LECTURE ON BOOTSTRAP

CHUNG-MING KUAN

Department of Finance
National Taiwan University

This version: May 11, 2009

Address for correspondence: Department of Finance, National Taiwan University, Taipei 106, Taiwan.
E-mail: ckuan@ntu.edu.tw; URL: homepage.ntu.edu.tw/~ckuan

† I am indebted to Yu-Chin Hsu for his help on early versions of this note. I also thank Hui-Ching Chuang for many useful comments on this note.

Contents

1 Introduction

To draw inferences from econometric studies, it is desirable to know the *exact* distributions of a parameter estimator and the associated test statistics. Yet, it is typically difficult to derive such distributions in practice. Researchers thus resort to approximations based on *asymptotic* distributions. For example, χ^2 distributions are used to approximate the null distributions of the standard large-sample tests, such as the Wald, LM and LR tests. A major problem with this approach is that an asymptotic distribution may provide a rather poor approximation to its finite-sample counterpart in many applications.

Efron (1979, 1982) introduce an alternative approach, namely, *bootstrap*, that offers convenient and quite accurate approximations. The bootstrap method treats the empirical distribution (or the fitted distribution) of sample data as the true distribution and obtains bootstrapped samples by re-sampling from this distribution. The exact distribution of a statistic is then approximated by the empirical distribution of the statistic, computed using the bootstrapped samples. This note is a brief introduction to the basic idea of bootstrap. Some materials in this note are taken freely from Efron (1982), Hall (1992), and Horowitz (2001).

2 Background

Let $\boldsymbol{X}_n = \{X_1, X_2, ..., X_n\}$ denote the collection of n independently and identically distributed (i.i.d.) random variables with the unknown distribution function $\boldsymbol{F}$, which is in the family of distribution functions $\mathcal{F}$. Let $R(\boldsymbol{X}_n)$ denote a statistic based on $\boldsymbol{X}_n$ and $H_n(\cdot, \boldsymbol{F})$ be its distribution function:

$$H_n(a, \boldsymbol{F}) = P_{\boldsymbol{F}}\big[R(\boldsymbol{X}_n) \leq a\big].$$

The statistic $R(\boldsymbol{X}_n)$ is said to be a *pivot* if $H_n(a, \boldsymbol{F})$ are identical for all $\boldsymbol{F} \in \mathcal{F}$.

Example 2.1 Suppose that X_i have the common distribution $\mathcal{N}(\mu, \sigma^2)$, so that $\boldsymbol{F}(a) = \Phi\big((a-\mu)/\sigma\big)$, where Φ is the distribution function of $\mathcal{N}(0, 1)$. Let $\hat{\mu}(\boldsymbol{X}_n) = \sum_{i=1}^n X_i/n$ be the estimator of μ and $\hat{\sigma}^2(\boldsymbol{X}_n) = \sum_{i=1}^n \big(X_i - \hat{\mu}(\boldsymbol{X}_n)\big)^2/(n-1)$ be the estimator of σ^2. To construct a confidence interval of μ, we utilize the following statistic:

$$R(\boldsymbol{X}_n) = \frac{\hat{\mu}(\boldsymbol{X}_n) - \mu}{\sqrt{\frac{\hat{\sigma}^2(\boldsymbol{X}_n)}{n}}}.$$

It is well known that $R(\boldsymbol{X}_n)$ is a pivot because it has the $t(n-1)$ distribution whenever $\boldsymbol{F}$ is normal. Then, given the confidence level $0 < \alpha < 1$,

$$\mathbb{P}_{\boldsymbol{F}}\left[t_{n-1,\frac{1-\alpha}{2}} < R(\boldsymbol{X}_n) < t_{n-1,\frac{1+\alpha}{2}}\right] = \alpha,$$

where $t_{n-1,s}$ is the s-th quantile of the $t(n-1)$ distribution. It is straightforward to derive the *exact* confidence interval of μ as

$$\left(\hat{\mu}(\boldsymbol{x}_n) + t_{n-1,\frac{1-\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}},\ \hat{\mu}(\boldsymbol{x}_n) + t_{n-1,\frac{1+\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}}\right),$$

where $\boldsymbol{x}_n$ is the realization of $\boldsymbol{X}_n$.

When $R(\boldsymbol{X}_n)$ is not a pivot, $H_n(\cdot, \boldsymbol{F})$ is usually unknown because $\boldsymbol{F}$ is unknown. Suppose that $R(\boldsymbol{X}_n)$ converges in distribution to $R_A(\boldsymbol{F})$, a random variable with the distribution function $H_A(\cdot, \boldsymbol{F})$. We have

$$\lim_{n\to\infty} H_n(a, \boldsymbol{F}) = H_A(a, \boldsymbol{F}), \tag{1}$$

for every a that is a continuity point of $H_A(\cdot, \boldsymbol{F})$. Given the convergence in distribution in (1), it is quite common to approximate $H_n(\cdot, \boldsymbol{F})$ by $H_A(\cdot, \boldsymbol{F})$. $R(\boldsymbol{X}_n)$ is said to be an *asymptotic pivot* if its limiting distribution function $H_A(\cdot, \boldsymbol{F})$ does not depend on $\boldsymbol{F}$.

Example 2.2 Suppose that X_i are i.i.d. with finite second moment. Without normality, the distribution of $R(\boldsymbol{X}_n)$ in Example 2.1 is not $t(n-1)$ and depends on $\boldsymbol{F}$ in general. On the other hand, it is well known that $R(\boldsymbol{X}_n)$ is an asymptotic pivot because it converges in distribution to $\mathcal{N}(0, 1)$ for all $\boldsymbol{F}$ with finite second moment. Thus, we can use the distribution function of $\mathcal{N}(0, 1)$, Φ, to approximate the distribution function of $R(\boldsymbol{X}_n)$, $H_n(\cdot, \boldsymbol{F})$. Given the confidence level $0 < \alpha < 1$,

$$\mathbb{P}_{\boldsymbol{F}}\left[q_{\frac{1-\alpha}{2}} < R(\boldsymbol{X}_n) < q_{\frac{1+\alpha}{2}}\right] \approx \alpha,$$

from which the approximated confidence interval of μ can be derived as:

$$\left(\hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1-\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}},\ \hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}}\right),$$

where q_s is the s-th percentile of the standard normal distribution.

3 Basic Idea of Bootstrap

When it is difficult to construct a pivot or an asymptotic pivot, other approximation methods are needed. The bootstrap method of Efron (1979) suggests to replace $\boldsymbol{F}$ with $\widehat{\boldsymbol{F}}_n$, the empirical distribution function based on the realization $\boldsymbol{x}_n$, and use $H_n(\cdot, \widehat{\boldsymbol{F}}_n)$ to approximate $H_n(\cdot, \boldsymbol{F})$. This results in a good approximation because $\widehat{\boldsymbol{F}}_n$ usually well approximates $\boldsymbol{F}$.

There are parametric and nonparametric methods to compute $\widehat{\boldsymbol{F}}_n$. Given the parametric form of $\boldsymbol{F}$, suppose that $\boldsymbol{F}$ is determined by the parameters $\boldsymbol{m} \in \mathbb{M} \subseteq \mathbb{R}^k$, so that $\mathcal{F} = \{\boldsymbol{F}(\cdot, \boldsymbol{m}) | \boldsymbol{m} \in \mathbb{M}\}$. Let $\widehat{\boldsymbol{m}}(\boldsymbol{x}_n)$ be an estimator of $\boldsymbol{m}$. The parametric, empirical distribution function is such that $\widehat{\boldsymbol{F}}_n(a) = \boldsymbol{F}(a, \widehat{\boldsymbol{m}}(\boldsymbol{x}_n))$. For example, when $\boldsymbol{F}$ is the normal distribution function with mean μ and variance σ^2, $\widehat{\boldsymbol{F}}_n(a) = \Phi\big((a - \hat{\mu}(\boldsymbol{x}_n))/\hat{\sigma}(\boldsymbol{x}_n)\big)$, where $\hat{\mu}(\boldsymbol{x}_n)$ and $\hat{\sigma}^2(\boldsymbol{x}_n)$ are the estimators of μ and σ^2, respectively. Write $\boldsymbol{F}^* := \widehat{\boldsymbol{F}}_n$ and let X_i^* be i.i.d. random variables with the distribution function $\boldsymbol{F}^*$. The distribution function of $R(\boldsymbol{X}_n^*)$ is then

$$H_n(a, \boldsymbol{F}^*) = H_n(a, \widehat{\boldsymbol{F}}_n) = \mathbb{P}_{\boldsymbol{F}^*}\big[R(\boldsymbol{X}_n^*) \leq a\big].$$

Unlike $\boldsymbol{F}$, $\boldsymbol{F}^*$ is known by construction. It follows that $H_n(\cdot, \boldsymbol{F}^*)$, the bootstrap distribution function of $R(\boldsymbol{X}_n)$, is also known and can be used to approximate $H_n(\cdot, \boldsymbol{F})$. The following examples illustrate the parametric method.

Example 3.1 Given $\boldsymbol{X}_n$ and $R(\boldsymbol{X}_n)$ in Example 2.1, we construct a confidence interval of μ using the bootstrap method. The parametric, empirical distribution function is $\boldsymbol{F}^*(a) = \Phi\big((a - \hat{\mu}(\boldsymbol{x}_n))/\hat{\sigma}(\boldsymbol{x}_n)\big)$. Let $\boldsymbol{X}_n^* = \{X_1^*, X_2^*, ..., X_n^*\}$ be i.i.d. random variables with the distribution $\boldsymbol{F}^*$ and

$$R(\boldsymbol{X}_n^*) = \frac{\hat{\mu}^*(\boldsymbol{X}_n^*) - \hat{\mu}(\boldsymbol{x}_n)}{\sqrt{\frac{\hat{\sigma}_*^2(\boldsymbol{X}_n^*)}{n}}},$$

where $\hat{\mu}^*(\boldsymbol{X}_n^*) = \sum_{i=1}^n X_i^*/n$ and $\hat{\sigma}_*^2(\boldsymbol{X}_n^*) = \sum_{i=1}^n \big(X_i^* - \hat{\mu}^*(\boldsymbol{X}_n^*)\big)^2/(n-1)$. The distribution of $R(\boldsymbol{X}_n)$ is $t(n-1)$, and so is the distribution of $R(\boldsymbol{X}_n^*)$ because $\boldsymbol{F}^*$ is a normal distribution function. Thus, the distribution function $H_n(\cdot, \boldsymbol{F}^*)$ agrees with $H_n(\cdot, \boldsymbol{F})$. Given the confidence level $0 < \alpha < 1$, we have

$$\mathbb{P}_{\boldsymbol{F}}\left[t_{n-1,\frac{1-\alpha}{2}} < R(\mathbf{X}_n) < t_{n-1,\frac{1+\alpha}{2}}\right] = \mathbb{P}_{\boldsymbol{F}^*}\left[t_{n-1,\frac{1-\alpha}{2}} < R(\mathbf{X}_n) < t_{n-1,\frac{1+\alpha}{2}}\right] = \alpha.$$

The confidence interval of μ based on the bootstrap method is thus exact, as in Example 2.1. This should not be surprising because $R(\boldsymbol{X}_n)$ is a pivot when X_i are i.i.d. normally distributed.

Example 3.2 Given $\boldsymbol{X}_n$ in Example 3.1, we now consider $R(\boldsymbol{X}_n) = \sqrt{n}(\hat{\mu}(\boldsymbol{X}_n) - \mu)$, whose distribution function is $H_n(a, \boldsymbol{F}) = \Phi(a/\sigma)$ and depends on σ, a parameter of $\boldsymbol{F}$. As in the previous example, $\boldsymbol{F}^*(a) = \Phi\big((a - \hat{\mu}(\boldsymbol{x}_n))/\hat{\sigma}(\boldsymbol{x}_n)\big)$. Thus, $R(\boldsymbol{X}_n^*)$ has the distribution function $H_n(\cdot, \boldsymbol{F}^*) = \Phi\big(a/\hat{\sigma}(\boldsymbol{x}_n)\big)$. Given the confidence level $0 < \alpha < 1$,

$$\mathbb{P}_{\boldsymbol{F}}\left[q_{\frac{1-\alpha}{2}}\sigma < R(\boldsymbol{X}_n) < q_{\frac{1+\alpha}{2}}\sigma\right] = \alpha,$$

which can be approximated by

$$\mathbb{P}_{\boldsymbol{F}^*}\left[q_{\frac{1-\alpha}{2}}\hat{\sigma}(\boldsymbol{x}_n) < R(\boldsymbol{X}_n) < q_{\frac{1+\alpha}{2}}\hat{\sigma}(\boldsymbol{x}_n)\right].$$

The approximated confidence interval of μ is thus

$$\left(\hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1-\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}},\ \hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}}\right).$$

If the MLE $\breve{\sigma}^2(\boldsymbol{x}_n) = \sum_{i=1}^n\big(x_i - \hat{\mu}(\boldsymbol{x}_n)\big)^2/n$ is used to estimate σ^2, $\boldsymbol{F}^*(a) = \Phi\big((a - \hat{\mu}(\boldsymbol{x}_n))/\breve{\sigma}(\boldsymbol{x}_n)\big)$. The resulting approximated confidence interval of μ is

$$\left(\hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1-\alpha}{2}}\frac{\breve{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}},\ \hat{\mu}(\boldsymbol{x}_n) + q_{\frac{1+\alpha}{2}}\frac{\breve{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}}\right).$$

This shows that the parametric bootstrap method depends not only on the choice of $R(\boldsymbol{X}_n)$ but also on the estimator of parameters.

In practice, the parametric form of $\boldsymbol{F}$ is rarely known, and a nonparametric bootstrap method that does not require prior information of $\boldsymbol{F}$ is more desirable. The nonparametric method is based on the following empirical distribution function:

$$\widehat{\boldsymbol{F}}_n(a) = \frac{1}{n}\sharp\{x_i \le a,\ i = 1, \ldots, n\}. \tag{2}$$

Let X_i^* be i.i.d. random variables with the distribution function $\boldsymbol{F}^* = \widehat{\boldsymbol{F}}_n$ given in (2). We again rely on $\boldsymbol{X}_n^*$ to construct $R(\boldsymbol{X}_n^*)$ and use the distribution function $H_n(\cdot, \boldsymbol{F}^*)$ to approximate $H_n(\cdot, \boldsymbol{F})$.

Example 3.3 Given $\boldsymbol{X}_n$ and $R(\boldsymbol{X}_n)$ in Example 2.2, we now use the nonparametric bootstrap method to construct the confidence interval of μ. For $\boldsymbol{X}_n^*$ with the distribution

$\boldsymbol{F}^*$ in (2), we calculate all possible values of $R(\boldsymbol{X}_n^*)$ over n^n different combinations of the realizations of $\boldsymbol{x}_n^* = \{x_1^*, x_2^*, ..., x_n^*\}$, and each value is assigned a probability $1/n^n$. Then,

$$H_n(a, \boldsymbol{F}^*) = \frac{1}{n^n} \sharp \big\{ R(\boldsymbol{x}_n^*) \leq a, \text{ for all } \boldsymbol{x}_n \big\}.$$

Given the confidence level $0 < \alpha < 1$, the approximated confidence interval of μ is determined by

$$\mathbb{P}_{\boldsymbol{F}^*} \left[p^*_{\frac{1-\alpha}{2}} < R(\boldsymbol{X}_n) < p^*_{\frac{1+\alpha}{2}} \right],$$

which leads to

$$\left(\hat{\mu}(\boldsymbol{x}_n) + p^*_{\frac{1-\alpha}{2}} \frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}}, \ \hat{\mu}(\boldsymbol{x}_n) + p^*_{\frac{1+\alpha}{2}} \frac{\hat{\sigma}(\boldsymbol{x}_n)}{\sqrt{n}} \right),$$

where p_s^* is the s-th quantile of $H_n(\cdot, \boldsymbol{F}^*)$.

4 Asymptotic Results

We say that $H_n(\cdot, \widehat{\boldsymbol{F}}_n)$ is consistent for $H_A(\cdot, \boldsymbol{F})$ if for every $\epsilon > 0$ and $\boldsymbol{F} \in \mathcal{F}$,

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{F}} \left[\sup_a |H_n(a, \widehat{\boldsymbol{F}}_n) - H_A(a, \boldsymbol{F})| > \epsilon \right] = 0.$$

To establish such consistency, Beran and Ducharme (1991) suggest the following conditions.

1. For every $\epsilon > 0$ and $\boldsymbol{F} \in \mathcal{F}$, $\widehat{\boldsymbol{F}}_n$ is such that $\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{F}} \big[\sup_a |\widehat{\boldsymbol{F}}_n(a) - \boldsymbol{F}(a)| > \epsilon \big] = 0$.

2. For each $\boldsymbol{F} \in \mathcal{F}$, $H_A(\cdot, \boldsymbol{F})$ is a continuous function.

3. For every a and any sequence $\{\boldsymbol{G}_n\} \in \mathcal{F}$ such that $\lim_{n\to\infty} \boldsymbol{G}_n(a) = \boldsymbol{F}(a)$, we have $\lim_{n\to\infty} H_n(a, \boldsymbol{G}_n) = H_A(a, \boldsymbol{F})$.

Note that these conditions do not require knowledge of the distribution function $\boldsymbol{F}$.

By Polya's Theorem,[1] conditions (ii) and (iii) ensure the following uniform convergence:

$$\lim_{n \to \infty} \sup_a |H_n(a, \boldsymbol{G}_n) - H_A(a, \boldsymbol{F})| = 0.$$

[1] Let F_{X_n} and F_X denote the distribution functions of X_n and X, respectively. Polya's theorem asserts that, if X_n converges in distribution to X and F_X is continuous, then $\lim_{n\to\infty} \sup_a |F_{X_n}(a) - F_X(a)| = 0$. That is, $F_{X_n}(a)$ converges to $F_X(a)$ uniformly in a.

Then for $\widehat{\boldsymbol{F}}_n$ satisfying condition (i), we have

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{F}}\left[\sup_a |H_n(a, \widehat{\boldsymbol{F}}_n) - H_A(a, \boldsymbol{F})| > \epsilon\right] = 0.$$

That is, $H_n(\cdot, \widehat{\boldsymbol{F}}_n)$ is consistent for $H_A(\cdot, \boldsymbol{F})$. When $R(\boldsymbol{X}_n)$ converges in distribution to $R_A(\boldsymbol{F})$, we have pointwise convergence of $H_n(a, \boldsymbol{F})$ to $H_A(a, \boldsymbol{F})$, as in (1). Polya's theorem ensures this convergence is uniform when $H_A(\cdot, \boldsymbol{F})$ is continuous. These results together yield

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{F}}\left[\sup_a |H_n(a, \widehat{\boldsymbol{F}}_n) - H_n(a, \boldsymbol{F})| > \epsilon\right] = 0. \tag{3}$$

That is, $H_n(\cdot, \widehat{\boldsymbol{F}}_n)$ can well approximate $H_n(\cdot, \boldsymbol{F})$ when n is sufficiently large. For more discussions of asymptotic results, see Horowitz (2001).

Remark: The Glivenko-Cantelli theorem ensures that the nonparametric empirical distribution function (2) satisfies condition (i) above and hence can serve as $\widehat{\boldsymbol{F}}_n$.[2] This justifies why the nonparametric bootstrap method works.

5 Bootstrap with Re-Sampling

We now describe how the bootstrap method can be implemented in practice. Example 3.3 shows that it is computationally formidable to calculate the bootstrap distribution. For example, when n is as small as 10, we need to compute 10^{10} (10 billion) different values for $R(\boldsymbol{X}_n^*)$. If $R(\boldsymbol{X}_n^*)$, such as t statistic, takes the same value regardless of the order of x_i^*, the required computation would be easier but still needs $\mathcal{C}_n^{2n-1}$ values of $R(\boldsymbol{X}_n^*)$. The number $\mathcal{C}_n^{2n-1}$ still grows much too fast when n becomes large.

As far as computation is concerned, re-sampling makes the bootstrap method computationally more tractable. When $\boldsymbol{X}_n$ contains n i.i.d. random variables, we can obtain another random sample by re-sampling from the realization $\boldsymbol{x}_n$. To this end, we randomly draw n observation from $\{x_1, x_2, ..., x_n\}$ with replacement and denote the b-th re-sampled observation as $\boldsymbol{x}_{n,b}^* = (x_{1,b}^*, x_{2,b}^*, ..., x_{n,b}^*)$, $b = 1, 2, ..., B$. The statistic $R(\boldsymbol{x}_{n,b}^*)$ is a realization based on $\boldsymbol{F}^*$, the empirical distribution of $\boldsymbol{x}_n$. We can then compute the empirical distribution function of $R(\boldsymbol{x}_n^*)$ based on the bootstrapped realizations as

$$\widetilde{H}_{n,B}(a, \boldsymbol{F}^*) = \frac{1}{B}\sharp\big\{R(\boldsymbol{x}_{n,b}^*) \leq a,\ b = 1, \ldots, B\big\}.$$

[2]Let $\boldsymbol{F}$ be the true distribution function and $\widehat{\boldsymbol{F}}_n$ be the nonparametric empirical distribution function. The Glivenko-Cantelli theorem asserts that $\lim_{n\to\infty} \sup_a |\boldsymbol{F}(a) - \boldsymbol{F}_n(a)| = 0$ almost surely.

Table 1: The coverage rates of the bootstrap and asymptotic methods.

	$n=10$		$n=20$		$n=50$		$n=100$	
$\boldsymbol{F}$	Boot	Asymp	Boot	Asymp	Boot	Asymp	Boot	Asymp
$e^{\mathcal{N}(0,1)}$	0.9074	0.8060	0.9192	0.8498	0.9280	0.8910	0.9346	0.9162
$t(5)$	0.9396	0.9256	0.9338	0.9296	0.9434	0.9490	0.9408	0.9454
$t(8)$	0.9430	0.9168	0.9458	0.9414	0.9470	0.9478	0.9460	0.9494
$t(11)$	0.9436	0.9194	0.9460	0.9368	0.9494	0.9478	0.9506	0.9498

Again by the Glivenko-Cantelli theorem,

$$\lim_{B\to\infty} \sup_a |\widetilde{H}_{n,B}(a,\boldsymbol{F}^*) - H_n(a,\boldsymbol{F}^*)| = 0,$$

almost surely. Thus, the bootstrap method with re-sampling works because $\widetilde{H}_{n,B}(\cdot,\boldsymbol{F}^*)$ can approximate $H_n(\cdot,\boldsymbol{F}^*)$ to any desired degree of accuracy when B is sufficiently large, and $H_n(\cdot,\boldsymbol{F}^*)$ in turn approximates $H_n(\cdot,\boldsymbol{F})$ arbitrarily well when n is large enough.

Example 5.1 Following Example 2.1, we simulate the coverage rates of 95% confidence intervals of μ for the data with the distribution $\boldsymbol{F}$. We compare the performance of the bootstrap and asymptotic methods. We consider four different distributions: log-normal ($\exp(\mathcal{N}(0,1))$), $t(5)$, $t(8)$ and $t(11)$, and four sample sizes n=10,20, 50 and 100. The number of replications is 5000, and the number of bootstrap $B = 1000$. The results are summarized in Table 1. It can be seen that the bootstrap method provides very good approximation for all cases considered and outperforms the asymptotic method when the sample size is small. In particular, for the log-normal distribution which has increasingly large moments, the bootstrap method clearly dominates the asymptotic method.

Example 5.2 This example illustrates the procedures for computing the coverage probabilities of the confidence intervals for regression coefficients. The regression data y_i are generated according to:

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1,\dots,n,$$

where x_i are i.i.d. standard normal random variables and ϵ_i are i.i.d. $t(5)$ random variables. We set $n = 30$, $\alpha = 1$ and $\beta = 1$.

The asymptotic method for approximating the coverage rate of β is described below; the coverage rate of α can be approximated similarly.

A1. Generate the data (y_i, x_i), $i = 1, \ldots, n$, as discussed above.

A2. Regress y_i on 1 and x_i to obtain the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$, the OLS residuals $\hat{e}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$, and the estimated standard deviations $\hat{\sigma}_{\hat{\alpha}}$ and $\hat{\sigma}_{\hat{\beta}}$. Specifically,

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\frac{\sum_{i=1}^{n} \hat{e}_i^2}{n\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]}}.$$

A3. Construct a two-sided 95% confidence interval for $\beta = 1$ based on the asymptotic distribution of $\hat{\beta}$:

$$CI_{AM} = \left(\hat{\beta} - q_{0.975}\, \hat{\sigma}_{\hat{\beta}},\ \hat{\beta} + q_{0.975}\, \hat{\sigma}_{\hat{\beta}}\right),$$

with $q_{0.975}$ the 0.975 quantile of the standard normal distribution.

A4. Set $AM_r^{\beta} = 1$ if $\beta = 1$ is in CI_{AM}; otherwise, $AM_r^{\beta} = 0$.

Repeat these steps R times (say, $R = 5000$) and calculate the percentage of $AM_r^{\beta} = 1$.

The nonparametric bootstrap method for approximating the coverage rate of β is described below; the coverage rate of α can be approximated similarly.

B1. Generate the data (y_i, x_i), $i = 1, \ldots, n$, as discussed above.

B2. Regress y_i on 1 and x_i and calculate the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ and their estimated standard deviations $\hat{\sigma}_{\hat{\alpha}}$ and $\hat{\sigma}_{\hat{\beta}}$ as in A2 above.

B3. Compute the bootstrapped confidence interval:

(i) Generate random indices from a uniform distribution over $\{1, .., n\}$ with replacement, denoted as $\{k_1^b, ..., k_n^b\}$.

(ii) Regress $\{y_{k_1^b}, ..., y_{k_n^b}\}$ on a constant term and $\{x_{k_1^b}, ..., x_{k_n^b}\}$ to obtain $\hat{\beta}_b^*$ and the estimated standard deviation $\hat{\sigma}_{\hat{\beta}_b^*}$. Compute the Studentized bootstrap statistic:

$$\hat{R}_b^* := \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{\sigma}_{\hat{\beta}_b^*}},$$

where $\hat{\sigma}_{\hat{\beta}_b^*}$ is computed as $\hat{\sigma}_{\hat{\beta}}$ using the b-th bootstrapped sample.[3]

[3] Alternatively, one may compute $\hat{R}_b^*$ as $(\hat{\beta}_b^* - \hat{\beta})/\hat{\sigma}_{\hat{\beta}}$.

(iii) Repeat the steps (i) and (ii) for $b = 1, ..., B$ (say, $B = 1000$) and rank the absolute value of $\hat{R}_b^*$ in an ascending order to form $\{\hat{R}_{r_1}^*, ..., \hat{R}_{r_B}^*\}$.

(iv) The symmetric, bootstrapped 95% confidence interval for β is:

$$CI_{BM,1} = \left(\hat{\beta} - p_{0.95}^* \, \hat{\sigma}_{\hat{\beta}}, \, \hat{\beta} + p_{0.95}^* \, \hat{\sigma}_{\hat{\beta}}\right),$$

where $p_{0.95}^*$ is the 0.95 quantile of $\{\hat{R}_{r_1}^*, ..., \hat{R}_{r_B}^*\}$.

B4. Set $BM_r^{\beta} = 1$ if $\beta = 1$ is in $CI_{BM,1}$; otherwise, $BM_r^{\beta} = 0$.

Repeat these steps R times and calculate the percentage of $BM_r^{\beta} = 1$.

There are different ways to construct a symmetric, bootstrapped confidence interval of β. For examples, the following confidence interval is also valid:

$$CI_{BM,2} = \left(\hat{\beta} - p_{0.95}^* \, \hat{s}_{\hat{\beta}^*}, \; \hat{\beta} + p_{0.95}^* \, \hat{s}_{\hat{\beta}^*}\right),$$

where $\hat{s}_{\hat{\beta}^*}$ is the sample standard deviation of $\{\hat{\beta}_b^*\}_{b=1}^B$, i.e.,

$$\hat{s}_{\hat{\beta}^*}^2 = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\beta}_b^* - \overline{\hat{\beta}^*}\right)^2, \quad \overline{\hat{\beta}^*} = \sum_{b=1}^{B} \hat{\beta}_b^* / B.$$

One may also compute the confidence interval as

$$CI_{BM,3} = \left(\hat{\beta} - \tilde{p}_{0.95}^*, \; \hat{\beta} + \tilde{p}_{0.95}^*\right),$$

where $\tilde{p}_{0.95}^*$ is the 0.95 quantile of the non-Studentized statistics $\hat{\beta}_b^* - \hat{\beta}$.

6 Stationary Bootstrap

The bootstrap method described in Section 5 is not valid for serially dependent data, for the independent draws in re-sampling destroys the dependence structure of the data. In this section, we introduce the stationary bootstrap of Politis and Romano (1994) that is applicable to stationary and weakly dependent data. In this method, observations are re-sampled in blocks of random size, where the size of each block is determined by a geometric distribution, so as to preserve the dependence in the original series.

Given the data $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ and the real number $0 < Q < 1$, the re-sampling scheme of stationary bootstrap can be implemented as follows.

S1. Randomly select an observation, say x_t, from the data $\boldsymbol{x}_n$ as the first bootstrapped observation $x^*_{1,b}$.

S2. with probability Q, $x^*_{2,b}$ is set to x_{t+1}, the observation following the previously sampled observation,[4] and with probability $1-Q$, the second bootstrapped observation $x^*_{2,b}$ is randomly selected from the original data $\boldsymbol{x}_n$.

S3. Repeat the second step to form $\boldsymbol{x}^*_{n,b}$, the b-th bootstrapped sample with n observations.

In this scheme, the block size is the (random) number of bootstrapped observations that are drawn consecutively. It can be shown that the block size indeed follows a geometric distribution with parameter Q, so that the expected block size is $1/(1-Q)$. Clearly, bootstrapped results depend on the choice of Q. When Q approaches zero, the resulting stationary bootstrap is like i.i.d. bootstrap.

Let $\bar{x}_n$ be the sample average of $\boldsymbol{x}_n$ and $\bar{x}^*_{n,b}$ the average of the b-th bootstrapped sample $\boldsymbol{x}^*_{n,b}$. To discuss the asymptotic property of stationary bootstrap, we shall denote the probability Q as $Q(n)$ to signify its dependence on the sample size n. The following result, due to Goncalves and de Jong (2003), holds under mild moment and dependence conditions on the data.[5]

Theorem 6.1 *Given strictly stationary and weakly dependent process $\{X_t\}$, suppose that $Q(n) \to 1$ and $n(1-Q(n))^2 \to \infty$. Then for any $\epsilon > 0$,*

$$\mathbb{P}\Big[\sup_{a\in\mathbb{R}} \big|\mathbb{P}^*[\sqrt{n}(\bar{X}^*_n - \bar{X}_n) \le a] - \mathbb{P}[\sqrt{n}(\bar{X}_n - \mu) \le a]\,\big| > \epsilon\Big] \to 0,$$

where $\mu = \mathbb{E}(X_t)$ and $\mathbb{P}^*$ is the probability measure generated by stationary bootstrap.

Remark: Theorem 6.1 holds for more general statistics $R(\boldsymbol{X}_n)$. Intuitively, the larger the expected block size (the larger the Q), the better can such re-sampling preserve the dependence structure in the data. Yet, when the expected block size is too big, the bootstrapped samples would have smaller variation and hence result in poor approximation. In Theorem 6.1, a proper expected block size is controlled by the conditions $Q(n) \to 1$ and $n(1-Q(n))^2 \to \infty$.

[4] The first observation x_1 is treated as the observation following the last observation x_n in re-sampling.

[5] Specifically, Goncalves and de Jong (2003) assume that the data $\{X_t\}$ is strictly stationary and α-mixing of size $-2(2+\delta)(r+\delta)/(r-2)$, for some $r > 2$ and $\delta > 0$, where $\mathbb{E}|X_t|^{r+\delta} < \infty$.

Example 6.1 As in Example 5.1, we examine the performance of i.i.d. bootstrap and stationary bootstrap by simulating the coverage rates of 95% confidence intervals of the mean. Suppose X_t follows an $AR(1)$ process such that

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., \tag{4}$$

where $|\rho| < 1$ and ε_t are i.i.d. standard normal. When $|\rho| \neq 0$, X_t are stationary and weakly dependent process, and when $\rho = 0$, $X_t = \varepsilon_t$ are i.i.d. standard normal.

In our simulations, we set $n = 200$ and generate $n + 100$ X_t's according to (4) with $X_0 = 0$. The last n generated observations are retained as our sample; ignoring the first 100 data points can reduce the effect of initial value on the generated sample. We consider $\rho = 0$, 0.3, 0.6, and 0.9 and $Q = 0$, 0.5, 0.7, 0.9, and 0.95; note that the stationary bootstrap simplifies to i.i.d. bootstrap when $Q = 0$. The number of replications is 5000, and the number of bootstrap $B = 1000$. The results are summarized in Table 2.

It is not surprising to see that stationary bootstrap performs worse than i.i.d. bootstrap when there is no dependence in the data. Yet, the difference is minor if Q is not too big. When the data are serially correlated, it can be seen that, for any Q, the coverage rate of the stationary bootstrapped confidence interval is higher than those based on i.i.d. bootstrap ($Q = 0$); the difference could be very substantial when correlation is strong. This illustrates the importance of a bootstrap method that can take into account data dependence. Note also that the bootstrapped coverage rates are all smaller than the nominal coverage rate (0.95) when $\rho \neq 0$. For a given Q, the difference between bootstrapped and nominal coverage rates are larger when ρ is larger. A reason for this observation is that $n = 200$ may be too small relative to ρ. Hence, better approximation of stationary bootstrap would be obtained when the sample becomes larger.

References

Beran, R. and G. R. Ducharme (1991). *Asymptotic Theory for Bootstrap Methods in Statistics*, Les Publication CRM, Centre de recherches mathématiques, Université de Montréal, Montréal, Canada.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Annals of Statistics*, **7**, 1–16.

Table 2: The coverage rates of the stationary bootstrap method.

ρ	Q=0	0.5	0.7	0.9	0.95
0	0.9514	0.9414	0.9466	0.9284	0.8982
0.3	0.8494	0.8944	0.9178	0.9150	0.8822
0.6	0.6726	0.8100	0.8502	0.8690	0.8822
0.9	0.3460	0.5314	0.6214	0.7562	0.7742

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society of Industrial and Applied Mathematics, CBMS–NSF Monograph 38.

Goncalves, S. and R. de Jong (2003). Consistency of the stationary bootstrap under weaker moment conditions, *Economics Letters*, **81**, 273–278.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Horowitz, J. L. (2001). The bootstrap, in J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5, pp. 3161–3228, Amsterdam: Elsevier.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap, *Journal of the American Statistical Association*, **89**, 1303–1313.