias and the variance of the estimator $\hat{\theta}_n$, φ_{3n} denotes the (one-sided) distribution function of $\sqrt{n}(\hat{\theta}_n - \theta)$ at a given point $x_0 \in \mathbb{R}$, and φ_{4n} denotes the two-sided distribution function of $\sqrt{n}(\hat{\theta}_n - \theta)$ at $y_0 \in (0, \infty)$. The latter is useful for constructing symmetric confidence intervals for θ (cf. Hall (1992)). Next, for $k = 1, 2, 3, 4$, let $\hat{\varphi}_{kn}(\ell)$ denote the MBB estimators of the level-2 parameter φ_{kn} based on blocks of length ℓ. We define the theoretical optimal block length ℓ_{kn}^0 as the minimizer of the MSE of $\hat{\varphi}_{kn}(\ell)$ over a set of values of the block size ℓ, depending on $k = 1, 2, 3, 4$. Specifically, we define

$$\ell_{kn}^0 = \text{argmin}\Big\{\text{MSE}(\hat{\varphi}_{kn}(\ell)) : \epsilon n^\epsilon < \ell < \epsilon^{-1} n^{1/2-\epsilon}\Big\}, \; k = 1, 2 \tag{7.6}$$

$$\ell_{kn}^0 = \text{argmin}\Big\{\text{MSE}(\hat{\varphi}_{kn}(\ell)) : \epsilon n^\epsilon \le \ell \le \epsilon^{-1} n^{1/3-\epsilon}\Big\}, \; k = 3, 4 \tag{7.7}$$

for some small $\epsilon > 0$. It will follow from the arguments and results below that the theoretical optimal block length ℓ_{kn}^0 is of the order $n^{1/3}$ for the bias and the variance functionals (with $k = 1, 2$), while the order of ℓ_{kn}^0 for the one- and the two-sided distribution functions, with $k = 3$ and $k = 4$, are of the orders $n^{1/4}$ and $n^{1/5}$, respectively. Thus, the ranges $[\epsilon n^\epsilon, \epsilon^{-1} n^{1/2-\epsilon}]$ and $[\epsilon n^\epsilon, \epsilon^{-1} n^{1/3-\epsilon}]$ of block lengths ℓ in (7.6) and (7.7), respectively, contain the optimal block lengths ℓ_{kn}^0 for all $k = 1, 2, 3, 4$. Indeed, it can be shown that under some additional regularity conditions, the theoretical optimal block lengths ℓ_{kn}^0 have the same order even when the ranges of ℓ values in (7.6) and (7.7) are replaced by the larger interval $[\epsilon n^\epsilon, \epsilon^{-1} n^{1-\epsilon}]$ for an arbitrarily small $\epsilon \in (0, 1)$. However, we will restrict

attention to the range of ℓ values specified by (7.6) and (7.7) and will not pursue such generalizations here.

For deriving expansions for the MSEs of the block bootstrap estimators $\hat{\varphi}_{kn}(\ell)$'s, $k = 1, 2, 3, 4$, we shall suppose that the level-1 parameter θ and its estimator $\hat{\theta}_n$ satisfy the requirements of the Smooth Function Model (cf. Section 4.2). Thus, there exists a function $H : \mathbb{R}^d \to \mathbb{R}$ such that

$$\hat{\theta}_n = H(\bar{X}_n), \; \theta = H(\mu) \tag{7.8}$$

and the function H is "smooth" in a neighborhood of μ, where $\mu = EX_1$ and $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Recall that we write $c_\alpha = D^\alpha H(\mu)/\alpha!$, D^α for the differential operator $\frac{\partial^{\alpha_1 + \cdots + \alpha_d}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ and $\alpha! = \prod_{i=1}^d \alpha_i!$ for $\alpha = (\alpha_1, \ldots, \alpha_d)' \in \mathbb{Z}_+^d$.

7.2.1 *Optimal Block Lengths for Bias and Variance Estimation*

Expansions of the MSEs of the MBB estimators of the bias and the variance of the estimator $\hat{\theta}_n$ under the Smooth Function Model (7.8) was given in Chapter 5. Here, we recast the relevant results in a slightly different form by expressing relevant population quantities in the time domain. Let Z_∞ be a d-dimensional Gaussian random vector with mean zero and covariance matrix $\Sigma_\infty = \sum_{j=-\infty}^{\infty} E\{(X_1 - \mu)(X_{1+j} - \mu)'\}$.

Theorem 7.1 *Suppose that $\ell^{-1} + n^{-1/2}\ell = o(1)$ as $n \to \infty$.*

(a) Suppose that Conditions (5.D_r) and (5.M_r) of Section 5.4 hold with $r = 3$ and $r = 3 + a_0$, respectively, where a_0 is as specified by (5.D_r). Then

$$MSE\Big(n \cdot \hat{\varphi}_{1n}(\ell)\Big) = \Big[(n^{-1}\ell)\frac{2}{3}\text{Var}\Big(\sum_{|\alpha|=2} c_\alpha Z_\infty^\alpha\Big) + \ell^{-2} A_1^2\Big] + o(n^{-1}\ell + \ell^{-2}) , \tag{7.9}$$

where

$$A_1 = -\sum_{|\alpha|=1}\sum_{|\beta|=1} c_{\alpha+\beta}\Big[\sum_{j=-\infty}^{\infty} |j| E(X_1 - \mu)^\alpha (X_{1+j} - \mu)^\beta\Big] .$$

(b) Suppose that Conditions (5.D_r) and (5.M_r) of Section 5.4 hold with $r = 2$ and $r = 4 + 2a_0$, respectively, where a_0 is as specified by Condition (5.D_r). Then,

$$MSE\Big(n \cdot \hat{\varphi}_{2n}(\ell)\Big) = \Big[(n^{-1}\ell)\frac{2}{3}\text{Var}\Big(\Big(\sum_{|\alpha|=1} c_\alpha Z_\infty^\alpha\Big)^2\Big) + \ell^{-2} A_2^2\Big] + o(n^{-1}\ell + \ell^{-2}) , \tag{7.10}$$

where

$$A_2 = -\sum_{|\alpha|=1}\sum_{|\beta|=1} c_\alpha c_\beta \Big[\sum_{j=-\infty}^{\infty} |j| E(X_1-\mu)^\alpha (X_{1+j}-\mu)^\beta\Big] .$$

Proof: Follows from the proofs of Theorems 5.1 and 5.2 for the case '$j = 1$' (corresponding to the MBB estimators). □

Note that under the regularity conditions of Theorem 7.1, both the bias and the variance of the estimator $\hat{\theta}_n$ are of the order $O(n^{-1})$. Hence, we state the MSEs of the scaled bootstrap bias estimator $n \cdot \hat{\varphi}_{1n}(\ell)$ and of the scaled bootstrap variance estimator $n \cdot \hat{\varphi}_{2n}(\ell)$, in Theorem 7.1. Alternatively, we may think of the scaled bootstrap estimators $n \cdot \hat{\varphi}_{kn}(\ell)$ as estimators of the limiting level-2 parameters $\varphi_{k,\infty} \equiv \lim_{n\to\infty} n \cdot \varphi_{kn}$, $k = 1, 2$, given by

$$\varphi_{1,\infty} = \sum_{|\alpha|=1}\sum_{|\beta|=1} c_{\alpha+\beta} \Big[\sum_{j=-\infty}^{\infty} E(X_1-\mu)^\alpha (X_{1+j}-\mu)^\beta\Big]$$

and

$$\varphi_{2,\infty} = \sum_{|\alpha|=1}\sum_{|\beta|=1} c_\alpha c_\beta \Big[\sum_{j=-\infty}^{\infty} E(X_1-\mu)^\alpha (X_{1+j}-\mu)^\beta\Big] .$$

Theorem 7.1 immediately yields expressions for the leading terms of the theoretical optimal block lengths for bias and variance estimation. We note these down in the following corollary.

Corollary 7.1 *Suppose that the respective set of conditions of Theorem 7.1 hold for the bias functional ($k = 1$) and the variance functional ($k = 2$), and that the constants A_1 and A_2 are nonzero. Then, for $k = 1, 2$,*

$$\ell_{kn}^0 = n^{1/3}(2A_k^2/v_k^2)^{1/3} + o(n^{1/3}) , \tag{7.11}$$

where $v_1^2 = \frac{2}{3}\mathrm{Var}(\sum_{|\alpha|=2} c_\alpha Z_\infty^\alpha)$ *and* $v_2^2 = \frac{2}{3}\mathrm{Var}([\sum_{|\alpha|=1} c_\alpha Z_\infty^\alpha]^2)$.

Künsch (1989) derived the leading term of the theoretical optional block length for the variance functional while Hall, Horowitz and Jing (1995) derived the leading terms for both the bias and the variance functionals φ_{1n} and φ_{2n}. The conclusions of Corollary 7.1 can be strengthened to some extent. A more detailed analysis of the remainder term in the proof of Theorem 7.1 can be used to show that under some additional smoothness and moment conditions, the $o(n^{1/3})$ term on the right side (7.11) is indeed $O(1)$ as $n \to \infty$, for both $k = 1$ and $k = 2$. Thus, the fluctuations of the true optimal block length from its leading term is bounded for both bias and variance functionals. In the next section, we consider theoretical optimal block lengths for the estimation of distribution functions.

7.2.2 Optimal Block Lengths for Distribution Function Estimation

First we consider the one-sided distribution function φ_{3n} of (7.4), given by

$$\varphi_{3n} = P(\sqrt{n}(\hat{\theta}_n - \theta_0)/\tau_n \le x_0)$$

for a given value $x_0 \in \mathbb{R}$. Hall, Horowitz and Jing (1995) consider both the NBB and the MBB estimators of φ_{3n} and derive expansions for the MSEs in the case of the sample mean, i.e., in the case where $\hat{\theta}_n = \bar{X}_n$ and $\theta = EX_1$. An expansion for the MSE of the MBB estimator $\hat{\varphi}_{3n}(\ell)$ (say) of φ_{3n} is obtained by Lahiri (1996d) under the Smooth Function Model (7.8). Here we follow the exposition of Lahiri (1996d) and describe an expansion for MSE $(\hat{\varphi}_{3n}(\ell))$ under the framework of Götze and Hipp (1983), introduced in Chapter 6. Suppose that $\{X_i\}_{i\in\mathbb{Z}}$ is defined on a probability space $(\Omega, \mathcal{F}, P)$, $\{X_i\}_{i\in\mathbb{Z}}$ is stationary, and that $\{\mathcal{D}_i\}_{i\in\mathbb{Z}}$ is a given sequence of sub-σ-fields of $\mathcal{F}$. For $-\infty \le a \le b \le \infty$, let $\mathcal{D}_a^b$ denote the smallest σ-field containing $\{\mathcal{D}_i : i \in [a, b] \cap \mathbb{Z}\}$. For easy reference, we now restate some of the conditions from Section 6.3, under the stationarity assumption on the process $\{X_i\}_{i\in\mathbb{Z}}$.

(C.1) There exists $\delta \in (0, 1)$ such that for all $n, m = 1, 2, \ldots$ with $m > \delta^{-1}$, there exists a $\mathcal{D}_{n-m}^{n+m}$-measurable random vector $X_{n,m}^\ddagger$ satisfying

$$E\big\|X_n - X_{n,m}^\ddagger\big\| \le \delta^{-1}\exp(-\delta m) .$$

(C.2) There exists $\delta \in (0, 1)$ such that for all $i \in \mathbb{Z}$, $m \in \mathbb{N}$, $A \in \mathcal{D}_{-\infty}^i$, and $B \in \mathcal{D}_{i+m}^\infty$,

$$\big|P(A \cap B) - P(A)P(B)\big| \le \delta^{-1}\exp(-\delta m) .$$

(C.3) There exists $\delta \in (0, 1)$ such that for all $m, n, k = 1, 2, \ldots$, and $A \in \mathcal{D}_{n-k}^{n+k}$

$$E\big|P(A \mid \mathcal{D}_j : j \ne n) - P(A \mid \mathcal{D}_j : 0 < |j-n| \le m+k)\big| \le \delta^{-1}\exp(-\delta m) .$$

(C.4) There exists $\delta \in (0, 1)$ such that for all $m, n = 1, 2, \ldots$ with $\delta^{-1} < m < n$, and for all $t \in \mathbb{R}^d$ with $\|t\| \ge \delta$,

$$E\Big|E\Big\{\exp(\iota t'[X_{n-m} + \cdots + X_{n+m}]) \mid \mathcal{D}_j : j \ne n\Big\}\Big| \le \delta^{-1}\exp(-\delta m) .$$

(C.5) $E\|X_1\|^{35+\delta} < \infty$ for some $\delta \in (0, 1)$.

Conditions (C.1)–(C.4) are restatements of Conditions (6.C.3)–(6.C.6) from Chapter 6, respectively. For a discussion of these conditions, see Chapter 6. We do not state Condition (6.C.2) separately here, as it follows from the conditional Cramer Condition (C.4) and the stationarity of $\{X_i\}_{i\in\mathbb{Z}}$. The moment Condition (C.5) is rather stringent. Lahiri (1996b) used this condition to prove negligibility of the remainder terms in the second-order Edgeworth expansion of the bootstrap distribution function estimator $\hat{\varphi}_{3n}(\ell)$ in the L^2-norm.

The following result gives an expansion for the MSE of $\hat{\varphi}_{3n}(\ell) \equiv \hat{\varphi}_{3n}(x_0;\ell)$ for a given $x_0 \in \mathbb{R}$.

Theorem 7.2 *Assume that Conditions (C.1)-(C.5) hold and that the smoothness Condition (5.D$_r$) of Section 5.4 on the function H holds with $r = 4$. Also, suppose that for some $\epsilon \in (0,3)$,*

$$\epsilon n^{\epsilon} \le \ell \le \epsilon^{-1} n^{1/3} \quad \text{for all} \quad n > \epsilon^{-1} . \tag{7.12}$$

Then, there exist constants $v_{31}, v_{32} \in (0,\infty)$ and $B_{31}, B_{32} \in \mathbb{R}$ such that for $|x_0| \neq 1$,

$$\begin{aligned} MSE(\hat{\varphi}_{3n}(x_0;\ell)) &= \Big[(x_0^2-1)\phi(x_0)\Big]^2 v_{31}^2 \cdot n^{-2}\ell^2 \\ &\quad + \Big[\phi(x_0)\{B_{31}+B_{32}(x_0^2-1)\}\Big]^2 n^{-1}\ell^{-2} \\ &\quad + o\Big(n^{-2}\ell^2 + n^{-1}\ell^{-2}\Big) , \end{aligned} \tag{7.13}$$

and for $|x_0| = 1$,

$$\begin{aligned} MSE(\hat{\varphi}_{3n}(x_0;\ell)) &= [\phi(x_0)]^2 v_{32}^2 \cdot n^{-2}\ell + [\phi(x_0)]^2 B_{31}^2 n^{-1}\ell^{-2} \\ &\quad + o\Big(n^{-2}\ell^2 + n^{-1}\ell^{-2}\Big) . \end{aligned} \tag{7.14}$$

Proof: See Lahiri (1996d). □

From the Edgeworth expansion results of Chapter 6 (cf. Theorem 6.7), it follows that

$$\begin{aligned} \hat{\varphi}_{3n}(x_0;\ell) &= \Phi(x_0) - n^{-1/2}\Big\{\hat{\mathcal{K}}_{31}(\ell) + (x_0^2-1)\hat{\mathcal{K}}_{32}(\ell)\Big\}\phi(x_0) \\ &\quad + O_p(n^{-1}) , \end{aligned}$$

where $\hat{\mathcal{K}}_{3i}(\ell) \equiv \hat{\mathcal{K}}_{3in}(\ell)$, $i = 1,2$ are smooth functions of certain bootstrap moments. For $|x_0| \neq 1$, the leading term of the variance of $\hat{\varphi}_{3n}(x_0;\ell)$ comes from the variance of the dominant term $n^{-1/2}(x_0^2-1)\hat{\mathcal{K}}_{32}(\ell)$, which is of the order $(n^{-1/2})^2 \cdot n^{-1}\ell^2$. In contrast, for $|x_0| = 1$, the term $n^{-1/2}(x_0^2-1)\hat{\mathcal{K}}_{31}(\ell)$ is zero and in this case, the leading term in the variance of $\hat{\varphi}_{3n}(x_0;\ell)$ is given by the variance of $n^{-1/2}\hat{\mathcal{K}}_{31}(\ell)$, which is of the order $(n^{-1/2})^2 \cdot n^{-1}\ell$.

On the other hand, the contribution to the bias of $\hat{\varphi}_{3n}(x_0;\ell)$ comes from both $\hat{\mathcal{K}}_{31}(\ell)$ and $\hat{\mathcal{K}}_{32}(\ell)$, each having a bias of the order ℓ^{-1}. This explains the sources of the various terms in the expansions for MSE$(\hat{\varphi}_{3n}(x_0;\ell))$ in (7.13) and (7.14). The exact forms of the population quantities v_{31}, v_{32}, B_{31}, and B_{32} are very complicated, and hence are not presented here. Interested readers are referred to Lahiri (1996d) for explicit expressions for these parameters. Interestingly, neither of the two empirical methods, that we describe in Sections 7.3 and 7.4 below for data-based selection of the optimal block sizes, requires explicit definitions of these parameters.

Theorem 7.2 readily yields the following asymptotic expressions for the optimal block lengths for estimating $\varphi_{3n}(x_0)$.

Corollary 7.2 *Assume that the conditions of Theorem 7.2 hold. Then, for $|x_0| \neq 1$,*

$$\begin{aligned} \ell_{3n}^0 \equiv \ell_{3n}^0(x_0) &= n^{1/4}\Big[\Big\{B_{31} + (x_0^2-1)B_{32}\Big\}^2 \Big/ \Big\{(x_0^2-1)v_{31}\Big\}^2\Big]^{1/4} \\ &\quad + o(n^{1/4}) \end{aligned} \tag{7.15}$$

and for $|x_0| = 1$,

$$\ell_{3n}^0 \equiv \ell_{3n}^0(x_0) = n^{1/3}\Big[2B_{31}^2/v_{32}^2\Big]^{1/3} + o(n^{1/3}) . \tag{7.16}$$

Thus, the optimal block length for estimating the distribution function of the normalized version of $\hat{\theta}_n$ is of the order $n^{1/4}$ at any given point $x_0 \in \mathbb{R}$, $|x_0| \neq 1$. For $|x_0| = 1$, the optimal order is $n^{1/3}$, the same as that for estimating the bias and variance parameters φ_{1n} and φ_{2n} (cf. (7.11)). Relations (7.15) and (7.16) give optional block lengths for *local* estimation of the distribution function of the pivotal quantity $\sqrt{n}(\hat{\theta}_n - \theta)/\tau_n$. The optimal block length for *global* estimation of the distribution function $\varphi_{3n}(\cdot) \equiv P(\sqrt{n}(\hat{\theta}_n - \theta)/\tau_n \le \cdot)$ can be obtained by minimizing an expansion for the (weighted) mean integrated squared error (MISE) of $\hat{\varphi}_{3n}(\cdot)$. An integration of the expansions (7.13) and (7.14) yields

$$\begin{aligned} \text{MISE}\Big(\hat{\varphi}_{3n}(\cdot;\ell)\Big) &\equiv E\int \Big[\hat{\varphi}_{3n}(x;\ell) - \varphi_{3n}(x;\ell)\Big]^2 w_0(x)dx \\ &= v_{33}^2 n^{-2}\ell^2 + B_{33}^2 n^{-1}\ell^{-2} + o(n^{-2}\ell^2 + n^{-1}\ell^{-2}) , \end{aligned} \tag{7.17}$$

where $w_0(\cdot) : \mathbb{R} \to (0,\infty)$ is a nonnegative weight function with $\int w_0(x)dx \in (0,\infty)$ and where $v_{33}^2 = v_{31}^2 \int (x^2-1)^2\phi(x)^2 w_0(x)dx$ and $B_{33}^2 = \int \phi(x)^2[B_{31} + B_{32}(x^2-1)]^2 w_0(x)dx$. Hence, the global optimal block length, defined as

$$\ell_{3n,\text{global}}^0 \equiv \text{argmin}\Big\{\text{MISE}\Big(\hat{\varphi}_{3n}(\cdot;\ell)\Big) : \epsilon n^{\epsilon} \le \ell \le \epsilon^{-1}n^{1/3}\Big\} \tag{7.18}$$

for a given $\epsilon \in (0, \frac{1}{3})$, is given by

$$\ell^0_{3n,\text{global}} = n^{1/4}\left[B_{33}^2/v_{33}^2\right]^{1/4} + o(n^{1/4}) . \quad (7.19)$$

Next, consider the two-sided distribution function $\varphi_{4n}(x_0) \equiv P(|\sqrt{n}(\hat{\theta}_n - \theta)/\tau_n| \le x_0)$, $x_0 \in (0, \infty)$. Hall, Horowitz and Jing(1995) give an expansion of the MSE of the MBB estimator $\hat{\varphi}_{4n}(x_0; \ell)$ for the case where $\hat{\theta}_n = \bar{X}_n$ and $\theta = EX_1$. In this case, they show that the optimal block length for estimating the level-2 parameter $\varphi_{4n}(x_0)$ is of the form

$$\ell^0_{4n} \equiv \ell^0_{4n}(x_0) = n^{1/5}C_0(x_0) + o(n^{+1/5}) \quad (7.20)$$

for some constant $C_0(x_0) \in (0, \infty)$. We refer the interested reader to Hall, Horowitz and Jing (1995) for further details in the two-sided case. Thus, one needs to use blocks of a smaller order (viz., $n^{1/5}$) for optimal estimation of probabilities assigned to symmetric intervals than those in the asymmetric case.

As pointed out in Chapter 1, the MSE and the optimal block length ℓ^0 are population-parameters that are determined by the sampling distributions of the bootstrap estimators of a level-2 parameter, and therefore, may be regarded as level-3 parameters. Thus, a general approach to the estimation of $\text{MSE}(\hat{\varphi}_n(\ell))$ and ℓ^0 is to apply two rounds of resampling methods iteratively. In Sections 7.3 and 7.4, we describe two such general methods, proposed by Hall, Horowitz and Jing (1995) and Lahiri, Furukawa and Lee (2003), respectively. The method proposed by Hall, Horowitz and Jing (1995) uses a combination of subsampling and bootstrapping the data. The other method, proposed by Lahiri, Furukawa and Lee (2003), is based on the Jackknife-After-Bootstrap method and it uses a combination of jackknifing and bootstrapping the data. In the same vein, one may use two rounds of block bootstrapping to estimate the level-3 parameters $\text{MSE}(\hat{\varphi}_n(\ell))$ and ℓ^0, although properties of this third alternative remain unexplored at this time. Estimation methods tailored to estimate the optimal block size for a specific functional are also known. For the case of the variance functional, Bühlmann and Künsch (1999b) propose some novel plug-in estimators of the optional block length for block bootstrap variance estimation and establish their convergence rates. For a more direct plug-in method in the variance functional case, see Politis and White (2003). They employ the "flat-top" kernel method of Politis and Romano (1995) to estimate the relevant population parameters in the leading term of the optimal block size given by Corollary 7.1 above.

7.3 A Method Based on Subsampling

In this section, we describe the Hall, Horowitz and Jing (1995) method for choosing the theoretical optimal block size. For concreteness, suppose that

$\hat{\varphi}_n(\ell)$ denotes the MBB estimator of the level-2 parameter φ_n, based on blocks of length ℓ, where n is the sample size. Furthermore, suppose that the MSE of $\hat{\varphi}_n(\ell)$ admits an expansion of the form

$$\text{MSE}(\hat{\varphi}_n(\ell)) = a_n\left[C_1 n^{-1}\ell^r + C_2\ell^{-2}\right](1 + o(1)) \quad \text{as} \quad n \to \infty \quad (7.21)$$

for some constants $C_1, C_2 \in (0, \infty)$, $r \in \mathbb{N}$, and for some sequence $\{a_n\}_{n\ge1}$ of positive real numbers, over a suitable set $\mathcal{J}_n \subset \mathbb{N}$ of block sizes. We shall assume that the set $\mathcal{J}_n$ contains the set $[n^{\frac{1}{r+2}-\epsilon}, n^{\frac{1}{r+2}+\epsilon}]$ for some small $\epsilon \in (0, 1)$. Next, define the optimal block length ℓ^0_n by

$$\ell^0_n \equiv \text{argmin}\left\{\text{MSE}(\hat{\varphi}_n(\ell)) : l \in \mathcal{J}_n\right\} . \quad (7.22)$$

Note that by (7.21) and (7.22), the optimal block length ℓ^0_n is of the order $n^{\frac{1}{r+2}}$. To define the Hall, Horowitz and Jing (1995) estimator of the theoretical optimal block length ℓ^0_n, we proceed as follows. Let $m \equiv m_n$ be a sequence of real numbers satisfying

$$m^{-1} + n^{-1}m = o(1) \quad \text{as} \quad n \to \infty . \quad (7.23)$$

Consider the subsamples $\mathcal{X}_{i,m} \equiv (X_i, \ldots, X_{i+m-1})$, $i = 1, \ldots, n - m + 1$ of length m. Let φ_m denote the level-2 parameter φ_n at $n = m$. For each $i = 1, \ldots, n - m + 1$, let $\hat{\varphi}_{m,i}(\ell)$ be the MBB estimator of φ_m obtained by resampling blocks of length ℓ from the m observations $\mathcal{X}_{i,m}$. Next define the subsampling estimator of $\text{MSE}(\hat{\varphi}_m(\ell))$, the mean squared error of the MBB estimator of the level-2 parameter φ_m based on a sample of size m, as

$$\widehat{\text{MSE}}_m(\ell) = (n - m + 1)^{-1}\sum_{i=1}^{n-m+1}\left[\hat{\varphi}_{m,i}(\ell) - \hat{\varphi}_n(\ell^*_n)\right]^2 , \quad (7.24)$$

where ℓ^*_n is a plausible pilot block size. Let

$$\hat{\ell}^0_m = \text{argmin}\left\{\widehat{\text{MSE}}_m(\ell) : \ell \in \mathcal{J}_m\right\} , \quad (7.25)$$

where we employ the set $\mathcal{J}_m$ (not $\mathcal{J}_n$) to define $\hat{\ell}^0_m$. Then, $\hat{\ell}^0_m$ is an estimator of the theoretical optimal block length when the sample size is m. We need to rescale this initial estimator to get an estimator of ℓ^0_n of (7.22). Since the optimal block length ℓ^0_n in (7.22) is of the order $n^{\frac{1}{r+2}}$, the right scaling factor here is $[n/m]^{\frac{1}{r+2}}$. The Hall, Horowitz and Jing (1995) estimator of ℓ^0_n is given by

$$\hat{\ell}^0_n = (\hat{\ell}^0_m) \cdot [n/m]^{\frac{1}{r+2}} . \quad (7.26)$$

Note that the Hall, Horowitz and Jing (1995) estimation method is applicable quite generally, requiring only that the MSE of the bootstrap estimator has (an expansion of) the form (7.21) for some $r \ge 1$ and that the subsampling estimator $\widehat{\text{MSE}}_m(\ell)$ of $\text{MSE}(\hat{\varphi}_m(\ell))$ converges in some suitable sense, say, in probability. In particular, the method is applicable even without an explicit expression for the constants C_1 and C_2 in (7.21). Similarly, the method can be applied when a block bootstrap method other than the MBB is employed. A set of sufficient conditions for the consistency of the subsampling estimator $\widehat{\text{MSE}}_m(\ell)$ are that the series $\{X_i\}_{i\in\mathbb{Z}}$ is stationary and has an absolutely summable strong mixing coefficient.

From the description of the method, it is clear that accuracy of the Hall, Horowitz and Jing (1995) method depends on the choices of the subsampling parameter m and the pilot block size ℓ_n^*. The optimal order of m is unknown at this stage. However, for the other smoothing parameter, viz., the pilot block size ℓ_n^*, Hall, Horowitz and Jing (1995) suggest a way to reduce the effect of ℓ_n^* on the optimal block length estimator ℓ_n^0 of (7.26). To reduce the effect of ℓ_n^*, they suggest iterating the main steps of the algorithm, by replacing the pilot block size ℓ_n^* with the estimated value $\hat{\ell}_n^0$ for the second iteration, and repeating this process until convergence. However, convergence of this iterative scheme is not guaranteed (see the numerical example below).

We now describe the results of a small simulation study on finite sample properties of the Hall, Horowitz and Jing (1995) method. We consider the time series model

$$X_i = (\epsilon_i + \epsilon_{i-1})/\sqrt{2},\ i \in \mathbb{Z} \tag{7.27}$$

where $\{\epsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of iid random variables with common distribution $(\chi^2(1) - 1)$, the centered Chi-squared distribution with one degree of freedom. Thus, $E\epsilon_1 = 0$ and $E\epsilon_1^2 = 2$. We took the level-1 parameter θ as EX_1, and the estimator $\hat{\theta}_n$ as $\hat{\theta}_n = \bar{X}_n$, the sample mean with sample size $n = 125$. The level-2 parameters of interest are given by (cf. (7.3) and (7.4))

$$\varphi_{2n} = n.\text{Var}(\bar{X}_n) \tag{7.28}$$

and

$$\begin{aligned}\varphi_{3n} &= P\Big(\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\tau_n} \le 0\Big) \\ &= P(\hat{\theta}_n \le \theta)\ .\end{aligned} \tag{7.29}$$

True values of φ_{2n} and φ_{3n} were found by 20,000 simulation runs. These are given by $\varphi_{2n} = 3.984$ and $\varphi_{3n} = .5226$.

To find the theoretical optimal block lengths for φ_{2n} and φ_{3n} , we applied the MBB method to generate block bootstrap estimators of the level-2 parameters φ_{2n} and φ_{3n} with several values of the block length ℓ. Table 7.1 below gives the expected value (Mean), the bias, the standard deviation (SD) and the MSE's of the MBB estimators based on 1000 simulation runs. From the table, it is evident that the optimal block lengths for estimating φ_{2n} and φ_{3n} are respectively given by $\ell_{2n}^0 = 3$ and $\ell_{3n}^0 = 2$. Next the subsampling-based method of Hall, Horowitz and Jing (1995) was applied to select an optional block size empirically. We chose the subsample size $m = 30$, and the pilot block size parameter $\ell_{kn}^* = 5$ for both φ_{2n} and φ_{3n}. Thus, for $k = 1, 2$, the MSE estimator $\widehat{\text{MSE}}_m(\ell)$ of (7.24) for the level-2 parameter φ_{kn} is now evaluated by resampling overlapping blocks of size ℓ from each of the 96 ($= 125 - 30 + 1$) overlapping subsamples of size $m = 30$ and then computing the MBB estimators $\hat{\varphi}_{km,i}(\ell)$ (say) of φ_{kn} for the ith subsample, for $i = 1, \ldots, 96$. The centering value $\hat{\varphi}_{kn}(\ell_{kn}^*)$ in $\widehat{\text{MSE}}_m(\ell)$ is computed using the full sample of size $n = 125$, with $\ell_{kn}^* = 5$ for both k. All bootstrap estimates (including those related to Table 7.2 below) were evaluated using 800 Monte-Carlo replicates.

TABLE 7.1. Determination of the true optimal block sizes for MBB estimation of the level-2 parameters φ_{2n} and φ_{3n} of (7.28) and (7.29) for model (7.27). The results are based on 1000 simulation runs. An asterisk(*) denotes the minimun MSE value for a functional.

		(a) Variance Estimation		
L	Mean	Bias	SD	MSE
1	1.947	−2.037	0.705	4.645
2	2.902	−1.082	1.089	2.358
3	3.204	−0.780	1.244	2.157*
4	3.320	−0.664	1.334	2.221
5	3.394	−0.590	1.412	2.341
6	3.437	−0.547	1.482	2.497
7	3.452	−0.532	1.542	2.660
8	3.460	−0.524	1.594	2.814
9	3.460	−0.524	1.648	2.990
10	3.469	−0.515	1.713	3.198

		(b) Distribution Function Estimation		
L	E.phi	Bias	SD	MSE
1	0.5099	−0.0126	0.0136	0.000345
2	0.5132	−0.0094	0.0132	0.000262*
3	0.5127	−0.0099	0.0142	0.000299
4	0.5136	−0.0089	0.0139	0.000272
5	0.5123	−0.0103	0.0144	0.000313
6	0.5125	−0.0100	0.0149	0.000322
7	0.5125	−0.0100	0.0150	0.000324
8	0.5121	−0.0105	0.0154	0.000347
9	0.5123	−0.0103	0.0157	0.000352
10	0.5103	−0.0122	0.0164	0.000419

Table 7.2 gives the frequency distribution of the optimal block size estimator $\hat{\ell}_{kn}^0$ for φ_{kn}, computed by formula (7.26) using 500 simulation runs. As in Hall, Horowitz and Jing (1995), in this simulation study also, the optimal block size estimators converged after a couple of iterations in a majority of the cases. However, in some instances, there was a circular behavior of the estimated optimal block size in successive iterations (e.g., the initial value 5 led to 8 which led to 3 and then, 3 led back to 5). The frequency of such cases is given under the value -1. This problem appeared to be more prevalent for distribution function estimation (i.e., for φ_{3n} of (7.29)) than for variance estimation (i.e., for φ_{2n} of (7.28)). In such a situation, one may pick a value of $\hat{\ell}_{kn}^0$ (from the set of all optimal block lengths in different iterations) that corresponds to the minimum estimated $\widehat{\text{MSE}}_m(\ell)$.

Parts (a) and (b) of Table 7.2 show that for both level-2 parameters φ_{2n} and φ_{3n}, the estimated optimal block sizes have a pronounced mode at the true optional block sizes, i.e., at $\ell_{2n}^0 = 3$ for φ_{2n} and at $\ell_{3n}^0 = 2$ for φ_{3n}. Furthermore, the distribution of the estimated optimal block size for variance estimation has a longer right tail compared to that for the distribution function estimation. However, the performance of this method improves as the sample size n increases. See Hall, Horowitz and Jing (1995) for further numerical examples and discussions.

TABLE 7.2. Frequency distribution of the optimal block sizes selected by the Hall, Horowitz and Jing (1995) method for model (7.27) with $n = 125$, $m = 30$, and initial block size $\ell_{kn}^* = 5$, $k = 1, 2$. Results are based on 500 simulation runs. The value -1 of $\hat{\ell}_{kn}^0$, $k = 1, 2$, corresponds to the cases where the iterations of the method failed to converge.

(a) Variance Estimation

$\hat{\ell}_{2n}^0$	-1	2	3	5	7	9	10	12	14	15	17	19	21
Freq.	35	137	200	63	18	12	13	8	6	2	3	1	2

(b) Distribution Function Estimation

$\hat{\ell}_{3n}^0$	-1	2	3	4	6	7	9	10	12	13
Freq.	137	276	50	5	7	13	8	1	2	1

7.4 A Nonparametric Plug-in Method

In this section, we describe a plug-in method for selecting the optimal block length based on a recent work of Lahiri, Furukawa and Lee (2003). The plug-in method estimates the leading term in the first-order expansion of the optimal block length using a resampling method, and does not require an explicit expression for the level-3 population parameters.

In Section 7.4.1, we describe the motivation and basic construction of the plug-in estimator and in Section 7.4.2, we describe estimation of the level-3 parameter associated with the bias part of the block bootstrap estimators. Estimation of the level-3 parameter associated with the variance part employ the Jackknife-After-Bootstrap (JAB) method of Efron (1992) and Lahiri (2002a). In Section 7.4.3, we describe the JAB method for dependent data. The nonparametric plug-in estimators of the optimal block lengths are presented in Section 7.4.4. Some finite sample results are given in Section 7.4.5. In all of Section 7.4, we restrict attention to the optimal block lengths for the MBB method.

7.4.1 Motivation

Let φ_n be a level-2 parameter of interest and let $\hat{\varphi}_n(\ell)$ be a block bootstrap estimator of φ_n based on blocks of length ℓ. From the discussion of Section 7.2, it follows that under suitable regularity conditions, the variance of $\hat{\varphi}_n(\ell)$ and the bias of $\hat{\varphi}_n(\ell)$ admit expansions of the form

$$n^{2a} \cdot \text{Var}(\hat{\varphi}_n(\ell)) = vn^{-1}\ell^r + o(n^{-1}\ell^r) \tag{7.30}$$

and

$$n^a \cdot \text{Bias}(\hat{\varphi}_n(\ell)) = B\ell^{-1} + o(\ell^{-1}) \tag{7.31}$$

for some population parameters $B \in \mathbb{R}$, $v \in (0, \infty)$ and for some known constants $a \in (0, \infty)$, $r \in \mathbb{N}$. For example, for the bias and variance functionals $\varphi_n = \varphi_{1n}, \varphi_{2n}$, $r = 1$, and $a = 1$, while for the distribution function (at a given point x_0) $\varphi_n = \varphi_{3n}(x_0)$ with $|x_0| \neq 1$, $r = 2$ and $a = 1/2$. In this case, the MSE-optimal block size $\ell_n^0 \equiv \ell_n^0(\varphi_n)$ is given by

$$\ell_n^0 = \left(\frac{2B^2}{rv}\right)^{\frac{1}{r+2}} n^{\frac{1}{r+2}} (1 + o(1)) . \tag{7.32}$$

Like any other plug-in method, the nonparametric plug-in method focuses on the leading term $\left(\frac{2B^2}{rv}\right)^{\frac{1}{r+2}} n^{\frac{1}{r+2}}$ but estimates the level-3 parameters B and v nonparametrically, as follows. Note that from (7.30) and (7.31), we have

$$\lim_{n\to\infty} (n^{-1}\ell^r)^{-1} n^{2a}\, \text{Var}(\hat{\varphi}_n(\ell)) = v \tag{7.33}$$

and

$$\lim_{n\to\infty} \ell \cdot n^a\, \text{Bias}(\hat{\varphi}_n(\ell)) = B . \tag{7.34}$$

This suggests that consistent estimators of v and B may be derived if we can estimate $\text{Var}(\hat{\varphi}_n(\ell))$ and $\text{Bias}(\hat{\varphi}_n(\ell))$ consistently. Let $\widehat{\text{VAR}}_n$ and $\widehat{\text{BIAS}}_n$ be nonparametric estimators of $\text{Var}(\hat{\varphi}_n(\ell))$ and $\text{Bias}(\hat{\varphi}_n(\ell))$, respectively, that are consistent in the following sense:

$$\frac{\widehat{\text{VAR}}_n}{\text{Var}(\hat{\varphi}_n(\ell_1))} \longrightarrow_p 1 \quad \text{as} \quad n \to \infty \tag{7.35}$$

and

$$\frac{\widehat{\mathrm{BIAS}}_n}{\mathrm{Bias}(\hat{\varphi}_n(\ell_1))} \longrightarrow_p 1 \quad \text{as} \quad n \to \infty \tag{7.36}$$

along some suitable sequence $\{\ell_1\} \equiv \{\ell_{1n}\}_{n\geq 1}$.

Then, using (7.33) and (7.34), we define estimators of the parameters v and B as

$$\hat{v}_n \equiv \hat{v}_n(\ell_1) = n^{2a}\, \widehat{\mathrm{VAR}}_n \cdot (n^{-1}\ell_1^r)^{-1}, \tag{7.37}$$

and

$$\hat{B}_n = \hat{B}_n(\ell_1) = n^a\, \widehat{\mathrm{BIAS}}_n \cdot \ell_1 . \tag{7.38}$$

The nonparametric plug-in estimator $\hat{\ell}_n^0$ of the optimal block length ℓ_n^0 is then given by replacing the level-3 parameters v and B in the leading term in (7.32) by the above estimators, i.e., by

$$\hat{\ell}_n^0 = \left[\frac{2\hat{B}_n^2}{r\hat{v}_n}\right]^{\frac{1}{r+2}} n^{\frac{1}{r+2}} . \tag{7.39}$$

It is clear that the performance of the estimator $\hat{\ell}_n^0$ depends on the sequence $\{\ell_{1n}\}_{n\geq 1}$, on the level-2 parameter φ_n, and on the basic estimators $\widehat{\mathrm{VAR}}_n$ and $\widehat{\mathrm{BIAS}}_n$ employed in the construction of $\hat{v}_n$ and $\hat{B}_n$ in (7.37) and (7.38), respectively. In the next section we describe the plug-in method of Lahiri, Furukawa and Lee (2003) who used the JAB method for estimating $\mathrm{Var}(\hat{\varphi}_n(\ell))$ and constructed an estimator of $\mathrm{Bias}(\hat{\varphi}_n(\ell))$ by combining two block bootstrap estimators suitably. The use of these basic estimators were prompted by considerations regarding computational efficacy and accuracy of the proposed plug-in method. As explained below, the JAB variance estimator has some computational advantage over other common resampling methods in that the JAB variance estimator can be computed by *reusing* the block bootstrap resamples used in the Monte-Carlo evaluation of $\hat{\varphi}_n(\ell_1)$, and thus, do not involve iterated levels of resampling. Similarly, the bias estimator proposed in Lahiri, Furukawa and Lee (2003) also involves a single level of resampling. In the section below, we describe further details of the construction.

7.4.2 The Bias Estimator

For constructing the bias estimator, we begin with relation (7.31), which gives an asymptotic representation for the bias part of the bootstrap estimator $\hat{\varphi}_n(\ell)$ and may be rewritten as

$$E\hat{\varphi}_n(\ell) = \varphi_n + \frac{B}{n^a \ell} + o(n^{-a}\ell^{-1}) \quad \text{as} \quad n \to \infty . \tag{7.40}$$

If (7.40) holds for the sequences $\{\ell_1\} \equiv \{\ell_{1n}\}_{n\geq 1}$ and $\{2\ell_1\} \equiv \{2\ell_{1n}\}_{n\geq 1}$, then we may combine the corresponding expansions to conclude that

$$\begin{aligned} & E\Big[\hat{\varphi}_n(\ell_1) - \hat{\varphi}_n(2\ell_1)\Big] \\ = \;& \Big[\Big\{\varphi_n + \frac{B}{n^a\ell_1} + o(n^{-a}\ell_1^{-1})\Big\} - \Big\{\varphi_n + \frac{B}{2n^a\ell_1} + o(n^{-a}\ell_1^{-1})\Big\}\Big] \\ = \;& \frac{B}{2n^a\ell_1} + o(n^{-a}\ell_1^{-1}) \quad \text{as} \quad n \to \infty . \end{aligned} \tag{7.41}$$

This suggests that a consistent estimator of $\mathrm{Bias}(\hat{\varphi}_n(\ell_1))$ satisfying (7.36) may be constructed as

$$\widehat{\mathrm{BIAS}}_n \equiv \widehat{\mathrm{BIAS}}_n(\ell_1) = 2\big(\hat{\varphi}_n(\ell_1) - \hat{\varphi}_n(2\ell_1)\big) . \tag{7.42}$$

Indeed, if the optimal order of the block length for estimating φ_n is $n^{\frac{1}{r+2}}$ (cf. (7.32)), then by Cauchy-Schwarz inequality, it follows that for any sequence $\{\ell_1\} = \{\ell_{1n}\}_{n\geq 1}$ satisfying the requirement

$$1 \ll \ell_1 \ll n^{\frac{1}{r+2}} \quad \text{as} \quad n \to \infty , \tag{7.43}$$

$\widehat{\mathrm{BIAS}}_n$ is consistent. A specific choice of $\{\ell_{1n}\}_{n\geq 1}$ will be suggested in Section 7.4.4 for the plug-in estimator $\hat{\ell}_n^0$ of (7.39). Note that, as pointed out earlier, the estimator $\widehat{\mathrm{BIAS}}_n$ is based on only two block bootstrap estimator of φ_n and may be computed using only one level of resampling.

In the next section, we describe the JAB method for dependent data. Readers familiar with the method may skip this section and proceed to Section 7.4.4.

7.4.3 The JAB Variance Estimator

The JAB method was proposed by Efron (1992) for assessing accuracy of bootstrap estimators based on the IID bootstrap method for independent data. A modified version of the method for block bootstrap estimators in the case of dependent data was proposed by Lahiri (2002a). The JAB method for dependent data applies a version of the block jackknife method to a block bootstrap estimator. For the sake of completeness, first we briefly describe the block jackknife method.

Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be the observations and let $\hat{\gamma}_n \equiv t_n(\mathcal{X}_n)$ be an estimator of a level-1 parameter of interest γ. The block jackknife method systematically deletes blocks of consecutive observations to define the jackknife copies (called the *block jackknife point values*) of $\hat{\gamma}_n$ and combines these to produce estimators of the bias and the variance of $\hat{\gamma}_n$. Like the block bootstrap methods, different versions of the block jackknife method, such as, overlapping, nonoverlapping, and weighted block jackknife methods have been proposed in the literature (cf. Künsch (1989), Liu and Singh

(1992)). Here we describe the overlapping version of the block jackknife or the moving blocks jackknife (MBJ) of Künsch (1989) and Liu and Singh (1992). (Like the term "MBB," the term MBJ was also introduced by Liu and Singh (1992)). Let $m \equiv m_n$ be a sequence of integers such that m goes to infinity but at a rate slower than n, i.e.,

$$m^{-1} + n^{-1}m = o(1) \quad \text{as} \quad n \to \infty . \tag{7.44}$$

Here m denotes the number of observations (or the size of the block) to be deleted for defining the MBJ point values. For $i = 1, \ldots, n-m+1$, let $\mathcal{X}_{n,i} = \mathcal{X}_n \backslash \{X_i, \ldots, X_{i+m-1}\}$ denote the set of observations after the block $\{X_i, \ldots, X_{i+m-1}\}$ of size m has been deleted from $\mathcal{X}_n$. Then, the ith *MBJ point value* $\hat{\gamma}_n^{(i)}$ is defined as

$$\hat{\gamma}_n^{(i)} = t_{n-m}(\mathcal{X}_{n,i}), \ i = 1, \ldots, n-m+1 . \tag{7.45}$$

The MBJ estimator of the variance of $\hat{\gamma}_n$ is given by

$$\widehat{\text{VAR}}_{\text{MBJ}}(\hat{\gamma}_n) = \frac{m}{(n-m)} \cdot \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \left(\tilde{\gamma}_n^{(i)} - \hat{\gamma}_n\right)^2 , \tag{7.46}$$

where $\tilde{\gamma}_n^{(i)} \equiv m^{-1}\big(n\hat{\gamma}_n - (n-m)\hat{\gamma}_n^{(i)}\big)$ is the ith *MBJ pseudo-value* corresponding to $\hat{\gamma}_n$. For consistency and finite sample properties of the MBJ and its other variants, we refer the reader to Künsch (1989), Liu and Singh (1992), Shao and Tu (1995), Davison and Hinkley (1997), and the references therein. Note that, if we set $m \equiv 1$, i.e., if we delete a single observation at a time, then the MBJ variance estimator in (7.46) reduces to the classical delete-1 jackknife variance estimator for independent data

$$\widehat{\text{VAR}}_J(\hat{\gamma}_n) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\tilde{\gamma}_n^{(i)} - \hat{\gamma}_n\right)^2 . \tag{7.47}$$

For properties of the jackknife method for independent data, see Miller (1974), Efron (1982), Wu (1990), Liu and Singh (1992), Efron and Tibshirani (1993), Shao and Tu (1995), Davison and Hinkley (1997), and the references therein.

Next we describe the JAB method for dependent data. Let $\hat{\varphi}_n \equiv \hat{\varphi}_n(\ell)$ be the MBB estimator of a level-2 parameter φ_n based on (overlapping) blocks of size ℓ from $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. Let $\mathcal{B}_i = \{X_i, \ldots, X_{i+\ell-1}\}, i = 1, \ldots, N$ (with $N = n-\ell+1$) denote the collection of all overlapping blocks contained in $\mathcal{X}_n$ that are used for defining the MBB estimator $\hat{\varphi}_n$. Also, let m be an integer such that (7.44) holds. Note that the MBB estimator $\hat{\varphi}_n(\ell)$ is defined in terms of the "basic building blocks" $\mathcal{B}_i$'s. Hence, instead of deleting blocks of original observations $\{X_i, \ldots, X_{i+m-1}\}$, as done in the MBJ method described above, the JAB method of Lahiri (2002a) defines the jackknife point-values by deleting blocks of $\mathcal{B}_i$'s. Later in this section, we will discuss how this simple modification plays an important role in ensuring computational efficacy of the JAB method.

Since there are N observed blocks of length ℓ, we can define $M \equiv N - m + 1$ many JAB point-values corresponding to the bootstrap estimator $\hat{\varphi}_n$, by deleting the *overlapping* "blocks of blocks" $\{\mathcal{B}_i, \ldots, \mathcal{B}_{i+m-1}\}$ of size m for $i = 1, \ldots, M$. Let $I_i^0 = \{1, \ldots, N\} \backslash \{i, \ldots, i+m-1\}$, $i = 1, \ldots, M$. To define the ith JAB point-value $\hat{\varphi}_n^{(i)} \equiv \hat{\varphi}_n^{(i)}(\ell)$, we need to resample $b = \lfloor n/\ell \rfloor$ blocks randomly and with replacement from the reduced collection $\{\mathcal{B}_j : j \in I_i^0\}$ and construct the MBB estimator of φ_n using these resampled blocks. More precisely, suppose that $T_n = t_n(\mathcal{X}_n; \theta)$ be a random variable with probability distribution G_n and let $\varphi_n = \varphi(G_n)$ for some functional φ. Let $J_{i1}, \ldots, J_{ib}$ be a collection of b random variables such that, conditional on $\mathcal{X}_n$, these are iid with common distribution

$$P_*(J_{i1} = j) = (N-m)^{-1} \quad \text{for all} \quad j \in I_i^0 . \tag{7.48}$$

Then, the resampled blocks to be used for defining the JAB point-value $\hat{\varphi}_n^{(i)}$ are given by

$$\left\{\mathcal{B}_j^{*(i)} \equiv \mathcal{B}_{J_{ij}} : j = 1, \ldots, b\right\} . \tag{7.49}$$

Let $\mathcal{X}_n^{*(i)}$ denote the resampled data obtained by concatenating $\{\mathcal{B}_j^{*(i)}, j = 1, \ldots, b\}$. Also, let $T_n^{*(i)} \equiv t_{n_1}(\mathcal{X}_n^{*(i)}; \tilde{\theta}_{n,i})$ be the MBB version of T_n, defined using the resampled data $\mathcal{X}_n^{*(i)}$ and using a suitable estimator $\tilde{\theta}_{n,i}$ of θ. Then, the JAB point-value $\hat{\varphi}_n^{(i)}$ is given by applying the functional φ to the conditional distribution $\hat{G}_{n,i}$ (say) of $T_n^{*(i)}$ as

$$\hat{\varphi}_n^{(i)} = \varphi(\hat{G}_{n,i}) . \tag{7.50}$$

For an example illustrating the definition of $T_n^{*(i)}$, suppose that

$$T_n = \sqrt{n}(\hat{\theta}_n - \theta) \tag{7.51}$$

with $\hat{\theta}_n = H(\bar{X}_n)$ and $\theta = H(\mu)$ for some (smooth) function $H : \mathbb{R}^d \to \mathbb{R}$, where $\{X_i\}_{i \in \mathbb{Z}}$ is a stationary sequence of $\mathbb{R}^d$-valued random vectors, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\mu = EX_1$. Let $\bar{X}_n^{*(i)}$ denote the MBB sample mean based on the $n_1 = b.\ell$ resampled values in $\{\mathcal{B}_j^{*(i)}, j = 1, \ldots, b\}$. Then, the MBB version $T_n^{*(i)}$ for the ith JAB point-value is defined as

$$T_n^{*(i)} = \sqrt{n_1}\left(\theta_n^{*(i)} - \tilde{\theta}_{n,i}\right) , \tag{7.52}$$

where $\theta_n^{*(i)} = H(\bar{X}_n^{*(i)})$ and where we set $\tilde{\theta}_{n,i} = H(\hat{\mu}_{n,i})$ with $\hat{\mu}_{n,i} = E_*\bar{X}_n^{*(i)}$, $i = 1, \ldots, M$.

Next we return to the general case of $T_n \equiv t_n(\mathcal{X}_n;\theta)$ and define the JAB variance estimator of $\hat\varphi_n$ as (cf. (7.46))

$$\widehat{\text{VAR}}_{\text{JAB}}(\hat\varphi_n) = \frac{m}{(N-m)} \cdot \frac{1}{M}\sum_{i=1}^{M}\left(\tilde\varphi_n^{(i)} - \hat\varphi_n\right)^2 , \tag{7.53}$$

where $\tilde\varphi_n^{(i)} = m^{-1}(N\hat\varphi_n - (N-m)\hat\varphi_n^{(i)})$ denotes the ith JAB *pseudo-value* corresponding to $\hat\varphi_n$ and where $\hat\varphi_n^{(i)}$'s are defined by (7.50). As with the given MBB estimator $\hat\varphi_n$, computation of the point-values $\hat\varphi_n^{(i)}$'s and hence, of the pseudo-values $\tilde\varphi_n^{(i)}$ are typically done using the Monte-Carlo method. A simple representation result, initially noted by Efron (1992) in the context of IID bootstrap, makes it possible to compute the JAB variance estimator by reusing the resampled blocks used in the computation of the given bootstrap estimator $\hat\varphi_n$. We now give a statement of this result below.

Proposition 7.1 *Let $J_1,\dots,J_b$ be iid random variables with the Discrete Uniform Distribution on $\{1,\dots,N\}$ and let $J_{i1},\dots,J_{ib}$ be iid random variables with the Discrete Uniform Distribution on I_i^0, $1 \le i \le M$. Let $\hat p_i = b^{-1}\sum_{j=1}^{b}\mathbb{1}(J_j \in I_i^0)$, $1 \le i \le M$. Then, for any $i = 1,\dots,M$, the conditional distribution of $(J_1,\dots,J_b)$ given $\hat p_i = 1$ is the same as the unconditional distribution of $(J_{i1},\dots,J_{ib})$.*

Proof: For any $j_1,\dots,j_b \in I_i^0$,

$$\begin{aligned} & P(J_1 = j_1,\dots,J_b = j_b \mid \hat p_i = 1) \\ = {} & P(J_1 = j_1,\dots,J_b \in j_b)/P(\hat p_i = 1) \\ = {} & [N^{-b}]/[(N-m)/N]^b \\ = {} & (N-m)^{-b} = P(J_{i1} = j_1,\dots,J_{ib} = j_b) . \end{aligned}$$

This completes the proof of the proposition. □

To appreciate the relevance of this result, suppose that $\{{}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*\}$, $k = 1,\dots,K$ denote the set of blocks drawn randomly, with replacement from the collection $\{\mathcal{B}_1,\dots,\mathcal{B}_N\}$ for the Monte-Carlo evaluation of the given block bootstrap estimator $\hat\varphi_n$. Let $\{{}_kJ_1,\dots,{}_kJ_b\}$ denote the random indices corresponding to $\{{}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*\}$, i.e., ${}_k\mathcal{B}_j^* = {}_k\mathcal{B}_{{}_kJ_j}$, $1 \le j \le b$, $k = 1,\dots,K$. Then for any k, if all b indices ${}_kJ_1,\dots,{}_kJ_b$ lie in I_i^0, by Proposition 7.2, we may consider $({}_kJ_1,\dots,{}_kJ_b)$ as a random sample of size b from the reduced index set $I_i^0 = \{1,\dots,N\}\backslash\{i,\dots,i+m-1\}$. Let

$$I_i^* = \{k : 1 \le k \le K,\ {}_kJ_j \in I_i^0 \quad \text{for all} \quad j = 1,\dots,b\}$$

denote the index set of all such random vectors $({}_kJ_1,\dots,{}_kJ_b)$. Then, $\{({}_kJ_1,\dots,{}_kJ_b) : k \in I_i^*\}$ gives us an iid collection of random vectors (of possibly different sizes for different $i \in \{1,\dots,M\}$), each having the same distribution as $(J_{i1},\dots,J_{ib})$ of the Proposition. Thus, the resamples for computing the ith JAB point-value $\hat\varphi_n^{(i)}$ may be obtained by *extracting* the subcollection $\{({}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*) : k \in I_i^*\}$ from the original resamples $\{({}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*) : 1 \le k \le K\}$, and no additional resampling is needed. The Monte-Carlo approximations generated by this method are close to the true values of $\hat\varphi_n^{(i)}$'s, provided K is large.

As an illustration, consider the random variable T_n of (7.51) and suppose that the level-2 parameter of interest is $\varphi_n = \varphi(G_n)$ for some functional φ where G_n is the sampling distribution of T_n. Figures 7.1 and 7.2 give a schematic description of the main steps involved in the computations of the MBB estimator $\hat\varphi_n$ and its JAB point-values $\hat\varphi_n^{(i)}$, $i = 1,\dots,M$. For computing $\hat\varphi_n$, we generate K iid sets of b many blocks $\{{}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*\}$ for $k = 1,\dots,K$, compute the bootstrap sample mean ${}_k\bar X_n^*$ and the bootstrap version ${}_kT_n^* = \sqrt{n}({}_k\theta_n^* - \tilde\theta_n)$ for each set with ${}_k\theta_n^* = H({}_k\bar X_n^*)$ and $\tilde\theta_n = H(\hat\mu_n)$. Then, the Monte-Carlo approximation to $\hat\varphi_n$ is given by $\varphi(G_n^*)$ where G_n^* denotes the empirical distribution of the bootstrap replicates $\{{}_kT_n^* : k = 1,\dots,K\}$. For computing $\hat\varphi_n^{(i)}$, we scan the K sets of resampled blocks $\{{}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*\}$, $k = 1,\dots,K$ and extract the ${}_k\theta_n^*$-values corresponding to the block-sets $\{{}_k\mathcal{B}_1^*,\dots,{}_k\mathcal{B}_b^*\}$ that do not contain any of the blocks $\mathcal{B}_i,\dots,\mathcal{B}_{i+m-1}$. Next, the bootstrap version of $T_n^{*(i)}$ are computed by employing these ${}_k\theta_n^*$'s in the formula ${}_kT_n^{*(i)} = \sqrt{n_1}({}_k\theta_n^* - \tilde\theta_{n,i})$ where $\tilde\theta_{n,i} \equiv H(\hat\mu_{n,i})$. Note that $\hat\mu_{n,i}$ is given by the average of block-averages in the reduced collection $\{\mathcal{B}_j : j \in I_i^0\}$ and can be computed without any resampling. The copies ${}_kT_n^{*(i)}$'s are now combined to generate the Monte-Carlo approximation to $\hat\varphi_n^{(i)}$, just in the same way the ${}_kT_n^*$'s are used for computing the original bootstrap estimate $\hat\varphi_n$.

7.4.4 The Optimal Block Length Estimator

We now return to the problem of choosing the optimal block length for block bootstrap methods using the nonparametric plug-in method. Let $\hat\varphi_n \equiv \hat\varphi_n(\ell)$ be an MBB estimator of a level-2 parameter φ_n with an MSE of the form (cf. (7.30),(7.31))

$$n^{2a} \cdot \text{MSE}(\hat\varphi_n(\ell)) = vn^{-1}\ell^r + B^2\ell^{-2} + o(n^{-1}\ell^r + \ell^{-2}) , \tag{7.54}$$

where $v \in (0,\infty)$, $B \in \mathbb{R}$, $B \neq 0$ are unknown parameters and where $r \in \mathbb{N}$, $a \in (0,\infty)$. Then, the theoretical optimal block length ℓ_n^0 is given by (cf. (7.32))

$$\ell_n^0 = \left(\frac{2B^2}{rv}\right)^{\frac{1}{r+2}} n^{\frac{1}{r+2}}(1 + o(1)) . \tag{7.55}$$

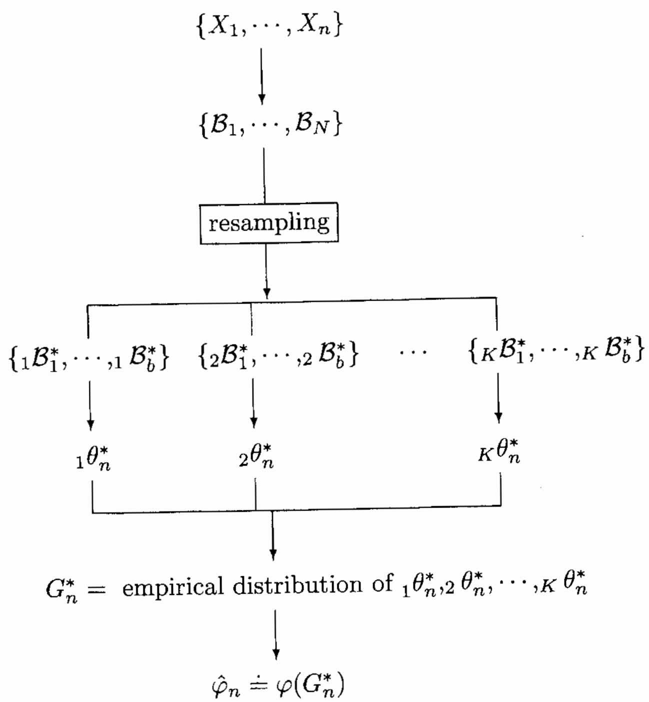

FIGURE 7.1. Monte-Carlo computation of the MBB estimator of $\varphi_n = \varphi(G_n)$ where G_n is the probability distribution of T_n of (7.51).

The nonparametric plug-in method, described in Section 7.4.1, suggests (cf. (7.39))

$$\hat{\ell}_n^0 = \left[\frac{2\hat{B}_n^2}{r\hat{v}_n}\right]^{\frac{1}{r+2}} n^{\frac{1}{r+2}} \tag{7.56}$$

as an estimator of the optional block length, where $\hat{B}_n = n^a \ell_1 \widehat{\text{BIAS}}_n$ and $\hat{v}_n = [n^{-1}\ell_1^r]^{-1} n^{2a} \widehat{\text{VAR}}_n$ are estimators of the level-3 parameters B and v, and $\widehat{\text{BIAS}}_n \equiv \widehat{\text{BIAS}}_n(\ell_1)$ and $\widehat{\text{VAR}}_n \equiv \widehat{\text{VAR}}_n(\ell_1)$ are some consistent estimators of the bias and the variance parts of the block bootstrap estimator $\hat{\varphi}_n(\ell_1)$ based on some suitable initial block length ℓ_1 (cf. (7.35), (7.36)). Lahiri, Furukawa and Lee (2003) suggest using the bias estimator $\widehat{\text{BIAS}}_n$ of (7.42) to define $\hat{B}_n$ and using the JAB variance estimator $\widehat{\text{VAR}}_{\text{JAB}}(\hat{\varphi}_n(\ell_1))$ for defining $\hat{v}_n$. With these choices, the plug-in estimator of the optimal block length ℓ_n^0 is given by

$$\tilde{\ell}_n^0 = \left[\frac{2\tilde{B}_n^2}{r\tilde{v}_n}\right]^{\frac{1}{r+2}} n^{\frac{1}{r+2}} , \tag{7.57}$$

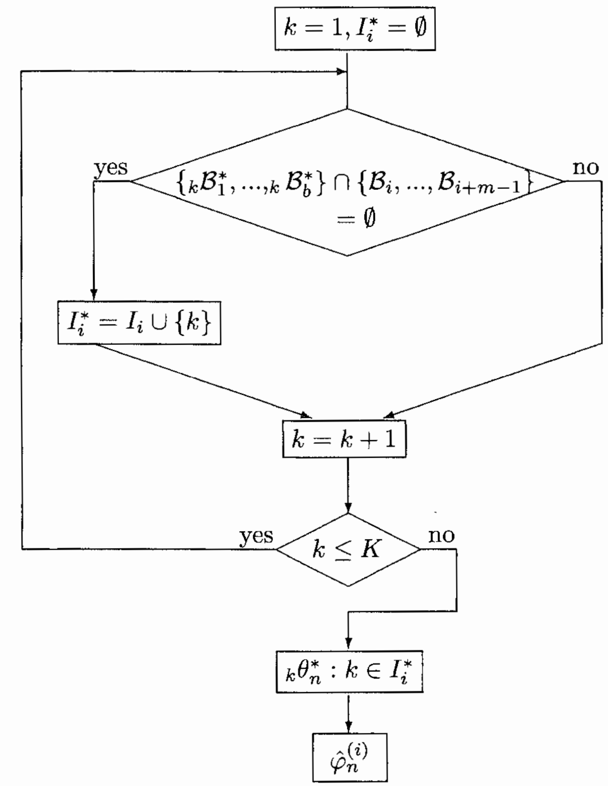

FIGURE 7.2. Computation of the ith JAB point value $\hat{\varphi}_n^{(i)}$ starting with the resampled blocks $\{{}_1\mathcal{B}_1^*, \ldots, {}_1\mathcal{B}_b^*\}, \ldots, \{{}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_b^*\}$ generated for the Monte-Carlo computation of the block bootstrap estimator $\hat{\varphi}_n$ of Figure 7.1.

where $\tilde{B}_n = 2\ell_1[\hat{\varphi}_n(\ell_1) - \hat{\varphi}_n(2\ell_1)]$ and $\tilde{v}_n = (n\ell_1^{-r}) \cdot \widehat{\text{VAR}}_{\text{JAB}}(\hat{\varphi}_n(\ell_1))$, and $\widehat{\text{VAR}}_{\text{JAB}}(\hat{\varphi}_n(\ell_1))$ is defined by (7.53) with $\ell = \ell_1$. The n^a and n^{2a} factors in the definitions of $\hat{B}_n$ and $\hat{v}_n$ are left out as they cancel from the numerator and the denominator of (7.56).

We now show that this naive construction yields consistent estimators of ℓ_n^0 for various functionals φ_n without explicit form of the constants B and v in (7.54). For this, we suppose that $\{X_i\}_{i\in\mathbb{Z}}$ is a sequence of stationary random vectors with values in $\mathbb{R}^d$ and the level-1 parameter θ and its estimator $\hat{\theta}_n$ satisfy the requirements of the Smooth Function Model (7.8), i.e., $\hat{\theta}_n = H(\bar{X}_n)$ and $\theta = H(\mu)$ for some smooth function $H : \mathbb{R}^d \to \mathbb{R}$, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\mu = EX_1$. First, we consider the bias and the variance functionals (cf. (7.2),(7.3))

$$\varphi_{1n} \equiv \text{Bias}(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta \tag{7.58}$$

$$\varphi_{2n} \quad = \quad \text{Var}(\hat{\theta}_n) \ . \tag{7.59}$$

For $k = 1, 2$, let ℓ_{kn}^0 denote the optional block length for estimating φ_{kn}, defined by (7.6). Then, we have the following result.

Theorem 7.3 *Suppose that Condition (5.D$_r$) of Section 5.4 holds with* $r = 4$,

$$\ell_1^{-1} + n^{-1/3}\ell_1 + m^{-1}\ell_1 = o(1) \ , \tag{7.60}$$

and

$$n^{-1}\ell_1^{-2}m^3 = O(1) \ . \tag{7.61}$$

Also, suppose that Condition (5.M$_r$) of Section 5.4 holds with $r = 2+k(2+a_0)$ *where* a_0 *is as in the statement of Condition (5.D$_r$). Then*

$$\tilde{\ell}_{kn}^0 / \ell_{kn}^0 \longrightarrow_p 1 \quad as \quad n \to \infty \ , \tag{7.62}$$

for $k = 1, 2$.

Proof: See Lahiri, Furukawa and Lee (2003). □

Under suitable regularity conditions, Lahiri, Furukawa and Lee (2003) also prove consistency of the plug-in estimator $\tilde{\ell}_n^0$ for bootstrap distribution function estimation and for bootstrap quantile estimation for certain studentized versions of $\hat{\theta}_n$. For these functionals, expansion (7.54) for the corresponding MSEs hold with $r = 2$ and $a = 1/2$, as in the case of the distribution function $\varphi_{3n}(x_0)$ of the normalized version of $\hat{\theta}_n$ for $|x_0| \neq 1$. Thus, the optimal block sizes for these functionals in the studentized case are of the order $n^{1/4}$ and the corresponding plug-in estimators $\tilde{\ell}_n^0$ are defined with $r = 2$ in such cases. For the *normalized* version of $\hat{\theta}_n$, consistency of $\tilde{\ell}_n^0$ for the distribution function estimator φ_{3n} of (7.4) also holds (cf. Lahiri (1996d)), provided we set $r = 2$ for $|x_0| \neq 1$, and $r = 1$ for $|x_0| = 1$. Thus, the plug-in estimator provides a consistent and computationally efficacious method for estimating the optimal block length for a variety of level-2 parameters.

Although the nonparametric plug-in method produces a valid (i.e., consistent) estimator of the optimal block length, finite sample performance of the estimator depends on the choice of the smoothing parameter ℓ_1, and on the JAB "blocks of blocks" deletion parameter m. It turns out that a reasonable choice of ℓ_1 in (7.57) depends on the functional φ_n. A careful analysis of the MSE of $\tilde{B}_n$ shows that the optimal choice of ℓ_1 is of the form

$$\ell_1 = C_3 n^{\frac{1}{r+4}} \ , \tag{7.63}$$

where r is as in (7.54), and C_3 is a population parameter. As for the other smoothing parameter, an heuristic argument in Lahiri (2002a) suggests that a reasonable choice of the JAB parameter m is given by

$$m = C_4 n^{1/3} \ell_1^{2/3} \tag{7.64}$$

for some constant C_4. Numerical results of Lahiri, Furukawa and Lee (2003) show that the choice $C_3 = 1$ in (7.63) for the initial block size ℓ_1 yields good results for both the variance and the distribution function estimation problems, while the corresponding values for C_4 in (7.64) are given by $C_4 = 1.0$ for the variance functional and $C_4 = 0.1$ for the distribution function. Below we report the results from a small simulation study with the above choices of C_3 and C_4. For more simulation results, see Lahiri, Furukawa and Lee (2003).

We consider the moving average model of Section 7.3, given by (cf. (7.27)) $X_i = (\epsilon_i + \epsilon_{i-1})/\sqrt{2}$, $i \in \mathbb{Z}$, where $\{\epsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of iid random variables having the centered Chi-squared distribution with one degree of freedom. As in Section 7.3, we also set the level-1 parameter to be $\theta = EX_1$, the estimator $\hat{\theta}_n$ to be the sample mean $\bar{X}_n$, and the level-2 parameters as $\varphi_{2n} = n.\text{Var}(\bar{X}_n)$ and $\varphi_{3n} = P(\sqrt{n}(\hat{\theta}_n - \theta)/\tau_n \leq 0)$. The true value of θ is zero. Also, we take the sample size n to be 125. As stated in Section 7.3, the true values of φ_{2n} and φ_{3n} are $\varphi_{2n} = 3.984$ and $\varphi_{3n} = 0.5226$. Furthermore, the theoretical optimal block sizes for estimating φ_{2n} and φ_{3n} by the MBB are $\ell_{2n}^0 = 3$ and $\ell_{3n}^0 = 2$, as shown in Table 7.1.

Next we applied the nonparametric plug-in method to estimate the target values ℓ_{2n}^0 and ℓ_{3n}^0. Table 7.3 gives the frequency distribution of the estimated optimal block sizes based on 500 simulation runs. The block boostrap estimators in each case were evaluated using 1000 Monte-Carlo replicates. Table 7.3 shows that more than 80% of the mass of the estimated block size $\hat{\ell}_{2n}^0$ for variance estimation lies in the interval [2,5] (the true value being $\ell_{2n}^0 = 3$). The method also produces very good results for distribution function estimation, with a pronounced mode at the true value $\ell_{3n}^0 = 2$, and a small support set $\{1, 2, 3\}$.

TABLE 7.3. Frequency distribution of the optimal block sizes selected by the nonparametric plug-in method for model (7.27) with $n = 125$.

(a) Variance Estimation

$\hat{\ell}_{2n}^0$	1	2	3	4	5	6	7	8	9	10
Frequency	50	114	125	94	71	29	10	3	2	2

(b) Distribution Function Estimation

$\hat{\ell}_{3n}^0$	1	2	3
Frequency	172	268	60