LECTURE ON BOOTSTRAP

CHUNG-MING KUAN

Department of Finance

National Taiwan University

April 18, 2011

Outline

1. Introduction
2. Basic Idea of Bootstrap
   - Confidence Intervals
   - Basic Idea
   - Examples
3. Asymptotic Results
4. Bootstrap with Re-Sampling
   - Re-Sampling
   - Examples
5. Stationary Bootstrap

Introduction

- Inference about a statistic should be based on its exact distribution.
  - The exact distribution is typically unknown.
  - Asymptotic distribution (based on first-order asymptotics) is usually easier to obtain under mild conditions and provides a reasonably good approximation to the exact distribution.
- Efron (1979): Bootstrap yields an alternative approximation to the exact distribution based on re-sampling of the data.
  - The approximation is usually more accurate than that of the first-order asymptotics.
  - It is computationally demanding.
- The results and discussion here are taken freely from Horowitz (2001, *Handbook of Econometrics*, Chap. 52).

Notations

- $\mathbf{X}_n = \{X_1, X_2, ..., X_n\}$, where X_i are i.i.d. with the distribution function (df) $\mathbf{F}$.
- $R(\mathbf{X}_n)$ is a statistic based on $\mathbf{X}_n$ with the exact df $H_n(\cdot, \mathbf{F})$:

$$H_n(a, \mathbf{F}) = P_{\mathbf{F}}\left[R(\mathbf{X}_n) \leq a\right].$$

  - $R(\mathbf{X}_n)$ is a pivot if $H_n(\cdot, \mathbf{F})$ are identical for all $\mathbf{F} \in \mathcal{F}$.
  - $R(\mathbf{X}_n)$ is an asymptotic pivot if its limiting df,

$$H_A(a, \mathbf{F}) := \lim_{n \to \infty} H_n(a, \mathbf{F}),$$

    does not depend on $\mathbf{F}$.
- It is common to approximate $H_n(a, \mathbf{F})$ by its limiting df $H_A(a, \mathbf{F})$.

Exact Confidence interval

- Given X_i i.i.d. $\mathcal{N}(\mu, \sigma^2)$, consider

$$R(\mathbf{X}_n) = \frac{\hat{\mu}(\mathbf{X}_n) - \mu}{\sqrt{\frac{\hat{\sigma}^2(\mathbf{X}_n)}{n}}} \sim t(n-1),$$

where $\hat{\mu}(\mathbf{X}_n) = \sum_{i=1}^n X_i/n$, $\hat{\sigma}^2(\mathbf{X}_n) = \sum_{i=1}^n \big(X_i - \hat{\mu}(\mathbf{X}_n)\big)^2/(n-1)$. As long as $\mathbf{F}$ is normal, $R(\mathbf{X}_n)$ is a pivot.

- The exact confidence interval of μ with the confidence coefficient α is

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

where $t_{n-1,\frac{1+\alpha}{2}}$ is the $(1+\alpha)/2$-th quantile of $t(n-1)$.

Asymptotic Confidence interval

- If X_i are i.i.d. with finite second moment (but not necessarily normally distributed), a CLT yields

$$R(\mathbf{X}_n) \xrightarrow{D} \mathcal{N}(0,1).$$

  $R(\mathbf{X}_n)$ is an asymptotic pivot because its limiting normal distribution does not depend on $\mathbf{F}$ (as long $\mathbf{F}$ has finite second moment).

- An approximate confidence interval of μ with the confidence coefficient α is

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

  where $q_{\frac{1+\alpha}{2}}$ is the $(1+\alpha)/2$-th quantile of $\mathcal{N}(0,1)$.

Basic Idea of Bootstrap

- Bootstrap approximates $H_n(\cdot, \mathbf{F})$ using $H_n(\cdot, \widehat{\mathbf{F}}_n)$, where $\widehat{\mathbf{F}}_n$ is an estimate of $\mathbf{F}$.
- Estimates of $\mathbf{F}$:
  - Parametric: Suppose $\mathbf{F}$ is determined by the parameters $\mathbf{m} \in \mathbb{M} \subseteq \mathbb{R}^k$, so that $\mathcal{F} = \{\mathbf{F}(\cdot, \mathbf{m}) | \mathbf{m} \in \mathbb{M}\}$. Then,

$$\widehat{\mathbf{F}}_n(a) = \mathbf{F}(a, \widehat{\mathbf{m}}(\mathbf{x}_n))$$

  - Nonparametric: Empirical distribution function of X_i is

$$\widehat{\mathbf{F}}_n(a) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i \leq a) = \frac{1}{n} \sharp\{x_i \leq a, \ i = 1, \ldots, n\},$$

where $\mathbf{1}(A)$ is the indicator function of the event A.

Example 3.1: Parametric Bootstrap

- X_i are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Suppose the normality is known but not μ and σ^2 which can be estimated by $\hat{\mu}(\mathbf{X}_n)$ and $\hat{\sigma}^2(\mathbf{X}_n)$.
- Consider the distribution $\mathcal{N}(\hat{\mu}(\mathbf{X}_n), \hat{\sigma}^2(\mathbf{X}_n))$, from which we can randomly draw $\mathbf{X}_n^* = \{X_1^*, X_2^*, ..., X_n^*\}$. Then, X_i^* are i.i.d. with

$$\mathbf{F}^*(a) := \widehat{\mathbf{F}}_n(a) = \Phi\big((a - \hat{\mu}(\mathbf{x}_n))/\hat{\sigma}(\mathbf{x}_n)\big).$$

- As R is a pivot, $R(\mathbf{X}_n) \sim t(n-1)$ and

$$R(\mathbf{X}_n^*) := \frac{\hat{\mu}^*(\mathbf{X}_n^*) - \hat{\mu}(\mathbf{x}_n)}{\sqrt{\frac{\hat{\sigma}_*^2(\mathbf{X}_n^*)}{n}}} \sim t(n-1).$$

This shows that $H_n(\cdot, \mathbf{F}^*)$ agrees with $H_n(\cdot, \mathbf{F})$.

- As $H_n(\cdot, \mathbf{F}^*)$ agrees with $H_n(\cdot, \mathbf{F})$,

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

- The bootstrapped CI is exact and reads

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

- Consider the statistic $R(\mathbf{X}_n) = \sqrt{n}(\hat{\mu}(\mathbf{X}_n) - \mu)$, where X_i are as in Example 3.1. Then, $R(\mathbf{X}_n) \sim H_n(a, \mathbf{F}) = \Phi(a/\sigma)$.
- Consider the distribution $\mathcal{N}(\hat{\mu}(\mathbf{X}_n), \hat{\sigma}^2(\mathbf{X}_n))$, from which we can randomly draw $\mathbf{X}_n^*$ with

$$\mathbf{F}^*(a) = \Phi\big((a - \hat{\mu}(\mathbf{x}_n))/\hat{\sigma}(\mathbf{x}_n)\big).$$

  Then, $R(\mathbf{X}_n^*) := \sqrt{n}\big(\hat{\mu}^*(\mathbf{X}_n^*) - \hat{\mu}(\mathbf{x}_n)\big) \sim H_n(\cdot, \mathbf{F}^*) = \Phi\big(a/\hat{\sigma}(\mathbf{x}_n)\big)$.
- $\mathbb{P}_{\mathbf{F}}\left[q_{\frac{1-\alpha}{2}}\sigma < R(\mathbf{X}_n) < q_{\frac{1+\alpha}{2}}\sigma\right] = \alpha$ can be approximated by

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

- The approximated confidence interval of μ is thus

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

- If the MLE $\check{\sigma}^2(\mathbf{x}_n) = \sum_{i=1}^n \big(x_i - \hat{\mu}(\mathbf{x}_n)\big)^2/n$ is used, we have

$$\mathbf{F}^*(a) = \Phi\big((a - \hat{\mu}(\mathbf{x}_n))/\check{\sigma}(\mathbf{x}_n)\big).$$

- The approximated confidence interval of μ is

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

Note: The parametric bootstrap method depends on the choice of $R(\mathbf{X}_n)$ as well as the estimator of parameters.

Example 3.3: Non-Parametric Bootstrap

- X_i^* are i.i.d. with the df $\mathbf{F}^* = \widehat{\mathbf{F}}_n$, the empirical distribution function.
- Calculate $R(\mathbf{X}_n^*)$ over n^n different combinations of $\{x_i^*\}_{i=1}^n$, so that

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

- Letting p_s^* denote the s-th quantile of $H_n(\cdot, \mathbf{F}^*)$, we have

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

The approximated confidence interval of μ is

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

Asymptotic Results

Definition: Consistency

$H_n(\cdot, \widehat{\mathbf{F}}_n)$ is said to be consistent for $H_A(\cdot, \mathbf{F})$ if for every $\epsilon > 0$ and $\mathbf{F} \in \mathcal{F}$,

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

The following conditions are due to Beran and Ducharme (1991):

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

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

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

- Polya's Theorem: If $X_n \xrightarrow{D} X$ and F_X is continuous, then

$$\lim_{n\to\infty} \sup_a |F_{X_n}(a) - F_X(a)| = 0.$$

- By Polya's Theorem, conditions (ii) and (iii) imply

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

- For $\widehat{\mathbf{F}}_n$ satisfying condition (i), we have with probability approaching one, $\widehat{\mathbf{F}}_n(a)$ is close to $F(a)$ uniformly in a. This, together with the result above, leads to

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

That is, $H_n(\cdot, \widehat{\mathbf{F}}_n)$ is consistent for $H_A(\cdot, \mathbf{F})$:

- When $R(\mathbf{X}_n) \xrightarrow{D} R_A(\mathbf{F})$ and $H_A(\cdot, \mathbf{F})$ is continuous, Polya's theorem again ensures the convergence of $H_n(a, \mathbf{F})$ to $H_A(a, \mathbf{F})$ is uniform.
- This shows that the bootstrap distribution $H_n(a, \widehat{\mathbf{F}}_n)$ is capable of approximating the exact distribution $H_n(a, \mathbf{F})$, in the sense that

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

Bootstrap with Re-Sampling

Nonparametric bootstrap is computationally burdensome (for $n = 10$, it requires 10^{10} values). This may be greatly simplified by re-sampling.

- Randomly draw n observation from $\{x_1, x_2, \dots, x_n\}$ with replacement: $\mathbf{x}^*_{n,b} = (x^*_{1,b}, x^*_{2,b}, \dots, x^*_{n,b})$, for $b = 1, 2, \dots, B$.
- Empirical distribution function of $R(\mathbf{x}^*_n)$ is

  $$\widetilde{H}_{n,B}(a, \mathbf{F}^*) = \frac{1}{B} \sharp \{ R(\mathbf{x}^*_{n,b}) \leq a, \;\; b = 1, \dots, B \},$$

  and by the Glivenko-Cantelli theorem,

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

- $\widetilde{H}_{n,B}(\cdot, \mathbf{F}^*)$ approximates $H_n(\cdot, \mathbf{F}^*)$ when B is large, and $H_n(\cdot, \mathbf{F}^*)$ in turn approximates $H_n(\cdot, \mathbf{F})$ when n is large.

Example 5.1

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

	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
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

Example 5.2

Given $y_i = \alpha + \beta x_i + \epsilon_i$, 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 $\hat{\sigma}_{\hat{\beta}}$. The i.i.d. bootstrap is

I1. Generate random indices from a uniform distribution over $\{1, .., n\}$ with replacement, denoted as $\{k_1^b, ..., k_n^b\}$.

I2. 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 statistic: $\hat{R}_b^* := (\hat{\beta}_b^* - \hat{\beta})/\hat{\sigma}_{\hat{\beta}_b^*}$.

I3. Repeat the steps (i) and (ii) for $b = 1, ..., B$ and rank the absolute value of $\hat{R}_b^*$ in an ascending order: $\{\hat{R}_{r_1}^*, ..., \hat{R}_{r_B}^*\}$.

Example 5.2 (Continued)

1. The bootstrapped 95% CI based on $\hat{\sigma}_{\hat{\beta}}$ 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 $\{\hat{R}^*_{r_1}, \ldots, \hat{R}^*_{r_B}\}$.

2. An alternative CI is

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

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

3. The CI based on non-Studentized statistic $(\hat{\beta}^*_b - \hat{\beta})$ is

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

   where $\tilde{p}^*_{0.95}$ is the 0.95-th quantile of $(\hat{\beta}^*_b - \hat{\beta})$.

Table: The coverage rates of β in simple linear regression.

	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
$\mathcal{F}_x/\mathcal{F}_\epsilon$	Boot	Asymp	Boot	Asymp	Boot	Asymp	Boot	Asymp
$\mathcal{N}(0,1)/t(5)$	0.918	0.911	0.937	0.925	0.942	0.932	0.945	0.945
$\mathcal{N}(0,1)^2/t(3)$	0.917	0.912	0.927	0.906	0.919	0.931	0.934	0.939
$e^{\mathcal{N}(0,5)}/\mathcal{N}(0,1)$	0.932	0.922	0.938	0.934	0.901	0.956	0.910	0.951

Stationary Bootstrap

- Stationary bootstrap of Politis and Romano (1994):
  - It is applicable to stationary and weakly dependent data.
  - Observations are re-sampled in blocks so as to capture the dependence in data.
  - Each block has a random size determined by the geometric distribution with parameter Q.
- Given $\mathbf{x}_n$ and $0 < Q < 1$, the procedure is:

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

S2. With prob Q, $x^*_{2,b}$ is set to x_{t+1}, the obs following the previously sampled obs, and with prob $1 - Q$, the second bootstrapped obs $x^*_{2,b}$ is randomly selected from the original data $\mathbf{x}_n$.

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

Goncalves and de Jong (2003)

Suppose that $Q(n) \to 1$ and $n(1 - Q(n))^2 \to \infty$. Then for any $\epsilon > 0$,

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

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

- Expected block size: $1/(1 - Q)$
- Stationary bootstrap is close to i.i.d. bootstrap when $Q \to 0$.
- The larger the expected block size (the larger the Q), the better can such re-sampling preserve the dependence in data. But when the expected block size is too big, the bootstrapped samples would have smaller variation and hence result in poor approximation.

Example 6.1

- $X_t = \rho X_{t-1} + \varepsilon_t$, with $|\rho| < 1$ and $\varepsilon_t \sim \mathcal{N}(0,1)$.
- Simulating the coverage rates of 95% confidence intervals of the mean of X_t; $B = 1000$ and $R = 5000$.

Table: 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