

29 The Bootstrap

The bootstrap is a resampling mechanism designed to provide information about the sampling distribution of a functional $T(X_1, X_2, \dots, X_n, F)$ where X_1, X_2, \dots, X_n are sample observations and F is the CDF from which X_1, X_2, \dots, X_n are independent observations. The bootstrap is not limited to the iid situation. It has been studied for various kinds of dependent data and complex situations. In fact, this versatile nature of the bootstrap is the principal reason for its popularity. There are numerous texts and reviews of bootstrap theory and methodology, at varied technical levels. We recommend Efron and Tibshirani(1993) and Davison and Hinkley(1997) for applications oriented broad expositions, and Hall(1992), and Shao and Tu(1995) for detailed theoretical development. Modern reviews include Hall(2003),Beran(2003),Bickel(2003),and Efron(2003). Bose and Politis(1992) is a well written nontechnical account and Lahiri(2003) is a rigorous treatment of the bootstrap for various kinds of dependent data.

Suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and $T(X_1, X_2, \dots, X_n, F)$ is a functional, e.g., $T(X_1, X_2, \dots, X_n, F) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$, where $\mu = E_F(X_1)$ and $\sigma^2 = Var_F(X_1)$. In statistical problems, we frequently need to know something about the sampling distribution of T , e.g., $P_F(T(X_1, X_2, \dots, X_n, F) \leq t)$. If we had replicated samples from the population, resulting in a series of values for the statistic T , then we could form estimates of $P_F(T \leq t)$ by counting how many of the T_i 's are $\leq t$. But statistical sampling is not done that way. We do not usually obtain replicated samples; we obtain just one set of data of some size n . However, let us think for a moment of a finite population. A large sample from a finite population should be well representative of the full population itself. So replicated samples (with replacement) from the original sample, which would just be an iid sample from the empirical CDF F_n , could be regarded as proxies for replicated samples from the population itself, provided n is large. Suppose for some number B , we draw B resamples of size n from the original sample. Denoting the resamples from the original sample as $(X_{11}^*, X_{12}^*, \dots, X_{1n}^*), (X_{21}^*, X_{22}^*, \dots, X_{2n}^*), \dots, (X_{B1}^*, X_{B2}^*, \dots, X_{Bn}^*)$, with corresponding values $T_1^*, T_2^*, \dots, T_B^*$ for the functional T , one can use simple frequency based estimates such as $\frac{\#\{j: T_j^* \leq t\}}{B}$ to estimate $P_F(T \leq t)$. This is the basic idea of the bootstrap. Over time, the bootstrap has found its use in estimating other quantities, e.g., $Var_F(T)$ or quantiles of T . The bootstrap is thus an omnibus mechanism for approximating sampling distributions or functionals of sampling distributions

of statistics. Since frequentist inference is mostly about sampling distributions of suitable statistics, the bootstrap is viewed as an immensely useful and versatile tool, further popularized by its automatic nature. However, it is also frequently misused in situations where it should not be used. In this chapter, we give a broad methodological introduction to various types of bootstrap, explain their theoretical underpinnings, discuss their successes and limitations, and try them out in some trial cases.

29.1 Bootstrap Distribution and Meaning of Consistency

The formal definition of the bootstrap distribution of a functional is the following.

Definition 29.1. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and $T(X_1, X_2, \dots, X_n, F)$ a given functional. The *ordinary bootstrap distribution* of T is defined as

$$H_{Boot}(x) = P_{F_n}(T(X_1^*, \dots, X_n^*, F_n) \leq x),$$

where (X_1^*, \dots, X_n^*) is an iid sample of size n from the empirical CDF F_n .

It is common to use the notation P_* to denote probabilities under the bootstrap distribution.

Remark: $P_{F_n}(\cdot)$ corresponds to probability statements corresponding to all the n^n possible with replacement resamples from the original sample (X_1, \dots, X_n) . Since recalculating T from all n^n resamples is basically impossible unless n is very small, one uses a smaller number of B resamples and recalculates T only B times. Thus $H_{Boot}(x)$ is itself estimated by a Monte Carlo, known as *Bootstrap Monte Carlo*. So the final estimate for $P_F(T(X_1, X_2, \dots, X_n, F_n) \leq x)$ absorbs errors from two sources: i) pretending $(X_{i1}^*, X_{i2}^*, \dots, X_{in}^*)$ to be bona fide resamples from F ; ii) estimating the true $H_{Boot}(x)$ by a Monte Carlo. By choosing B adequately large, the Monte Carlo error is generally ignored. The choice of B which would let one ignore the Monte Carlo error is a hard mathematical problem; Hall (1986,1989) are two key references. It is customary to choose $B \approx 300$ for variance estimation and a somewhat larger value for estimating quantiles. It is hard to give any general reliable prescriptions on B .

It is important to note that the resampled data need not necessarily be obtained from the empirical CDF F_n . Indeed, it is a natural question whether resampling from a smoothed nonparametric distribution estimator can result in better performance. Examples of such smoothed distribution estimators are integrated kernel density

estimates. It turns out that in some problems, smoothing does lead to greater accuracy, typically in the second order. See Silverman and Young (1987) and Hall et al. (1989) for practical questions and theoretical analysis of the benefits of using a smoothed bootstrap. Meanwhile, bootstrapping from F_n is often called the *naive or orthodox bootstrap* and we will sometimes use this terminology.

Remark: At first glance, the idea appears to be a bit too simple to actually work. But one has to have a definition for what one means by the bootstrap working in a given situation. It depends on what one wants the bootstrap to do. For estimating the CDF of a statistic, one should want $H_{Boot}(x)$ to be numerically close to the true CDF $H_n(x)$ of T . This would require consideration of metrics on CDFs. For a general metric ρ , the definition of “the bootstrap working” is the following.

Definition 29.2. Let F, G be two CDFs on a sample space \mathcal{X} . Let $\rho(F, G)$ be a metric on the space of CDFs on \mathcal{X} . For $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$, and a given functional $T(X_1, X_2, \dots, X_n, F)$, let

$$\begin{aligned} H_n(x) &= P_F(T(X_1, X_2, \dots, X_n, F) \leq x), \\ H_{Boot}(x) &= P_*(T(X_1^*, X_2^*, \dots, X_n^*, F_n) \leq x). \end{aligned}$$

We say that the bootstrap is weakly consistent under ρ for T if $\rho(H_n, H_{Boot}) \xrightarrow{P} 0$ as $n \rightarrow \infty$. We say that the bootstrap is strongly consistent under ρ for T if $\rho(H_n, H_{Boot}) \xrightarrow{a.s.} 0$.

Remark: Note that the need for mentioning convergence to zero in probability or *a.s.* in this definition is due to the fact that the bootstrap distribution H_{Boot} is a random CDF. That H_{Boot} is a random CDF has nothing to do with bootstrap Monte Carlo; it is a random CDF because as a function it depends on the original sample (X_1, X_2, \dots, X_n) . Thus, the bootstrap uses a random CDF to approximate a deterministic but unknown CDF, namely the true CDF H_n of the functional T .

Example 29.1. How does one apply the bootstrap in practice? Suppose for example, $T(X_1, \dots, X_n, F) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$. In the orthodox bootstrap scheme, we take iid samples from F_n . The mean and the variance of the empirical distribution F_n are \bar{X} and $s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (note the n rather than $n - 1$ in the denominator). The bootstrap is a device for estimating $P_F(\frac{\sqrt{n}(\bar{X} - \mu(F))}{\sigma} \leq x)$ by $P_{F_n}(\frac{\sqrt{n}(\bar{X}_n^* - \bar{X})}{s} \leq x)$. We will further approximate $P_{F_n}(\frac{\sqrt{n}(\bar{X}_n^* - \bar{X})}{s} \leq x)$ by resampling only B times from the original sample set $\{X_1, \dots, X_n\}$. In other words, finally we will report as our estimate for $P_F(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq x)$ the number $\#\{j : \frac{\sqrt{n}(\bar{X}_{n,j}^* - \bar{X})}{s} \leq x\} / B$.

29.2 Consistency in the Kolmogorov and Wasserstein Metric

We start with the case of the sample mean of iid random variables. If $X_1, \dots, X_n \stackrel{iid}{\sim} F$ and if $Var_F(X_i) < \infty$, then $\sqrt{n}(\bar{X} - \mu)$ has a limiting normal distribution by the CLT. So a probability like $P_F(\sqrt{n}(\bar{X} - \mu) \leq x)$ could be approximated by, e.g., $\Phi(\frac{x}{s})$, where s is the sample standard deviation. An interesting property of the bootstrap approximation is that even when the CLT approximation $\Phi(\frac{x}{s})$ is available, the bootstrap approximation may be more accurate. We will later describe theoretical results to this regard. But first we present two consistency results corresponding to two specific metrics that have earned a special status in this literature. The two metrics are

(i) Kolmogorov metric

$$K(F, G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|;$$

(ii) Mallows-Wasserstein metric

$$\ell_2(F, G) = \inf_{\Gamma_{2,F,G}} (E|Y - X|^2)^{\frac{1}{2}},$$

where $X \sim F, Y \sim G$ and $\Gamma_{2,F,G}$ is the class of all joint distributions of (X, Y) with marginals F and G , each with a finite second moment.

ℓ_2 is a special case of the more general metric

$$\ell_p(F, G) = \inf_{\Gamma_{p,F,G}} (E|Y - X|^p)^{\frac{1}{p}},$$

with the infimum being taken over the class of joint distributions with marginals as F, G , and the p th moment of F, G being finite.

Of these, the Kolmogorov metric is universally regarded as a natural one. But how about ℓ_2 ? ℓ_2 is a natural metric for many statistical problems because of its interesting property that $\ell_2(F_n, F) \rightarrow 0$ iff $F_n \xrightarrow{\mathcal{L}} F$ and $E_{F_n}(X^i) \rightarrow E_F(X^i)$ for $i = 1, 2$. Since one might want to use the bootstrap primarily for estimating the CDF, mean and variance of a statistic, consistency in ℓ_2 is just the right result for that purpose.

Theorem 29.1. Suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and suppose $E_F(X_1^2) < \infty$. Let $T(X_1, \dots, X_n, F) = \sqrt{n}(\bar{X} - \mu)$. Then $K(H_n, H_{Boot})$ and $\ell_2(H_n, H_{Boot}) \xrightarrow{a.s.} 0$ as

$n \rightarrow \infty$.

Remark: Strong consistency in K is proved in Singh (1981) and that for ℓ_2 is proved in Bickel and Freedman (1981). Notice that $E_F(X_1^2) < \infty$ guarantees that $\sqrt{n}(\bar{X} - \mu)$ admits a CLT. And the theorem above says that the bootstrap is strongly consistent (wrt K and ℓ_2) under that assumption. This is in fact a very good rule of thumb: if a functional $T(X_1, X_2, \dots, X_n, F)$ admits a CLT, then the bootstrap would be at least weakly consistent for T . Strong consistency might require a little more assumption.

We sketch a proof of the strong consistency in K . The proof requires use of the Berry-Esseen inequality, Polya's theorem (see Chapter 1 or Chapter 2), and a strong law known as the Zygmund-Marcinkiewicz strong law, which we state below.

Lemma (Zygmund-Marcinkiewicz SLLN) Let Y_1, Y_2, \dots be iid random variables with cdf F and suppose, for some $0 < \delta < 1$, $E_F|Y_1|^\delta < \infty$. Then $n^{-1/\delta} \sum_{i=1}^n Y_i \xrightarrow{a.s.} 0$.

We are now ready to sketch the proof of strong consistency of H_{Boot} under K .

$$\begin{aligned} & \text{Using the definition of } K, \text{ we can write } K(H_n, H_{Boot}) = \sup_x |P_F \{T_n \leq x\} - P_* \{T_n^* \leq x\}| \\ &= \sup_x \left| P_F \left\{ \frac{T_n}{\sigma} \leq \frac{x}{\sigma} \right\} - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| \\ &= \sup_x \left| P_F \left\{ \frac{T_n}{\sigma} \leq \frac{x}{\sigma} \right\} - \Phi \left(\frac{x}{\sigma} \right) + \Phi \left(\frac{x}{\sigma} \right) - \Phi \left(\frac{x}{s} \right) + \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| \\ &\leq \sup_x \left| P_F \left\{ \frac{T_n}{\sigma} \leq \frac{x}{\sigma} \right\} - \Phi \left(\frac{x}{\sigma} \right) \right| + \sup_x \left| \Phi \left(\frac{x}{\sigma} \right) - \Phi \left(\frac{x}{s} \right) \right| \\ &+ \sup_x \left| \Phi \left(\frac{x}{s} \right) - P_* \left\{ \frac{T_n^*}{s} \leq \frac{x}{s} \right\} \right| \\ &= A_n + B_n + C_n, \quad \text{say.} \end{aligned}$$

That $A_n \rightarrow 0$ is a direct consequence of Polya's Theorem. Also, s^2 converges almost surely to σ^2 and so, by the Continuous Mapping Theorem, s converges almost surely to σ . Then $B_n \Rightarrow 0$ almost surely by the fact that $\Phi(\cdot)$ is a uniformly continuous function. Finally, we can apply the Berry-Esseen Theorem to show that C_n goes to zero:

$$\begin{aligned} C_n &\leq \frac{4}{5\sqrt{n}} \cdot \frac{E_{F_n} |X_1^* - \bar{X}_n|^3}{[\text{Var}_{F_n}(X_1^*)]^{3/2}} \\ &= \frac{4}{5\sqrt{n}} \cdot \frac{\sum_{i=1}^n |X_i - \bar{X}_n|^3}{ns^3} \\ &\leq \frac{4}{5n^{3/2}s^3} \cdot 2^3 \left[\sum_{i=1}^n |X_i - \mu|^3 + n|\mu - \bar{X}_n|^3 \right] \\ &= \frac{M}{s^3} \left[\frac{1}{n^{3/2}} \sum_{i=1}^n |X_i - \mu|^3 + \frac{|\bar{X}_n - \mu|^3}{\sqrt{n}} \right], \end{aligned}$$

where $M = \frac{32}{5}$.

Since $s \Rightarrow \sigma > 0$ and $\bar{X}_n \Rightarrow \mu$, it is clear that $|\bar{X}_n - \mu|^3/(\sqrt{n}s^3) \Rightarrow 0$ almost surely. As regards the first term, let $Y_i = |X_i - \mu|^3$ and $\delta = 2/3$. Then the $\{Y_i\}$ are iid and

$$E|Y_i|^\delta = E_F|X_i - \mu|^{3 \cdot 2/3} = \text{Var}_F(X_1) < \infty.$$

It now follows from the Zygmund-Marcinkiewicz SLLN that

$$\frac{1}{n^{3/2}} \sum_{i=1}^n |X_i - \mu|^3 = n^{-1/\delta} \sum_{i=1}^n Y_i \Rightarrow 0, \text{ a.s.}, \quad \text{as } n \rightarrow \infty.$$

Thus, $A_n + B_n + C_n \Rightarrow 0$ almost surely, and hence $K(H_n, H_{Boot}) \Rightarrow 0$.

We now proceed to a proof of convergence under the Wasserstein-Kantorovich-Mallows metric ℓ_2 . Recall that convergence in ℓ_2 allows us to conclude more than weak convergence. We start with a sequence of results that enumerate useful properties of the ℓ_2 metric.

These facts (see Bickel and Doksum (1981)) are needed to prove consistency of H_{Boot} in the ℓ_2 metric.

Lemma. Let $G_n, G \in \Gamma_2$. Then $\ell_2(G_n, G) \rightarrow 0$ if and only if

$$G_n \xrightarrow{\ell_2} G \quad \text{and} \quad \lim_{n \rightarrow \infty} \int x^k dG_n(x) = \int x^k dG(x), \quad k = 1, 2.$$

Lemma. Let $G, H \in \Gamma_2$ and suppose Y_1, \dots, Y_n are iid G and Z_1, \dots, Z_n are iid H . If $G^{(n)}$ is the cdf of $\sqrt{n}(\bar{Y} - \mu_G)$ and $H^{(n)}$ is the cdf of $\sqrt{n}(\bar{Z} - \mu_H)$, then $\ell_2(G^{(n)}, H^{(n)}) \leq \ell_2(G, H)$, $\forall n \geq 1$.

Lemma. (Glivenko-Cantelli) Let X_1, X_2, \dots, X_n be iid F and let F_n be the empirical cdf. Then $F_n(x) \rightarrow F(x)$ almost surely, uniformly in x .

Lemma. Let X_1, X_2, \dots, X_n be iid F and let F_n be the empirical cdf. Then $\ell_2(F_n, F) \Rightarrow 0$ almost surely.

The proof that $\ell_2(H_n, H_{Boot})$ converges to zero almost surely follows on simply putting together the above lemmas. We omit this easy verification.

It is natural to ask if the bootstrap is consistent for $\sqrt{n}(\bar{X} - \mu)$ even when $E_F(X_1^2) = \infty$. If we insist on strong consistency, then the answer is negative. The point is that the sequence of bootstrap distributions is a sequence of random CDFs and so it cannot be a priori expected that it will converge to a fixed CDF. It may very

well converge to a random CDF, depending on the particular realization X_1, X_2, \dots . One runs into this problem if $E_F(X_1^2)$ does not exist. We state the result below.

Theorem 29.2. Suppose X_1, X_2, \dots are iid random variables. There exist $\mu_n(X_1, X_2, \dots, X_n)$, an increasing sequence c_n and a fixed CDF $G(x)$ such that

$$P_* \left(\frac{\sum_{i=1}^n (X_i^* - \mu(X_1, \dots, X_n))}{c_n} \leq x \right) \xrightarrow{a.s.} G(x),$$

if and only if $E_F(X_1^2) < \infty$, in which case $\frac{c_n}{\sqrt{n}} \rightarrow 1$.

Remark: The moral of this theorem is that the existence of a nonrandom limit itself would be a problem if $E_F(X_1^2) = \infty$. See Athreya (1987), Giné and Zinn (1989) and Hall (1990) for proofs and additional examples.

The consistency of the bootstrap for the sample mean under finite second moments is also true for the multivariate case. We record consistency under the Kolmogorov metric next; see Shao and Tu (1995) for a proof.

Theorem 29.3. Let X_1, \dots, X_n, \dots be iid F , with $Cov_F(X_1) = \Sigma$, Σ finite. Let $T(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n, F) = \sqrt{n}(\tilde{X} - \mu)$. Then $K(H_{Boot}, H_n) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

29.3 Delta Theorem for the Bootstrap

We know from the ordinary delta theorem that if T admits a CLT and $g(\cdot)$ is a smooth transformation, then $g(T)$ also admits a CLT. If we were to believe in our rule of thumb, then this would suggest that the bootstrap should be consistent for $g(T)$ if it is already consistent for T . For the case of sample mean vectors, the following result holds; again, see Shao and Tu (1995) for a proof.

Theorem 29.4. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$, and let $\Sigma_{p \times p} = Cov_F(X_1)$ be finite. Let $T(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n, F) = \sqrt{n}(\tilde{X} - \mu)$ and for some $m \geq 1$, let $g : \mathbb{R}^p \rightarrow \mathbb{R}^m$. If $\nabla g(\cdot)$ exists in a neighborhood of μ , $\nabla g(\mu) \neq 0$, and if $\nabla g(\cdot)$ is continuous at μ , then the bootstrap is strongly consistent wrt K for $\sqrt{n}(g(\tilde{X}) - g(\mu))$.

Example 29.2. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and suppose $E_F(X_1^4) < \infty$. Let $Y_i = \begin{pmatrix} X_i \\ X_i^2 \end{pmatrix}$. Then with $p = 2$, Y_1, Y_2, \dots, Y_n are iid p -dimensional vectors with $Cov(Y_1)$

finite. Note that $\tilde{Y} = \left(\frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n X_i^2} \right)$. Consider the transformation $g : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ defined as $g(u, v) = v - u^2$. Then $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = g(\tilde{Y})$. If we let $\mu = E(Y_1)$, then $g(\mu) = \sigma^2 = Var(X_1)$. Since $g(\cdot)$ satisfies the conditions of the above theorem, it follows that the bootstrap is strongly consistent wrt K for $\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)$.

29.4 Second Order Accuracy of Bootstrap

One philosophical question about the use of the bootstrap is whether the bootstrap has any advantages at all when a CLT is already available. To be specific, suppose $T(X_1, \dots, X_n, F) = \sqrt{n}(\bar{X} - \mu)$. If $\sigma^2 = Var_F(X) < \infty$, then $\sqrt{n}(\bar{X} - \mu) \xrightarrow{L} N(0, \sigma^2)$ and $K(H_{Boot}, H_n) \xrightarrow{a.s.} 0$. So two competitive approximations to $P_F(T(X_1, \dots, X_n, F) \leq x)$ are $\Phi\left(\frac{x}{\sigma}\right)$ and $P_{F_n}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x)$. It turns out that for certain types of statistics, the bootstrap approximation is (theoretically) more accurate than the approximation provided by the CLT. The CLT, because any normal distribution is symmetric, cannot capture information about the skewness in the finite sample distribution of T . The bootstrap approximation does so. So the bootstrap succeeds in correcting for skewness, just as an Edgeworth expansion would do. This is called Edgeworth correction by the bootstrap and the property is called second order accuracy of the bootstrap. It is important to remember that second order accuracy is not automatic; it holds for certain types of T but not for others. It is also important to understand that practical accuracy and theoretical higher order accuracy can be different things. The following heuristic calculation will illustrate when second order accuracy can be anticipated. The first result on higher order accuracy of the bootstrap is due to Singh(1981). In addition to the references we provided in the beginning, Lehmann (1999) gives a very readable treatment of higher order accuracy of the bootstrap.

Suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and $T(X_1, \dots, X_n, F) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$; here $\sigma^2 = Var_F(X_1) <$

∞. We know that T admits the Edgeworth expansion:

$$\begin{aligned}
P_F(T \leq x) &= \Phi(x) + \frac{p_1(x|F)}{\sqrt{n}}\varphi(x) + \frac{p_2(x|F)}{n}\varphi(x) \\
&\quad + \text{smaller order terms,} \\
P_*(T^* \leq x) &= \Phi(x) + \frac{p_1(x|F_n)}{\sqrt{n}}\varphi(x) + \frac{p_2(x|F_n)}{n}\varphi(x) \\
&\quad + \text{smaller order terms,} \\
H_n(x) - H_{Boot}(x) &= \frac{p_1(x|F) - p_1(x|F_n)}{\sqrt{n}} + \frac{p_2(x|F) - p_2(x|F_n)}{n} \\
&\quad + \text{smaller order terms.}
\end{aligned}$$

Recall now that the polynomials p_1, p_2 are given as

$$\begin{aligned}
p_1(x|F) &= \frac{\gamma}{6}(1 - x^2), \\
p_2(x|F) &= x \left[\frac{\kappa - 3}{24}(3 - x^2) - \frac{\kappa^2}{72}(x^4 - 10x^2 + 15) \right],
\end{aligned}$$

where $\gamma = \frac{E_F(X_1 - \mu)^3}{\sigma^3}$ and $\kappa = \frac{E_F(X_1 - \mu)^4}{\sigma^4}$. Since $\gamma_{F_n} - \gamma = O_p(\frac{1}{\sqrt{n}})$ and $\kappa_{F_n} - \kappa = O_p(\frac{1}{\sqrt{n}})$, just from the CLT for γ_{F_n} and κ_{F_n} under finiteness of four moments, one obtains $H_n(x) - H_{Boot}(x) = O_p(\frac{1}{n})$. If we contrast this to the CLT approximation, in general, the error in the CLT is $O(\frac{1}{\sqrt{n}})$, as is known from the Berry Esseen theorem. The $\frac{1}{\sqrt{n}}$ rate cannot be improved in general even if there are four moments. Thus, by looking at the standardized statistic $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$, we have succeeded in making the bootstrap one order more accurate than the CLT. This is called second order accuracy of the bootstrap. If one does not standardize, then

$$P_F(\sqrt{n}(\bar{X} - \mu) \leq x) = P_F\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq \frac{x}{\sigma}\right) \rightarrow \Phi\left(\frac{x}{\sigma}\right)$$

and the leading term in the bootstrap approximation in this unstandardized case would be $\Phi(\frac{x}{\sigma})$. So the bootstrap approximates the true CDF $H_n(x)$ also at the rate $\frac{1}{\sqrt{n}}$, i.e., if one does not standardize, then $H_n(x) - H_{Boot}(x) = O_p(\frac{1}{\sqrt{n}})$. We have now lost the second order accuracy. The following second rule of thumb often applies.

Rule of Thumb Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and $T(X_1, \dots, X_n, F)$ a functional. If $T(X_1, \dots, X_n, F) \stackrel{\mathcal{L}}{\rightarrow} N(0, \tau^2)$ where τ is independent of F , then second order accuracy is likely. Proving it will depend on the availability of an Edgeworth expansion for T . If τ depends on F , i.e., $\tau = \tau(F)$, then the bootstrap should be just first order accurate.

Thus, as we will now see, orthodox bootstrap is second order accurate for the standardized mean $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$, although from an inferential point of view, it is not particularly useful to have an accurate approximation to the distribution of $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$, because σ would usually be unknown, and the accurate approximation could not really be used to construct a confidence interval for μ . Still, the second order accuracy result is theoretically insightful.

We state a specific result below for the case of standardized and non-standardized sample means. Let $H_n(x) = P_F(\sqrt{n}(\bar{X} - \mu) \leq x)$, $H_{n,0}(x) = P_F(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq x)$, $H_{Boot}(x) = P_*(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x)$, $H_{Boot,0}(x) = P_{F_n}(\frac{\sqrt{n}(\bar{X}^* - \bar{X})}{s} \leq x)$.

Theorem 29.5. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$.

a) If $E_F|X_1|^3 < \infty$, and F is non-lattice, then $K(H_{n,0}, H_{Boot,0}) = o_p(\frac{1}{\sqrt{n}})$;

b) If $E_F|X_1|^3 < \infty$, and F is lattice, then $\sqrt{n}K(H_{n,0}, H_{Boot,0}) \xrightarrow{P} c$, $0 < c < \infty$.

Remark: See Lahiri (2003) for a proof. The constant c in the lattice case equals $\frac{h}{\sigma\sqrt{2\pi}}$, where h is the span of the lattice $\{a + kh, k = 0, \pm 1, \pm 2, \dots\}$ on which the X_i are supported. Note also that part a) says that higher order accuracy for the standardized case obtains with three moments; Hall (1988) showed that finiteness of three absolute moments is in fact necessary and sufficient for higher order accuracy of the bootstrap in the standardized case. Bose and Babu (1991) investigate the *unconditional probability* that the Kolmogorov distance between H_{Boot} and H_n exceeds a quantity of the order $o(n^{-\frac{1}{2}})$ for a variety of statistics and show that with various assumptions, this probability goes to zero at a rate faster than $O(n^{-1})$.

Example 29.3. How does the bootstrap compare with the CLT approximation in actual applications? The question can only be answered by case by case simulation. The results are mixed in the following numerical table. The X_i are iid Exp(1) in this example and $T = \sqrt{n}(\bar{X} - 1)$, with $n = 20$. For the bootstrap approximation, $B = 250$ was used.

t	$H_n(t)$	CLT approximation	$H_{Boot}(t)$
-2	0.0098	0.0228	0.0080
-1	0.1563	0.1587	0.1160
0	0.5297	0.5000	0.4840
-1	0.8431	0.8413	0.8760
-2	0.9667	0.9772	0.9700

29.5 Other Statistics

The ordinary bootstrap which resamples with replacement from the empirical CDF F_n is consistent for many other natural statistics besides the sample mean and even higher order accurate for some, but under additional conditions. We mention a few such results below; see Shao and Tu (1995) for further details on the theorems in this section.

Theorem 29.6. (Sample percentiles)

Let X_1, \dots, X_n be $\overset{iid}{\sim} F$ and let $0 < p < 1$, Let $\xi_p = F^{-1}(p)$ and suppose F has a positive derivative $f(\xi_p)$ at ξ_p . Let $T_n = T(X_1, \dots, X_n, F) = \sqrt{n}(F_n^{-1}(p) - \xi_p)$ and $T_n^* = T(X_1^*, \dots, X_n^*, F_n) = \sqrt{n}(F_n^{*-1}(p) - F_n^{-1}(p))$ where F_n^* is the empirical CDF of X_1^*, \dots, X_n^* . Let $H_n(x) = P_F(T_n \leq x)$ and $H_{Boot}(x) = P_*(T_n^* \leq x)$. Then, $K(H_{Boot}, H_n) = O(n^{-1/4} \sqrt{\log \log n})$ almost surely.

Remark: So again, we see that under certain conditions that ensure the existence of a CLT, the bootstrap is consistent.

Next we consider the class of one-sample U-statistics.

Theorem 29.7. (U-statistics)

Let $U_n = U_n(X_1, \dots, X_n)$ be a U-statistic with a kernel h of order 2. Let $\theta = E_F(U_n) = E_F[h(X_1, X_2)]$, where $X_1, X_2 \overset{iid}{\sim} F$. Assume:

- (i) $E_F(h^2(X_1, X_2)) < \infty$
- (ii) $\tau^2 = Var_F(\tilde{h}(X)) > 0$, where $\tilde{h}(x) = E_F[h(X_1, X_2)|X_2 = x]$
- (iii) $E_F|h(X_1, X_1)| < \infty$

Let $T_n = \sqrt{n}(U_n - \theta)$ and $T_n^* = \sqrt{n}(U_n^* - U_n)$, where $U_n^* = U_n(X_1^*, \dots, X_n^*)$, $H_n(x) = P_F(T_n \leq x)$ and $H_{Boot}(x) = P_*(T_n^* \leq x)$. Then $K(H_n, H_{Boot}) \xrightarrow{\text{a.S.}} 0$.

Remark: Under conditions (i) and (ii), $\sqrt{n}(U_n - \theta)$ has a limiting normal distribution. Condition (iii) is a new additional condition and actually, it cannot be relaxed. Condition (iii) is vacuous if the kernel h is bounded or a function of $|X_1 - X_2|$. Under additional moment conditions on the kernel h , there is also a higher order accuracy result; see Helmers (1991).

Previously, we observed that the bootstrap is consistent for smooth functions of a sample mean vector. That lets us handle statistics such as the sample variance.

Under some more conditions, even higher order accuracy obtains. Here is a result in that direction.

Theorem 29.8. (Higher Order Accuracy for Functions of Means)

Let $X_1, \dots, X_n \stackrel{iid}{\sim} F$ with $E_F(X_1) = \mu$ and $Cov_F(X_1) = \Sigma_{p \times p}$. Let $g : \mathbb{R}^p \rightarrow \mathbb{R}$ be such that $g(\cdot)$ is twice continuously differentiable in some neighborhood of μ and $\nabla g(\mu) \neq 0$. Assume also:

- (i) $E_F \|X_1 - \mu\|^3 < \infty$
- (ii) $\limsup_{\|t\| \rightarrow \infty} |E_F(e^{it'X_1})| < 1$.

Let $T_n = \frac{\sqrt{n}(g(\bar{X}) - g(\mu))}{\sqrt{(\nabla g(\mu))' \Sigma (\nabla g(\mu))}}$ and $T_n^* = \frac{\sqrt{n}(g(\bar{X}^*) - g(\bar{X}))}{\sqrt{(\nabla g(\bar{X}))' S (\nabla g(\bar{X}))}}$ where $S = S(X_1, \dots, X_n)$ is the sample variance-covariance matrix. Let also $H_n(x) = P_F(T_n \leq x)$ and $H_{Boot}(x) = P_*(T_n^* \leq x)$. Then $\sqrt{n}K(H_n, H_{Boot}) \xrightarrow{\text{a.s.}} 0$.

Finally, let us describe the case of the t -statistic. By our previous rule of thumb, we would expect the bootstrap to be higher order accurate simply because the t -statistic is already studentized, and has an asymptotic variance function independent of the underlying F .

Theorem 29.9. (Higher Order Accuracy for the t -statistic)

Let $X_1, \dots, X_n \stackrel{iid}{\sim} F$. Suppose F is non-lattice and that $E_F(X^6) < \infty$. Let $T_n = \frac{\sqrt{n}(\bar{X} - \mu)}{s}$ and $T_n^* = \frac{\sqrt{n}(\bar{X}^* - \bar{X})}{s^*}$, where s^* is the standard deviation of X_1^*, \dots, X_n^* . Let $H_n(x) = P_F(T_n \leq x)$ and $H_{Boot}(x) = P_*(T_n^* \leq x)$. Then $\sqrt{n}K(H_n, H_{Boot}) \xrightarrow{\text{a.s.}} 0$.

29.6 Some Numerical Examples

The bootstrap is used in practice for a variety of purposes. It is used to estimate a CDF, or a percentile, or the bias or variance of a statistic T_n . For example, if T_n is an estimate for some parameter θ , and if $E_F(T_n - \theta)$ is the bias of T_n , the bootstrap estimate $E_{F_n}(T_n^* - T_n)$ can be used to estimate the bias. Likewise, variance estimates can be formed by estimating $Var_{F_n}(T_n)$ by $Var_{F_n}(T_n^*)$. How accurate are the bootstrap based estimates in reality?

This can only be answered on the basis of case by case simulation. Some overall qualitative phenomena have emerged from these simulations. They are:

- (a) The bootstrap captures information about skewness that the CLT will miss;

(b) The bootstrap tends to underestimate the variance of a statistic T_n .

Here are a few numerical examples:

Example 29.4. Let $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Cauchy}(\mu, 1)$. Let M_n be the sample median and $T_n = \sqrt{n}(M_n - \mu)$. If n is odd, say $n = 2k + 1$, then there is an exact variance formula for M_n . Indeed

$$\text{Var}(M_n) = \frac{2n!}{(k!)^2 \pi^n} \int_0^{\pi/2} x^k (\pi - x)^k (\cot x)^2 dx$$

See David (1981). Because of this exact formula, we can easily gauge the accuracy of the bootstrap variance estimate. In this example, $n=21$ and $B=200$. For comparison, the CLT based variance estimate is also used which is

$$\widehat{\text{Var}}(M_n) = \frac{\pi^2}{4n}.$$

The exact variance, the CLT based estimate and the bootstrap estimate for the specific simulation are 0.1367, 0.1175 and 0.0517 respectively. Note the obvious underestimation of variance by the bootstrap. Of course one cannot be sure if it is the idiosyncrasy of the specific simulation.

A general useful result on consistency of the bootstrap variance estimate for medians under very mild conditions is Ghosh et al. (1984).

Example 29.5. Suppose X_1, \dots, X_n are iid $Poi(\mu)$ and let T_n be the t -statistic $T_n = \sqrt{n}(\bar{X} - \mu)/s$. In this example $n = 20$ and $B = 200$ and for the actual data, μ was chosen to be 1. Apart from the bias and the variance of T_n , in this example we also report percentile estimates for T_n . The bootstrap percentile estimates are found by calculating T_n^* for the B resamples and calculating the corresponding percentile value of the B values of T_n^* . The bias and the variance are estimated to be -0.18 and 1.614 respectively. The estimated percentiles are reported in the table.

α	Estimated 100α Percentile
0.05	-2.45
0.10	-1.73
0.25	-0.76
0.50	-0.17
0.75	0.49
0.90	1.25
0.95	1.58

On observing the $100(1 - \alpha)\%$ estimated percentiles, it is clear that there seems to be substantial skewness in the distribution of T . Whether the skewness is truly as serious can be assessed by a large scale simulation.

Example 29.6. Suppose (X_i, Y_i) , $i = 1, 2, \dots, n$ are iid $BVN(0, 0, 1, 1, \rho)$ and let r be the sample correlation coefficient. Let $T_n = \sqrt{n}(r - \rho)$. We know that $T_n \xrightarrow{\mathcal{L}} N(0, (1 - \rho^2)^2)$; see Chapter 3. Convergence to normality is very slow. There is also an exact formula for the density of r . For $n \geq 4$, the exact density is,

$$f(r|\rho) = \frac{2^{n-3}(1 - \rho^2)^{(n-1)/2}}{\pi(n-3)!} (1 - r^2)^{(n-4)/2} \sum_{k=0}^{\infty} \Gamma\left(\frac{n+k-1}{2}\right)^2 \frac{(2\rho r)^k}{k!};$$

see Tong (1990). In the table above, we give simulation averages of the estimated standard deviation of r by using the bootstrap. We used $n = 20$, and $B = 200$. The bootstrap estimate was calculated for 1,000 independent simulations; the table reports the average of the standard deviation estimates over the 1,000 simulations.

n	True ρ	True s.d. of r	CLT estimate	Bootstrap Estimate
20	0.0	0.230	0.232	0.217
	0.5	0.182	0.175	0.160
	0.9	0.053	0.046	0.046

Again, except when ρ is large the bootstrap underestimates the variance and the CLT estimate is better.

29.7 Failure of Bootstrap

In spite of the many consistency theorems in the previous sections, there are instances where the ordinary bootstrap based on with replacement sampling from F_n actually does not work. Typically, these are instances where the functional T_n fails to admit a CLT. Before seeing a few examples we list a few situations where the ordinary bootstrap fails to estimate the CDF of T_n consistently:

- (a) $T_n = \sqrt{n}(\bar{X} - \mu)$, when $Var_F(X_1) = \infty$.
- (b) $T_n = \sqrt{n}(g(\bar{X}) - g(\mu))$ and $\nabla g(\mu) = 0$.
- (c) $T_n = \sqrt{n}(g(\bar{X}) - g(\mu))$ and g is not differentiable at μ .
- (d) $T_n = \sqrt{n}(F_n^{-1}(p) - F^{-1}(p))$ and $f(F^{-1}(p)) = 0$, or, F has unequal right and left derivatives at $F^{-1}(p)$.

- (e) The underlying population F_θ is indexed by a parameter θ and the support of F_θ depends on the value of θ .
- (f) The underlying population F_θ is indexed by a parameter θ and the true value θ_0 belongs to the boundary of the parameter space Θ .

Example 29.7. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ and $\sigma^2 = Var_F(X) = 1$. Let $g(x) = |x|$ and $T_n = \sqrt{n}(g(\bar{X}) - g(\mu))$. If the true value of $\mu = 0$, then by CLT for \bar{X} and the continuous mapping theorem, $T_n \xrightarrow{\mathcal{L}} |Z|$ with $Z \sim N(0, \sigma^2)$. To show that the bootstrap does not work in this case, we first need to observe a few subsidiary facts.

- (a) For almost all sequences $\{X_1, X_2, \dots\}$, the conditional distribution of $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$ given \bar{X}_n , converges in law to $N(0, \sigma^2)$ by use of the triangular array CLT (see van der Vaart, 1998);
- (b) The joint asymptotic distribution of $(\sqrt{n}(\bar{X}_n - \mu), \sqrt{n}(\bar{X}_n^* - \bar{X}_n)) \xrightarrow{\mathcal{L}} (Z_1, Z_2)$ where Z_1, Z_2 are iid $N(0, \sigma^2)$.

In fact a more general version of part (b) is true. Suppose (X_n, Y_n) is a sequence of random vectors such that $X_n \xrightarrow{\mathcal{L}} Z \sim H$ (some Z) and $Y_n|X_n \xrightarrow{\mathcal{L}} Z$ (the same Z) almost surely. Then $(X_n, Y_n) \xrightarrow{\mathcal{L}} (Z_1, Z_2)$ where Z_1, Z_2 are iid $\sim H$.

Therefore, returning to the example, when the true $\mu = 0$,

$$\begin{aligned} T_n^* &= \sqrt{n}(|\bar{X}_n^*| - |\bar{X}_n|) \\ &= |\sqrt{n}(\bar{X}_n^* - \bar{X}_n) + \sqrt{n}\bar{X}_n| - |\sqrt{n}\bar{X}_n| \\ &\xrightarrow{\mathcal{L}} |Z_1 + Z_2| - |Z_1| \end{aligned} \tag{29.1}$$

where Z_1, Z_2 are iid $N(0, \sigma^2)$. But this is not distributed as the absolute value of $N(0, \sigma^2)$. The sequence of bootstrap CDFs is therefore not consistent when $\mu = 0$.

Example 29.8. Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} U(0, \theta)$ and let $T_n = n(\theta - X_{(n)})$, $T_n^* = n(X_{(n)} - X_{(n)}^*)$. The ordinary bootstrap will fail in this example in the sense that the conditional distribution of T_n^* given $X_{(n)}$ does not converge to the $Exp(\theta)$ a.s.. Let us assume $\theta = 1$. Then for $t \geq 0$,

$$\begin{aligned} P_{F_n}(T_n^* \leq t) &\geq P_{F_n}(T_n^* = 0) \\ &= P_{F_n}(X_{(n)}^* = X_{(n)}) \\ &= 1 - P_{F_n}(X_{(n)}^* < X_{(n)}) \\ &= 1 - \left(\frac{n-1}{n}\right)^n \\ &\xrightarrow{n \rightarrow \infty} 1 - e^{-1} \end{aligned}$$

For example, take $t = 0.0001$; then $\lim_n P_{F_n}(T_n^* \leq t) \geq 1 - e^{-1}$ while $\lim_n P_F(T_n \leq t) = 1 - e^{-0.0001} \approx 0$. So $P_{F_n}(T_n^* \leq t) \not\rightarrow P_F(T_n \leq t)$.

The phenomenon of this example can be generalized essentially to any CDF F with a compact support $[\underline{\omega}(F), \bar{\omega}(F)]$ with some conditions on F , such as existence of a smooth and positive density. This is one of earliest examples of the failure of the ordinary bootstrap. We will revisit this issue in the next section.

29.8 m out of n Bootstrap

In the particular problems presented above and several other problems where the ordinary bootstrap fails to be consistent, resampling fewer than n observations from F_n , say m observations, cures the inconsistency problem. This is called m out of n bootstrap. Typically, consistency will be regained if $m = o(n)$; in some general theorems to this regard, one requires $m^2 = o(n)$ or some similar stronger condition than $m = o(n)$. If the n out of n ordinary bootstrap is already consistent, then there can still be m out of n schemes with m going to ∞ slower than n which too are consistent, but the m out of n scheme will perform somewhat worse than the n out of n . See Bickel et al. (1997) for an overall review.

We will now present a collection of results which show that the m out of n bootstrap, written as m/n bootstrap, solves the orthodox bootstrap's inconsistency problem in a number of cases; see Shao and Tu (1995) for proofs and details on all of the theorems in this section.

Theorem 29.10. Let X_1, X_2, \dots be iid F where F is a cdf on \mathbb{R}^d , $d \geq 1$. Suppose $\mu = E_F(X_1)$ and $\Sigma = \text{Cov}_F(X_1)$ exist, and suppose Σ is positive definite. Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be such that $\nabla g(\mu) = 0$ and the Hessian matrix $\nabla^2 g(\mu)$ is not the zero matrix. Let $T_n = n(g(\bar{X}_n) - g(\mu))$ and $T_{m,n}^* = m(g(\bar{X}_{m,n}^*) - g(\bar{X}_n))$ and define $H_n(x) = P_F\{T_n \leq x\}$ and $H_{Boot,m,n}(x) = P_*\{T_{m,n}^* \leq x\}$. Here $\bar{X}_{m,n}^*$ denotes the mean of an iid sample of size $m = m(n)$ from F_n , where $m \rightarrow \infty$ with n .

(a) If $m = o(n)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathcal{P}} 0$.

(b) If $m = o\left(\frac{n}{\log \log n}\right)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.

Theorem 29.11. Let X_1, X_2, \dots be iid F where F is a cdf on \mathbb{R} . For $0 < p < 1$, let $\xi_p = F^{-1}(p)$. Suppose F has finite and positive left and right derivatives $f(\xi_p+)$, $f(\xi_p-)$ and that $f(\xi_p+) \neq f(\xi_p-)$. Let $T_n = \sqrt{n}(F_n^{-1}(p) - \xi_p)$ and $T_{m,n}^* = \sqrt{m}(F_m^{*-1}(p) - F_n^{-1}(p))$ and define $H_n(x) = P_F\{T_n \leq x\}$ and $H_{Boot,m,n}(x) = P_*\{T_{m,n}^* \leq x\}$. Here, $F_m^{*-1}(p)$ denotes the p th quantile of an iid sample of size m from F_n .

(a) If $m = o(n)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathcal{P}} 0$.

(b) If $m = o\left(\frac{n}{\log \log n}\right)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.

Theorem 29.12. Suppose F is a cdf on \mathbb{R} and let X_1, X_2, \dots be iid F . Suppose $\theta = \theta(F)$ is such that $F(\theta) = 1$ and $F(x) < 1$ for all $x < \theta$. Suppose for some $\delta > 0$, $P_F\{n^{1/\delta}(\theta - X_{(n)}) > x\} \rightarrow e^{-(x/\theta)^\delta}$, $\forall x$. Let $T_n = n^{1/\delta}(\theta - X_{(n)})$ and $T_{m,n}^* = m^{1/\delta}(X_{(n)} - X_{(m)}^*)$ and define $H_n(x) = P_F\{T_n \leq x\}$ and $H_{Boot,m,n}(x) = P_*\{T_{m,n}^* \leq x\}$.

(a) If $m = o(n)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathcal{P}} 0$.

(b) If $m = o\left(\frac{n}{\log \log n}\right)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.

Remark: Clearly an important practical question is the choice of the bootstrap resample size m . This is a difficult question to answer, and no precise prescriptions that have any sort of general optimality are possible. A rule of thumb is to take $m \approx 2\sqrt{n}$.

29.9 Bootstrap Confidence Intervals

The standard method to find a confidence interval for a parameter θ is to find a studentized statistic, sometimes called a pivot, say $T_n = \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$, such that $T_n \xrightarrow{\mathcal{L}} T$, with T having some known CDF G . An equal tailed confidence interval for θ , asymptotically correct, is constructed as

$$\hat{\theta}_n - G^{-1}(1 - \alpha/2)\hat{\sigma}_n \leq \theta \leq \hat{\theta}_n - G^{-1}(\alpha/2)\hat{\sigma}_n$$

This agenda requires the use of a standard deviation estimate $\hat{\sigma}_n$ for the standard deviation of $\hat{\theta}_n$ and the knowledge of the function $G(x)$. Furthermore, in many cases, the limiting CDF G may depend on some unknown parameters too, which will have to be estimated in turn to construct the confidence interval. The bootstrap

methodology offers an omnibus, sometimes easy to implement, and often a more accurate method of constructing confidence intervals.

Bootstrap confidence intervals and lower and upper one sided confidence limits of various types have been proposed in great generality. Although, as a matter of methodology, they can be used in an automatic manner, a theoretical evaluation of their performance requires specific structural assumptions. The theoretical evaluation involves an Edgeworth expansion for the relevant statistic and an expansion for their quantiles, called Cornish- Fisher expansions. Necessarily, we are limited to the cases where the underlying statistic admits a known Edgeworth and Cornish-Fisher expansion. The main reference is Hall (1988). See also Göetze (1989), Hall and Martin (1989), Bickel (1992), Konishi (1991), DiCiccio and Efron (1996) and Lee (1999), of which the article by DiCiccio and Efron is a survey article and Lee (1999) discusses m/n bootstrap confidence intervals. There are also confidence intervals based on more general *subsampling* methods, which work asymptotically under the mildest conditions. These intervals and their extensions to higher dimensions are discussed in Politis, Romano, and Wolf (1999).

Over time, various bootstrap confidence limits have been proposed. Generally, the evolution is from the algebraically simplest to progressively more complicated and computer-intensive formulae for the limits. Many of these limits have, however, now been incorporated into standard statistical software. We present below a selection of these different bootstrap confidence limits and bounds. Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ be a specific estimate of the underlying parameter of interest θ .

- (A) *The Bootstrap Percentile Lower Bound (BP)*. Let $G(x) = G_n(x) = P_F\{\hat{\theta}_n \leq x\}$ be the exact distribution and let $\hat{G}(x) = P_*\{\hat{\theta}_n^* \leq x\}$ be the bootstrap distribution. The lower $1 - \alpha$ bootstrap percentile confidence bound would be $\hat{G}^{-1}(\alpha)$. So the reported interval would be $[\hat{G}^{-1}(\alpha), \infty)$. This was present in Efron (1979) itself, but it is seldom used because it tends to have a significant coverage bias.
- (B) *Transformation-based Bootstrap Percentile Confidence Bound*. Suppose there is a suitable 1-1 transformation $\varphi = \varphi_n$ of $\hat{\theta}_n$ such that $P_F\{\varphi(\hat{\theta}_n) - \varphi(\theta) \leq x\} = \psi(x)$, with ψ being a known continuous, strictly increasing and symmetric CDF (e.g., the $N(0, 1)$ CDF). Then a transformation-based bootstrap percentile lower confidence bound for θ is $\varphi^{-1}(\hat{\varphi}_n + z_\alpha)$, where $\hat{\varphi}_n = \varphi(\hat{\theta}_n)$ and $z_\alpha = \psi^{-1}(\alpha)$. Transforming may enhance the quality of the confidence bound in

some problems. But, on the other hand, it is rare that one can find such a 1-1 transformation with a known ψ .

(C) *Bootstrap t (BT)*. Let $t_n = \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$, where $\hat{\sigma}_n$ is an estimate of the standard error of $\hat{\theta}_n$, and let $t_n^* = \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}_n^*}$ be its bootstrap counterpart. As usual, let $H_{Boot}(x) = P_*\{t_n^* \leq x\}$. The bootstrap t lower bound is $\hat{\theta}_n - H_{Boot}^{-1}(1 - \alpha)\hat{\sigma}_n$, and the two-sided BT confidence limits are $\hat{\theta}_n - H_{Boot}^{-1}(1 - \alpha_1)\hat{\sigma}_n$ and $\hat{\theta}_n - H_{Boot}^{-1}(\alpha_2)\hat{\sigma}_n$, where $\alpha_1 + \alpha_2 = \alpha$, the nominal confidence level.

(D) *Bias-corrected Bootstrap Percentile Bound (BC)*. The derivation of the BC bound involves quite a lot of calculation; see Efron (1981) and Shao and Tu (1995). The BC lower confidence bound is given by $\underline{\theta}_{BC} = \hat{G}^{-1}[\psi(z_\alpha + 2\psi^{-1}(\hat{G}(\hat{\theta}_n)))]$, where \hat{G} is the bootstrap distribution of $\hat{\theta}_n^*$, ψ is as above, and $z_\alpha = \psi^{-1}(\alpha)$.

(E) *Hybrid Bootstrap Confidence Bound (BH)*. Suppose for some deterministic sequence $\{c_n\}$, $c_n(\hat{\theta}_n - \theta) \sim H_n$ and let H_{Boot} be the bootstrap distribution; i.e. the distribution of $c_n(\hat{\theta}_n^* - \hat{\theta}_n)$ under F_n . We know that $P_F\{c_n(\hat{\theta}_n - \theta) \leq H_n^{-1}(1 - \alpha)\} = 1 - \alpha$.

If we knew H_n , then we could turn this into a $100(1 - \alpha)\%$ lower confidence bound, $\theta \geq \hat{\theta}_n - \frac{1}{c_n}H_n^{-1}(1 - \alpha)$. But H_n is, in general, not known; so we approximate it by H_{Boot} . That is, the hybrid bootstrap lower confidence bound is defined as $\underline{\theta}_{BH} = \hat{\theta}_n - \frac{1}{c_n}H_{Boot}^{-1}(1 - \alpha)$

(F) *Accelerated Bias-corrected Bootstrap Percentile Bound (BC_a)*. The ordinary bias-corrected bootstrap bound is based on the assumption that we can find $z_0 = z_0(F, n)$ and ψ (for known ψ), such that

$$P_F\{\hat{\varphi}_n - \varphi + z_0 \leq x\} = \psi(x).$$

The accelerated bias-corrected bound comes from the modified assumption that there exists a constant $a = a(F, n)$ such that $P_F\left\{\frac{\hat{\varphi}_n - \varphi}{1 + a\varphi} + z_0 \leq x\right\} = \psi(x)$. In applications, it is rare that even this modification holds exactly for any given F and n . Manipulation of this probability statement results in a lower bound: $\underline{\theta}_{BC_a} = \hat{G}^{-1}\left(\psi\left(z_0 + \frac{z_\alpha + z_0}{1 - a(z_\alpha - z_0)}\right)\right)$, where $z_\alpha = \psi^{-1}(\alpha)$, a is the acceleration parameter, and \hat{G} is as before. We repeat that, of these, z_0 and a both depend on F and n . They will have to be estimated. Moreover, the cdf ψ will generally have to be replaced by an asymptotic version; e.g. an asymptotic normal cdf

of $(\hat{\varphi}_n - \varphi)/(1 + a\varphi)$. The exact manner in which z_0 and a depend on F and n is a function of the specific problem. For example, suppose that the problem to begin with is a parametric problem, $F = F_\theta$. In such a case, $z_0 = z_0(\theta, n)$ and $a = a(\theta, n)$. The exact form of $z_0(\theta, n)$ and $a(\theta, n)$ depend on F_θ , $\hat{\theta}_n$ and φ .

Remark: As regards computational simplicity, BP , BT and BH are the simplest to apply; BC and BC_a are harder to apply and, in addition, are based on assumptions that will rarely exactly hold for finite n . Furthermore, BC_a involves estimation of a very problem-specific acceleration constant a . The bootstrap t intervals are popular in practice, provided an estimate $\hat{\sigma}_n$ is readily available. The BP method usually suffers from large bias in coverage and is seldom used.

Remark: If the model is parametric, $F = F_\theta$, and $\hat{\theta}_n$ is the MLE, then one can show the following general and useful formula: $a = z_0 = \frac{1}{6} \times \text{skewness coefficient of } \dot{\ell}(\theta)$ where $\dot{\ell}(\theta)$ is the score function, $\dot{\ell}(\theta) = \frac{d}{d\theta} \log f(x_1, \dots, x_n | \theta)$. This expression allows for estimation of a and z_0 by plug-in estimates. Nonparametric estimates of a and z_0 have also been suggested; see Efron (1987) and Loh and Wu (1987).

We now state the theoretical coverage properties of the various one-sided bounds and two-sided intervals.

Definition 29.3. Let $0 < \alpha < 1$ and $I_n = I_n(X_1, \dots, X_n)$ be a confidence set for the functional $\theta(F^{(n)})$, where $F^{(n)}$ is the joint distribution of (X_1, \dots, X_n) . Then I_n is called k^{th} order accurate if $P_{F^{(n)}} \{I_n \ni \theta(F^{(n)})\} = 1 - \alpha + O(n^{-k/2})$.

The theoretical coverage properties below are derived by using Edgeworth expansions as well as Cornish-Fisher expansions for the underlying estimate $\hat{\theta}_n$. If X_1, X_2, \dots are iid F on \mathbb{R}^d , $1 \leq d < \infty$, and if $\theta = \varphi(\mu)$, $\hat{\theta} = \varphi(\bar{X})$, for a sufficiently smooth map $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, then such Edgeworth and Cornish-Fisher expansions are available. In the results below, it is assumed that θ and $\hat{\theta}$ are the images of μ and \bar{X} , respectively, under such a smooth mapping φ . See Hall (1988) for the exact details.

Theorem 29.13. The CLT, BP , BH and the BC one-sided confidence bounds are first-order accurate. The BT and BC_a one-sided bounds are second-order accurate. The CLT, BP , BH , BT and BC_a two-sided intervals are all second-order accurate.

Remark: For two-sided intervals, the higher-order accuracy result is expected because the coverage bias for the two tails cancel in the $n^{-1/2}$ term, as can be seen from the Edgeworth expansion. The striking part of the result is that the BT and BC_a can achieve higher-order accuracy, even for one-sided bounds.

Discussion. The second order accuracy of the BT lower bound is driven by an Edgeworth expansion for H_n and an analogous one for H_{Boot} . One can invert these expansions for the CDFs to get expansions for their quantiles, i.e., to obtain Cornish-Fisher expansions. Under suitable conditions on F , H_n^{-1} and H_{Boot}^{-1} admit expansions of the form:

$$H_n^{-1}(t) = z_t + \frac{q_{11}(z_t, F)}{\sqrt{n}} + \frac{q_{12}(z_t, F)}{n} + o\left(\frac{1}{n}\right)$$

and

$$H_{Boot}^{-1}(t) = z_t + \frac{q_{11}(z_t, F_n)}{\sqrt{n}} + \frac{q_{12}(z_t, F_n)}{n} + o\left(\frac{1}{n}\right) \text{ (a.s.)}$$

where $q_{11}(\cdot, F)$ and $q_{12}(\cdot, F)$ are polynomials with coefficients depending on the moments of F . The exact polynomials depend on what the statistic $\hat{\theta}_n$ is. For example, if $\hat{\theta}_n = \bar{X}$, and $\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}$, then $q_{11}(x, F) = -\frac{\gamma}{6}(1+2x^2)$, $q_{12} = x[\frac{x^2+3}{4} - \frac{\kappa(x^2-3)}{12} + \frac{5\gamma^2}{72}(4x^2-1)]$, where $\gamma = E_F \frac{(X-\mu)^3}{\sigma^3}$ and $\kappa = E_F \frac{(X-\mu)^4}{\sigma^4} - 3$. For given $t, 0 < t < 1$, On subtraction,

$$\begin{aligned} H_n^{-1}(t) - H_{Boot}^{-1}(t) &= \frac{1}{\sqrt{n}}[q_{11}(z_t, F) - q_{11}(z_t, F_n)] + \frac{1}{n}[q_{12}(z_t, F) - q_{12}(z_t, F_n)] + o\left(\frac{1}{n}\right) \text{ (a.s.)} \\ &= \frac{1}{\sqrt{n}}O_p\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{n}O_p\left(\frac{1}{\sqrt{n}}\right) + o\left(\frac{1}{n}\right) \text{ (a.s.)} = O_p\left(\frac{1}{n}\right). \end{aligned}$$

The actual confidence bounds obtained from H_n, H_{Boot} are: $\underline{\theta}_{H_n} = \hat{\theta}_n - \hat{\sigma}_n H_n^{-1}(1-\alpha)$ and $\underline{\theta}_{BT} = \hat{\theta}_n - \hat{\sigma}_n H_{Boot}^{-1}(1-\alpha)$. On subtraction,

$$|\underline{\theta}_{H_n} - \underline{\theta}_{BT}| = \hat{\sigma}_n O_p\left(\frac{1}{n}\right) \stackrel{\text{typically}}{=} O_p(n^{-\frac{3}{2}}).$$

Thus, the bootstrap-t lower bound is approximating the idealized lower bound with third order accuracy. In addition, it can be shown that $P(\theta \geq \underline{\theta}_{BT}) = 1 - \alpha + \frac{p(z_\alpha)\varphi(z_\alpha)}{n} + o\left(\frac{1}{n}\right)$, where $p(\cdot)$ is again a polynomial depending on the specific statistic and F . For the case of \bar{X} , as an example, $p(x) = \frac{x}{6}(1+2x^2)(\kappa - \frac{3}{2}\gamma^2)$. Notice the second order accuracy in this coverage statement, inspite of the fact that the confidence bound is one sided. Again, see Hall (1988) for full details.

29.10 Some Numerical Examples

How accurate are the bootstrap confidence intervals in practice? Only case by case numerical investigation can give an answer to that question. We report in a Table

$\theta(F)$	Type of CI	<u>F</u>					
		N(0,1)		t(5)		Weibull	
		coverage	length	coverage	length	coverage	length
μ	Regular t	.9	.76	.91	1.8	.75	2.8
	BP	.91	.71	.84	1.7	.73	2.6
	BT	.92	.77	.83	2.7	.83	5.5
σ^2	BP	.79	.86	.68	1.1	.65	1.3
	BT	.88	1.5	.85	3.2	.83	5.5

results of simulation averages of coverage and length in two problems. The sample size in each case is $n = 20$, in each case $B = 200$, the simulation size is 500 and the nominal coverage $1 - \alpha = .9$.

Discussion From the table, the bootstrap- t interval seems to buy more accuracy (i.e., a smaller bias in coverage) with a larger length than the BP interval. But the BP interval has such a serious bias in coverage that the bootstrap- t may be preferable. To kill the bias, modifications of the BP method have been suggested: some of these are the bias corrected BP and the accelerated bias corrected BP intervals. Extensive numerical comparisons are reported in Shao and Tu(1995).

29.11 Bootstrap Confidence Intervals for Quantiles

Another interesting problem is the estimation of quantiles of a cdf F on \mathbb{R} . We know, for example, that if X_1, X_2, \dots are iid F , if $0 < p < 1$ and if $f = F'$ exists and is strictly positive at $\xi_p = F^{-1}(p)$, then $\sqrt{n}(F_n^{-1}(p) - \xi_p) \xrightarrow{L} N(0, p(1-p)[f(\xi_p)]^{-2})$. So, a standard CLT based interval is

$$F_n^{-1}(p) \pm \frac{z_{\alpha/2}}{\sqrt{n}} \cdot \frac{\sqrt{p(1-p)}}{\widehat{f(\xi_p)}},$$

where $\widehat{f(\xi_p)}$ is some estimate of the unknown $f = F'$ at the unknown ξ_p .

For a bootstrap interval, let H_n be the cdf of $\sqrt{n}(F^{-1}(p) - \xi_p)$ and H_{Boot} its bootstrap counterpart. Using the terminology from before, a hybrid bootstrap two-sided confidence interval for ξ_p is

$$\left[F_n^{-1}(p) - H_{Boot}^{-1}(1 - \frac{\alpha}{2})/\sqrt{n}, F_n^{-1}(p) - H_{Boot}^{-1}(\frac{\alpha}{2})/\sqrt{n} \right].$$

It turns out that this interval is not only asymptotically correct, but also comes

with a surprising asymptotic accuracy. The main references are Hall and Martin (1988) and Falk and Kaufman (1991).

Theorem 29.14. Let X_1, X_2, \dots be iid F , a cdf on \mathbb{R} . For $0 < p < 1$, let $\xi_p = F^{-1}(p)$ and suppose $0 < f(\xi_p) = F'(\xi_p) < \infty$. If I_n is the two-sided hybrid bootstrap interval, then $P_F\{I_n \ni \xi_p\} = 1 - \alpha + O(n^{-1/2})$.

Remark: Actually, the best result available is stronger and says that $P_F\{I_n \ni \xi_p\} = 1 - \alpha + \frac{c(F, \alpha, p)}{\sqrt{n}} + o(n^{-1/2})$, where $c(F, \alpha, p)$ has an explicit but complicated formula. That the bias of the hybrid interval is $O(n^{-1/2})$ is *still* a surprise in view of the fact that the bootstrap distribution of $F_n^{-1}(p)$ is consistent at a very slow rate; see Singh (1981).

29.12 Bootstrap in Regression

Regression models are among the key ones that differ from the iid set up and are also among the most widely used. Bootstrap for regression cannot be model free; the particular choice of the bootstrap scheme depends on whether the errors are iid or not. We will only talk about the linear model with deterministic X and iid errors. Additional moment conditions will be necessary depending on the specific problem to which the bootstrap will be applied. The results here are available in Freedman (1981). First let us introduce some notation.

Model: $y_i = \beta'x_i + \epsilon_i$, where β is a $p \times 1$ vector and so is x_i and ϵ_i are iid with mean 0 and variance $\sigma^2 < \infty$.

X is the $n \times p$ design matrix with i -th row equal to x_i' ; $H = X(X'X)^{-1}X'$ and $h_i = H_{ii} = x_i'(X'X)^{-1}x_i$.

$\hat{\beta} = \hat{\beta}_{LS} = (X'X)^{-1}X'y$ is the least squares estimate of β , where $y = (y_1, \dots, y_n)'$ and $(X'X)^{-1}$ is assumed nonsingular.

The bootstrap scheme is defined below.

29.13 Residual Bootstrap

Let e_1, e_2, \dots, e_n denote the residuals obtained from fitting the model, i.e., $e_i = y_i - x_i'\hat{\beta}$; $\bar{e} = 0$ if $x_i = (1, x_{i1}, \dots, x_{i,p-1})'$ but not otherwise. Define $\tilde{e}_i = e_i - \bar{e}$ and let e_1^*, \dots, e_n^* be a with-replacement sample of size n from $\{\tilde{e}_1, \dots, \tilde{e}_n\}$. Let

$y_i^* = x_i' \hat{\beta} + e_i^*$ and let β^* be the LSE of β computed from (x_i, y_i^*) , $i = 1, \dots, n$. This is the bootstrapped version of $\hat{\beta}$ and the scheme is called Residual Bootstrap (RB).

Remark: The more direct approach of resampling the pairs (x_i, y_i) is known as the paired bootstrap and is necessary when the errors are not iid, for example, the case when the errors are still independent, but their variances depend on the corresponding covariate values (called the heteroscedastic case). In such a case, the residual bootstrap scheme would not work.

By simple matrix algebra, it can be shown that,

$$\begin{aligned} E_*(\beta^*) &= \hat{\beta} \\ Cov_*(\beta^*) &= \hat{\sigma}^2 (X'X)^{-1} \end{aligned}$$

where $\hat{\sigma}^2 = (1/n) \sum_{i=1}^n (e_i - \bar{e})^2$. Note that $E(\hat{\sigma}^2) < \sigma^2$. So on an average the bootstrap covariance matrix estimate will somewhat underestimate $Cov(\hat{\beta})$. However, $Cov_*(\beta^*)$ is still consistent under some mild conditions. See Shao and Tu (1995) or Freedman (1981) for the following result.

Theorem 29.15. Suppose $|X'X| \rightarrow \infty$ and $\max_{1 \leq i \leq n} h_i \rightarrow 0$ as $n \rightarrow \infty$. Then $[Cov_*(\beta^*)]^{-1} Cov(\hat{\beta}) \Rightarrow I_{p \times p}$, almost surely.

Example 29.9. The only question is when do the conditions $|X'X| \rightarrow \infty$, $\max_{1 \leq i \leq n} h_i \rightarrow 0$ hold? As an example, take the basic regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with one covariate. Then, $|X'X| = n \sum_i (x_i - \bar{x})^2$ and $h_i = (\sum_j x_j^2 - 2x_i \sum_j x_j + nx_i^2) / (n \sum_j (x_j - \bar{x})^2)$.

$$\therefore h_i \leq \frac{4n \max_j x_j^2}{n \sum_j (x_j - \bar{x})^2} = \frac{4 \max_j x_j^2}{\sum_j (x_j - \bar{x})^2}.$$

Therefore for the theorem to apply, it is enough to have $\max |x_j| / \sqrt{\sum (x_j - \bar{x})^2} \rightarrow 0$ and $n \sum (x_i - \bar{x})^2 \rightarrow \infty$.

29.14 Confidence Intervals

We present some results on bootstrap confidence intervals for a linear combination $\theta = c' \beta_1$, where $\beta' = (\beta_0, \beta_1')$, i.e., there is an intercept term in the model. Correspondingly, $x'_i = (1, t'_i)$. The confidence interval for θ or confidence bounds (lower or upper) are going to be in terms of the studentized version of the LSE of θ , namely, $\hat{\theta} = c' \hat{\beta}_1$. In fact, $\hat{\beta}_1 = S_{tt}^{-1} S_{ty}$, where $S_{tt} = \sum_i (t_i - \bar{t})(t_i - \bar{t})'$

and $S_{ty} = \sum_i (t_i - \bar{t})(y_i - \bar{y})'$. The bootstrapped version of $\hat{\theta}$ is $\theta^* = c'\beta_1^*$, where $\beta^{*'} = (\beta_0^*, \beta_1^{*'})$, as before. Since the variance of $\hat{\theta}$ is $\sigma^2 c' S_{tt}^{-1} c$, the bootstrapped version of the studentized $\hat{\theta}$ is,

$$\theta_s^* = \frac{\theta^* - \hat{\theta}}{\sqrt{\frac{1}{n} \sum_i (y_i - x_i' \beta^*)^2 c' S_{tt}^{-1} c}}$$

The bootstrap distribution is defined as $H_{Boot}(x) = P_*(\theta_s^* \leq x)$. For given α , let $H_{Boot}^{-1}(\alpha)$ be the α th quantile of H_{Boot} . We consider the bootstrap- t (BT) confidence bounds and intervals for θ . They are obtained as

$$\begin{aligned} \underline{\theta}_{BT}^{(\alpha)} &= \hat{\theta} - H_{Boot}^{-1}(1 - \alpha) \sqrt{\hat{\sigma}^2 c' S_{tt}^{-1} c}; \\ \bar{\theta}_{BT}^{(\alpha)} &= \hat{\theta} - H_{Boot}^{-1}(\alpha) \sqrt{\hat{\sigma}^2 c' S_{tt}^{-1} c}. \end{aligned}$$

and the intervals $\theta_{L,BT} = \underline{\theta}_{BT}^{(\alpha/2)}$ and $\theta_{U,BT} = \bar{\theta}_{BT}^{(\alpha/2)}$.

There are some remarkable results on the accuracy in coverage of the BT one sided bounds and confidence intervals. We state one key result below.

Theorem 29.16. (a) $P(\theta \geq \underline{\theta}_{BT}) = (1 - \alpha) + O(n^{-3/2})$

(b) $P(\theta \leq \bar{\theta}_{BT}) = (1 - \alpha) + O(n^{-3/2})$

(c) $P(\theta_{L,BT} \leq \theta \leq \theta_{U,BT}) = (1 - \alpha) + O(n^{-2})$

These results are derived in Hall (1989).

Remark: It is remarkable that one already gets third order accuracy for the one sided confidence bounds and fourth order accuracy for the two sided bounds. There seems to be no intuitive explanation for this phenomenon. It just happens that certain terms cancel in the Cornish-Fisher expansions used in the proof for the regression case.

29.15 Distribution Estimates in Regression

The residual bootstrap is also consistent for estimating the distribution of the least squares estimate $\hat{\beta}$ of the full vector β . The metric chosen is the Mallows-Wasserstein metric we used earlier for sample means of iid data. See Freedman (1981) for the result below. We first state the model and the required assumptions below.

Let $y_i = x_i' \beta + \varepsilon_i$, where x_i is the p -vector of covariates for the i^{th} sample unit. Write the design matrix as X_n . We assume that the ε_i 's are iid with mean 0 and

variance $\sigma^2 < \infty$ and that $\{X_n\}$ is a sequence of non-stochastic matrices. We assume that for every n ($n > p$), $X_n'X_n$ is positive definite. Let $h_i = x_i'(X_n'X_n)^{-1}x_i$ and let $h_{\max} = \max\{h_i\}$. We assume, for the consistency theorem below, that

(C1) **Stability:** $\frac{1}{n}X_n'X_n \rightarrow V$ where V is a $p \times p$ positive definite matrix.

(C2) **Uniform Asymptotic Negligibility:** $h_{\max} \rightarrow 0$.

Under these conditions we have the following theorem of Freedman (1981) for RB.

Theorem 29.17. Under conditions C1 and C2 above, we have the following:

- (a) $\sqrt{n}(\hat{\beta} - \beta) \stackrel{L}{\Rightarrow} N_p(0, \sigma^2 V^{-1})$;
- (b) For almost all $\{\varepsilon_i : i \geq 1\}$, $\sqrt{n}(\beta^* - \hat{\beta}) \stackrel{L}{\Rightarrow} N_p(0, \sigma^2 V^{-1})$;
- (c) $\frac{1}{\sigma}(X_n'X_n)^{1/2}(\hat{\beta} - \beta) \stackrel{L}{\Rightarrow} N_p(0, I_p)$;
- (d) For almost all $\{\varepsilon_i : i \geq 1\}$, $\frac{1}{\sigma}(X_n'X_n)^{1/2}(\beta^* - \hat{\beta}) \stackrel{L}{\Rightarrow} N_p(0, I_p)$;
- (e) If H_n and H_{Boot} are the true and bootstrap distributions of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{n}(\beta^* - \hat{\beta})$, respectively, then for almost all $\{\varepsilon_i : i \geq 1\}$, $\ell_2(H_n, H_{Boot}) \rightarrow 0$.

Remark: This theorem gives a complete picture of the consistency issue for the case of a nonstochastic design matrix and iid errors, using the residual bootstrap. If the errors are iid, but the design matrices are random, the same results hold as long as the conditions of stability and uniform asymptotic negligibility stated earlier hold with probability one. See Shao and Tu (1995) for the case of independent but not iid errors (for example, the heteroscedastic case).

29.16 Bootstrap for Dependent Data

The orthodox bootstrap does not work when the sample observations are dependent. This was already pointed out in Singh (1981). It took some time before consistent bootstrap schemes were offered for dependent data. There are consistent schemes that are meant for specific dependence structures (e.g. stationary autoregression of a known order) and there are also general bootstrap schemes that work for large classes of stationary time series without requiring any particular dependence structure. The model based schemes are better for the specific models, but can completely fall apart if some assumption about the specific model does not hold.

We start with examples of some standard short-range dependence time series models. As opposed to these, there are models that have a long memory or long-range dependence. Bootstrap runs into problems for long-memory data. See Lahiri (2006).

Standard time series models for short-range dependent processes include:

- (a) *Autoregressive Processes*. The observations y_t are assumed to satisfy

$$y_t = \mu + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t,$$

where $1 \leq p < \infty$ and the ε_t 's are iid white noise with mean 0 and variance $\sigma^2 < \infty$. The $\{y_t\}$ process is stationary if the solutions of the polynomial equation

$$1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_p z^p = 0$$

lie strictly *outside* the unit circle in the complex plane. This process is called autoregression of order p and is denoted by AR(p).

- (b) *Moving Average Processes*. Given a white noise process $\{\varepsilon_t\}$ with mean 0 and variance $\sigma^2 < \infty$, the observations are assumed to satisfy

$$y_t = \mu + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q},$$

where $1 \leq q < \infty$. The process $\{y_t\}$ is stationary if the roots of

$$1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_q z^q = 0$$

lie strictly *outside* the unit circle. This process is called a moving average process of order q , and is denoted by MA(q).

- (c) *Autoregressive Moving Average Processes*. This combines the two previously mentioned models. The observations are assumed to satisfy

$$y_t = \mu + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \dots - \varphi_q \varepsilon_{t-q}.$$

The process $\{y_t\}$ is called an autoregressive moving average process of order (p, q) , and is denoted by ARMA(p, q).

For all of these processes, the autocorrelation sequence dies off quickly; in particular, if ρ_k is the autocorrelation of lag k , then $\sum_k |\rho_k| < \infty$.

29.17 Consistent Bootstrap for Stationary Autoregression

A version of the residual bootstrap (RB) was offered in Bose (1988) and shown to be consistent and even higher-order accurate for the least squares estimate (LSE) of the vector of regression coefficients in the stationary AR(p) case. For ease of presentation, we assume $\mu = 0$ and $\sigma = 1$. In this case, the LSE of $\theta = (\theta_1, \dots, \theta_p)'$ is defined as $\hat{\theta} = \arg \min_{\theta} \sum_{t=1}^n \left[y_t - \sum_{j=1}^p \theta_j y_{t-j} \right]^2$, where $y_{1-p}, \dots, y_0, y_1, \dots, y_n$ is the observed data sequence. There is a closed form expression of $\hat{\theta}$; specifically, $\hat{\theta} = S_{nn}^{-1} (\sum_{t=1}^n y_t y_{t-1}, \sum_{t=1}^n y_t y_{t-2}, \dots, \sum_{t=1}^n y_t y_{t-p})$, where $S_{nn} = ((S_{nn}^{ij}))_{p \times p}$ and $S_{nn}^{ij} = \sum_{t=1}^n y_{t-i} y_{t-j}$. Let $\sigma_k = \text{Cov}(y_i, y_{i+k})$ and let

$$\Sigma = \begin{vmatrix} \sigma_0 & \sigma_1 & \dots & \sigma_{p-1} \\ \sigma_1 & \sigma_0 & \dots & \sigma_{p-2} \\ \vdots & & \ddots & \\ \sigma_{p-1} & \sigma_{p-2} & \dots & \sigma_0 \end{vmatrix}.$$

Assume Σ is positive definite. It is known that under this condition $\sqrt{n}\Sigma^{-1/2}(\hat{\theta} - \theta) \xrightarrow{L} N(0, I)$. So we may expect that with a suitable bootstrap scheme, $\sqrt{n}\hat{\Sigma}^{-1/2}(\theta^* - \hat{\theta})$ converges a.s. in law to $N(0, I)$. Here $\hat{\Sigma}$ denotes the sample autocovariance matrix. We now describe the bootstrap scheme given in Bose (1988).

Let $\hat{y}_t = \sum_{j=1}^p \hat{\theta}_j y_{t-j}$ and let the residuals be $e_t = y_t - \hat{y}_t$. To obtain the bootstrap data, define $\{y_{1-2p}^*, y_{2-2p}^*, \dots, y_{-p}^*\} \equiv \{y_{1-p}, y_{2-p}, \dots, y_0\}$. Obtain bootstrap residuals by taking a random sample with replacement from $\{e_t - \bar{e}\}$. Then obtain the “starred” data by using the equation $y_t^* = \sum_{j=1}^p \hat{\theta}_j y_{t-j}^* + e_t^*$. Then θ^* is the LSE obtained by using $\{y_t^*\}$. Bose (1988) proves the following result.

Theorem 29.18. Assume that ε_1 has a density with respect to Lebesgue measure and that $E(\varepsilon_1^8) < \infty$. If $H_n(x) = P\{\sqrt{n}\Sigma^{-1/2}(\hat{\theta} - \theta) \leq x\}$ and $H_{Boot}(x) = P_*\{\sqrt{n}\hat{\Sigma}^{-1/2}(\theta^* - \hat{\theta}) \leq x\}$, then $\|H_n - H_{Boot}\|_{\infty} = o(n^{-1/2})$, almost surely.

Remark: This was the first result on higher order accuracy of a suitable form of the bootstrap for dependent data. One possible criticism of the otherwise important result is that it assumes a specific dependence structure and that it assumes the order p to be known. More flexible consistent bootstrap schemes involve some form of block resampling, which we describe next.

29.18 Block Bootstrap Methods

The basic idea of the block bootstrap method is that if the underlying series is a stationary process with short-range dependence, then blocks of observations of suitable lengths should be approximately independent and the joint distribution of the variables in different blocks would be (about) the same, due to stationarity. So, if we resample blocks of observations, rather than observations one at a time, then that should bring us back to the nearly iid situation, a situation in which the bootstrap is known to succeed. Block bootstrap was first suggested in Carlstein (1986) and Künsch (1989). Various block bootstrap schemes are now available. We only present three such schemes, for which the block length is non-random. A small problem with some of the blocking schemes is that the “starred” time series is not stationary, although the original series is, by hypothesis, stationary. A version of the block bootstrap which resamples blocks of *random* length allows the “starred” series to be provably stationary. This is called the *stationary bootstrap*, proposed in Politis and Romano (1994), and Politis et al. (1999). However, later theoretical studies have established that the auxiliary randomization to determine the block lengths can make the stationary bootstrap less accurate. For this reason, we only discuss three blocking methods with non-random block lengths.

- (a) *Non-overlapping Block Bootstrap (NBB)*. In this scheme, one splits the observed series $\{y_1, \dots, y_n\}$ into non-overlapping blocks

$$B_1 = \{y_1, \dots, y_h\}, B_2 = \{y_{h+1}, \dots, y_{2h}\}, \dots, B_m = \{y_{(m-1)h+1}, \dots, y_{mh}\}$$

where it is assumed that $n = mh$. The common block length is h . One then resamples $B_1^*, B_2^*, \dots, B_m^*$ at random, with replacement, from $\{B_1, \dots, B_m\}$. Finally, the B_i^* 's are pasted together to obtain the “starred” series y_1^*, \dots, y_n^* .

- (b) *Moving Block Bootstrap (MBB)*. In this scheme, the blocks are

$$B_1 = \{y_1, \dots, y_h\}, B_2 = \{y_2, \dots, y_{h+1}\}, \dots, B_N = \{y_{n-h+1}, \dots, y_n\}$$

where $N = n - h + 1$. One then resamples B_1^*, \dots, B_m^* from B_1, \dots, B_N , where still $n = mh$.

- (c) *Circular Block Bootstrap (CBB)*. In this scheme, one periodically extends the observed series as $y_1, y_2, \dots, y_n, y_1, y_2, \dots, y_n, \dots$. Suppose we let z_i be the members of this new series, $i = 1, 2, \dots$. The blocks are defined as

$$B_1 = \{z_1, \dots, z_h\}, B_2 = \{z_{h+1}, \dots, z_{2h}\}, \dots, B_n = \{z_n, \dots, z_{n+h-1}\}.$$

One then resamples B_1^*, \dots, B_m^* from B_1, \dots, B_n .

Next we give some theoretical properties of the three block bootstrap methods described above. The results below are due to Lahiri (1999).

Suppose $\{y_i : -\infty < i < \infty\}$ is a d -dimensional stationary process with a finite mean μ and spectral density f . Let $h : \mathbb{R}^d \rightarrow \mathbb{R}^1$ be a sufficiently smooth function. Let $\theta = h(\mu)$ and $\hat{\theta}_n = h(\bar{y}_n)$, where \bar{y}_n is the mean of the realized series. We propose to use the block bootstrap schemes to estimate the bias and variance of $\hat{\theta}_n$. Precisely, let $b_n = E(\hat{\theta}_n - \theta)$ be the bias and let $\sigma_n^2 = \text{var}(\hat{\theta}_n)$ be the variance. We use the block bootstrap based estimates of b_n and σ_n^2 , denoted by \hat{b}_n and $\hat{\sigma}_n^2$, respectively.

Next, let $T_n = \hat{\theta}_n - \theta = h(\bar{y}_n) - h(\mu)$, and let $T_n^* = h(\bar{y}_n^*) - h(E_*\bar{y}_n^*)$. The estimates \hat{b}_n and $\hat{\sigma}_n^2$ are defined as $\hat{b}_n = E_*T_n^*$ and $\hat{\sigma}_n^2 = \text{Var}_*(T_n^*)$. Then the following asymptotic expansions hold; see Lahiri (1999).

Theorem 29.19. Let $h : \mathbb{R}^d \rightarrow \mathbb{R}^1$ be a sufficiently smooth function.

(a) For each of the NBB, MBB, and CBB, there exists $c_1 = c_1(f)$ such that

$$E\hat{b}_n = b_n + \frac{c_1}{nh} + o((nh)^{-1}), \quad n \rightarrow \infty.$$

(b) For the NBB, there exists $c_2 = c_2(f)$ such that

$$\text{Var}(\hat{b}_n) = \frac{2\pi^2 c_2 h}{n^3} + o(hn^{-3}), \quad n \rightarrow \infty,$$

and for the MBB and CBB,

$$\text{Var}(\hat{b}_n) = \frac{4\pi^2 c_2 h}{3n^3} + o(hn^{-3}), \quad n \rightarrow \infty.$$

(c) For each of NBB, MBB and CBB, there exists $c_3 = c_3(f)$ such that $E(\hat{\sigma}_n^2) = \sigma_n^2 + \frac{c_3}{nh} + o((nh)^{-1})$, $n \rightarrow \infty$.

(d) For NBB, there exists $c_4 = c_4(f)$ such that $\text{var}(\hat{\sigma}_n^2) = \frac{2\pi^2 c_4 h}{n^3} + o(hn^{-3})$, $n \rightarrow \infty$, and for the MBB and CBB, $\text{var}(\hat{\sigma}_n^2) = \frac{4\pi^2 c_4 h}{3n^3} + o(hn^{-3})$, $n \rightarrow \infty$.

These expansions are used in the next section.

29.19 Optimal Block Length

The asymptotic expansions for the bias and variance of the block bootstrap estimates, given in the theorem above, can be combined to produce mse-optimal block lengths. For example, for estimating b_n by \hat{b}_n , the leading term in the expansion for the mse is

$$m(h) = \frac{4\pi^2 c_2 h}{3n^3} + \frac{c_1^2}{n^2 h^2}.$$

To minimize $m(\cdot)$, we solve $m'(h) = 0$ to get

$$h_{\text{opt}} = \left(\frac{3c_1^2}{2\pi^2 c_2} \right)^{1/3} n^{1/3}.$$

Similarly, a mse-optimal block length can be derived for estimating σ_n^2 by $\hat{\sigma}_n^2$. We state the following optimal block length result of Lahiri (1999) below .

Theorem 29.20. For the MBB and the CBB, the mse-optimal block length for estimating b_n by \hat{b}_n satisfies

$$h_{\text{opt}} = \left(\frac{3c_1^2}{2\pi^2 c_2} \right)^{1/3} n^{1/3} (1 + o(1)),$$

and the mse-optimal block length for estimating σ_n^2 by $\hat{\sigma}_n^2$ satisfies

$$h_{\text{opt}} = \left(\frac{3c_3^2}{2\pi^2 c_4} \right)^{1/3} n^{1/3} (1 + o(1)).$$

Remark: Recall that the constants c_i depend on the spectral density f of the process. So, the optimal block lengths cannot be directly used. Plug-in estimates for the c_i may be substituted. Or, the formulae can be used to try block lengths proportional to $n^{1/3}$, with flexible proportionality constants. There are also other methods in the literature on selection of block lengths; see Hall et al. (1995) and Politis and White (2004).

29.20 Exercises

Exercise 29.1. For $n = 10, 20, 50$ take a random sample from a $N(0, 1)$ distribution, and bootstrap the sample mean \bar{X} using a bootstrap Monte Carlo size $B = 200$. Construct a histogram and superimpose on it the exact density of \bar{X} . Compare.

Exercise 29.2. For $n = 5, 25, 50$, take a random sample from an $Exp(1)$ density, and bootstrap the sample mean \bar{X} using a bootstrap Monte Carlo size $B = 200$. Construct a histogram and superimpose on it the exact density of \bar{X} and the CLT approximation. Compare, and discuss if the bootstrap is doing something that the CLT answer does not.

Exercise 29.3. * By using combinatorial coefficient matching cleverly, derive a formula for the number of distinct orthodox bootstrap samples with a general value of n .

Exercise 29.4. * For which, if any, of the sample mean, the sample median, and the sample variance, is it possible to explicitly obtain the bootstrap distribution $H_{Boot}(x)$?

Exercise 29.5. * For $n = 3$, write an expression for the exact Kolomogorov distance between H_n and H_{Boot} when the statistic is \bar{X} and $F = N(0, 1)$.

Exercise 29.6. For $n = 5, 25, 50$, take a random sample from an $Exp(1)$ density, and bootstrap the sample mean \bar{X} using a bootstrap Monte Carlo size $B = 200$, using both the canonical bootstrap and the natural parametric bootstrap. Construct the corresponding histograms and superimpose them on the exact density. Is the parametric bootstrap more accurate ?

Exercise 29.7. * Prove that under appropriate moment conditions, the bootstrap is consistent for the sample correlation coefficient r between two jointly distributed variables X, Y .

Exercise 29.8. * Give examples of three statistics for which the condition in the rule of thumb on second order accuracy of the bootstrap does not hold.

Exercise 29.9. * By gradually increasing the value of n , numerically approximate the constant c in the limit theorem for the Kolomogorov distance for the $Poisson(1)$ case (see the Text for the definition of c).

Exercise 29.10. * For samples from a uniform distribution, is the bootstrap consistent for the second largest order statistic? Prove your assertion.

Exercise 29.11. For $n = 5, 25, 50$, take a random sample from an $Exp(1)$ density, and compute the bootstrap-t, bootstrap-percentile, and the usual t 95% lower confidence bounds on the population mean. Use $B = 300$. Compare meaningfully.

Exercise 29.12. * Give an example of

- a) a density such that the bootstrap is not consistent for the median;
- b) a density such that the bootstrap is not consistent for the mean;
- c) a density such that the bootstrap is consistent, but not second order accurate for the mean.

Exercise 29.13. For simulated independent samples from the $U[0, \theta]$ density, let $T_n = n(\theta - X_{(n)})$. For $n = 20, 40, 60$, numerically approximate $K(H_{Boot,m,n}, H_n)$ with varying choices of m and investigate the choice of an optimal m .

Exercise 29.14. * Suppose (X_i, Y_i) are iid samples from a bivariate normal distribution. Simulate $n = 25$ observations taking $\rho = .5$, and compute

- a. The usual 95% confidence interval;
- b. The interval based on the variance stabilizing transformation (Fisher's z) (see Chapter 4);
- c. The bootstrap percentile interval;
- d. The bootstrap hybrid percentile interval;
- e. The bootstrap t interval with $\hat{\sigma}_n$ as the usual estimate;
- f. The accelerated bias-corrected bootstrap interval using φ as Fisher's z , $z_0 = \frac{r}{2\sqrt{n}}$ (the choice coming from theory), and three different values of a near zero.

Write a discussion on your findings.

Exercise 29.15. * In which of the following cases, the results in Hall (1988) are not applicable, and why?

- a. estimating the 80th percentile of a density on \mathcal{R} ;
- b. estimating the variance of a Gamma density with known scale and unknown shape parameter;

- c. estimating θ in the $U[0, \theta]$ density;
- d. estimating $P(X > 0)$ in a location parameter Cauchy density;
- e. estimating the variance of the t -statistic for Weibull data;
- f. estimating a Binomial success probability.

Exercise 29.16. Using simulated data, compute a standard CLT based 95% confidence interval and the hybrid bootstrap interval for the 90th percentile of a i) standard Cauchy distribution; ii) a Gamma distribution with scale parameter 1 and shape parameter 3. Compare and comment. Use $n = 20, 40$.

Exercise 29.17. * Are the centers of the CLT based interval and the hybrid bootstrap interval for a population quantile always the same? Sometimes the same?

Exercise 29.18. * Simulate a series of length 50 from a stationary $AR(p)$ process with $p = 2$ and then obtain the starred series by using the scheme in Bose (1988).

Exercise 29.19. * For the simulated data in the problem above, obtain the actual blocks in the NBB and the MBB scheme, with $h = 5$. Hence, generate the starred series by pasting the resampled blocks.

Exercise 29.20. For $n = 25$, take a random sample from a bivariate normal distribution with zero means, unit variances, and correlation .6. Implement the residual bootstrap using $B = 150$. Compute a bootstrap estimate of the variance of the LSE of the regression slope parameter. Comment on the accuracy of this estimate.

Exercise 29.21. For $n = 25$, take a random sample from a bivariate normal distribution with zero means, unit variances, and correlation .6. Implement the paired bootstrap using $B = 150$. Compute a bootstrap estimate of the variance of the LSE of the regression slope parameter. Compare with the preceding exercise.

Exercise 29.22. * Give an example of two design matrices that do not satisfy the conditions C1 and C2 in text.

Exercise 29.23. * Suppose the values of the covariates are $x_i = \frac{1}{i}, i = 1, 2, \dots, n$ in a simple linear regression set up. Prove or disprove that the residual bootstrap consistently estimates the distribution of the LSE of the slope parameter if the errors are i) iid $N(0, \sigma^2)$, ii) iid $t(m, 0, \sigma^2)$, where m denotes the degree of freedom.

Exercise 29.24. * Suppose \bar{X}_n is the sample mean of an iid sample from a CDF F with a finite variance, and \bar{X}_n^* is the mean of a bootstrap sample. Consistency of the bootstrap is a statement about the bootstrap distribution, conditional on the observed data. What can you say about the unconditional limit distribution of $\sqrt{n}(\bar{X}_n^* - \mu)$, where μ is the mean of F ?

29.21 References

Athreya, K. (1987). Bootstrap of the mean in the infinite variance case, *Ann. Stat.*, 15, 2, 724-731.

Beran, R. (2003). The impact of the bootstrap on statistical algorithms and theory, *Statist. Sc.*, 18, 2, 175-184.

Bickel, P. J. (1992). Theoretical comparison of different bootstrap t confidence bounds, in *Exploring the Limits of Bootstrap*, 65-76, John Wiley, New York.

Bickel, P. J. (2003). Unorthodox bootstraps, Invited paper, *J. Korean Statist. Soc.*, 32, 3, 213-224.

Bickel, P. J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap, *Ann. Stat.*, 9, 6, 1196-1217.

Bickel, P. J., Götze, F. and van Zwet, W. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses, *Stat. Sinica*, 1, 1-31.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions, *Ann. Stat.*, 16, 4, 1709 - 1722.

Bose, A. and Babu, G. (1991). Accuracy of the bootstrap approximation, *Prob. Theory and Rel. Fields*, 90, 3, 301 - 316.

Bose, A. and Politis, D. (1992). A review of the bootstrap for dependent samples, in *Stochastic Processes and Statistical Inference*, New Age, New Delhi.

Carlstein, E. (1986). The use of subsample values for estimating the variance of a general statistic from a stationary sequence, *Ann. Stat.*, 14, 3, 1171 - 1179.

David, H. A. (1981). *Order Statistics*, Wiley, New York.

Davison, A. C. and Hinkley, D. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge.

- DiCiccio, T. and Efron, B. (1996). Bootstrap confidence intervals, with discussion, *Stat. Sc.*, 11, 3, 189 - 228.
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife, *Ann. Stat.*, 7, 1, 1 - 26.
- Efron, B. (1981). Nonparametric standard errors and confidence intervals, with discussion, *Canad. J. Stat.*, 9, 2, 139 - 172.
- Efron, B. (1987). Better bootstrap confidence intervals, with comments, *JASA*, 82, 397, 171 - 200.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Efron, B. (2003). Second thoughts on the bootstrap, *Statist. Sc.*, 18, 2, 135-140.
- Falk, M. and Kaufman, E. (1991). Coverage probabilities of bootstrap confidence intervals for quantiles, *Ann. Stat.*, 19, 1, 485 - 495.
- Freedman, D. (1981). Bootstrapping regression models, *Ann. Stat.*, 9, 6, 1218-1228.
- Ghosh, M., Parr, W., Singh, K. and Babu, J. (1984). A note on bootstrapping the sample median, *Ann. Stat.*, 12, 1130-1135.
- Giné, E. and Zinn, J. (1989). Necessary conditions for bootstrap of the mean, *Ann. Stat.*, 17, 2, 684-691.
- Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval, *Ann. Stat.*, 14, 4, 1453-1462.
- Hall, P. (1988). Rate of convergence in bootstrap approximations, *Ann. Prob.*, 16, 4, 1665-1684.
- Hall, P. (1989). On efficient bootstrap simulation, *Biometrika*, 76, 3, 613 - 617.
- Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems, *Prob. Th. and Related Fields*, 81, 2, 247-273.
- Hall, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions, *Ann. Prob.*, 18, 3, 1342-1360.
- Hall, P. (1992). *Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.
- Hall, P. (2003). A Short Prehistory of the Bootstrap, *Statist. Sc.*, 18, 2, 158-167.
- Hall, P., DiCiccio, T. and Romano, J. (1989). On smoothing and the bootstrap,

Ann. Stat., 17, 2, 692 - 704.

Hall, P. and Martin, M. A. (1989). A note on the accuracy of bootstrap percentile method confidence intervals for a quantile, Stat. and Prob. Letters, 8, 3, 197 - 200.

Hall, P., Horowitz, J. and Jing, B. (1995). On blocking rules for the bootstrap with dependent data, Biometrika, 82, 3, 561 - 574.

Helmers, R. (1991). On the Edgeworth expansion and bootstrap approximation for a studentized U -statistic, Ann. Stat., 19, 1, 470-484.

Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals, Ann. Stat., 19, 4, 2209-2225.

Künsch, H. R. (1989). The Jackknife and the bootstrap for general stationary observations, Ann. Stat., 17, 3, 1217 - 1241.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods, Ann. Stat., 27, 1, 386 - 404.

Lahiri, S. N. (2003). Resampling Methods for Dependent Data, Springer-Verlag, New York.

Lahiri, S. N. (2006). Bootstrap methods, a review, in Frontiers in Statistics, J. Fan and H. Koul Eds., 231 - 256, Imperial College Press, London.

Lee, S. (1999). On a class of m out of n bootstrap confidence intervals, JRSSB, 61, 4, 901 - 911.

Lehmann, E. L. (1999). Elements of Large Sample Theory, Springer, New York.

Politis, D. and Romano, J. (1994). The stationary bootstrap, JASA, 89, 428, 1303 - 1313.

Politis, D., Romano, J. and Wolf, M. (1999). Subsampling, Springer, New York.

Politis, D. and White, A. (2004). Automatic block length selection for the dependent bootstrap, Econ. Rev., 23, 1, 53 - 70.

Silverman, B. and Young, G. (1987). The bootstrap: To smooth or not to smooth?, Biometrika, 74, 469 - 479.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, Ann. Stat., 9, 6, 1187-1195.

Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap, Springer-Verlag, New York.

Tong, Y. L. (1990). *The Multivariate Normal Distribution*, Springer, New York.
van der Vaart, A. (1998). *Asymptotic Statistics*, Cambridge.