Session 07: Bootstrap

Professor: Olivier Scaillet (olivier.scaillet@unige.ch)
Teaching Assistant: Afrae Hassouni (afrae.hassouni@ulb.be)

Introduction

Asymptotic properties of estimators are valid when sample size is large

But what happens when the size of the sample is small? → Asymptotic properties may provide poor approximations of the real distribution of estimators

- In particular, confidence intervals based on asymptotic normality may be too wide or too narrow

Solution: when samples are small, use simulations to get larger samples and hence better approximation for the estimators

Steps to Build Bootstrap Confidence Interval for Parameter θ

1. Estimate $\hat{\theta}$ from the original sample

2. Generate B (very large) random samples of size T **from the original data** **with replacement**

If we draw without replacement, at the end we will get the same sample as the original one.
With replacement, we will get the same values several times but with small probabilities.

The samples are drawn from the original data.
If we do it sequentially, we will end up with a constant at the end.

3. …

4. $- \hat{\theta}$)is given by:

$$\leq q_{(1-\alpha/2)} - \hat{\theta}\big) = 1 - \alpha$$

In general, we do not necessarily know anything about the distribution of the new parameter estimates $\hat{\theta}^{(b)}$. But, although we may not have a perfect idea of the shape of this distribution, we can calculate quantiles $q_{(\alpha)}$, such that a fraction α of the bootstrap statistics are less than or equal to $q_{(\alpha)}$. For example, if we have 1000 bootstrap values, the quantile $q_{(0.05)}$ will be the 50th largest bootstrap statistic.

Steps to Build Bootstrap Confidence Interval for Parameter θ

1. Estimate $\hat{\theta}$ from the original sample

2. Generate B (very large) random samples of size T from the original data with replacement

3. For each simulated sample $b = 1, \ldots, B$ of size T, compute the estimate $\hat{\theta}^{(b)}$

4. The confidence interval for $\left(\hat{\theta}^{(b)} - \hat{\theta}\right)$is given by:

$$P\left(q_{(\alpha/2)} - \hat{\theta} \leq \hat{\theta}^{(b)} - \hat{\theta} \leq q_{(1-\alpha/2)} - \hat{\theta}\right) = 1 - \alpha$$

In general, we do not necessarily know anything about the distribution of the new parameter estimates $\hat{\theta}^{(b)}$. But, although we may not have a perfect idea of the shape of this distribution, we can calculate quantiles $q_{(\alpha)}$, such that a fraction α of the bootstrap statistics are less than or equal to $q_{(\alpha)}$. For example, if we have 1000 bootstrap values, the quantile $q_{(0.05)}$will be the 50$^{\text{th}}$ largest bootstrap statistic.

Steps to Build Bootstrap Confidence Interval for Parameter θ

5. In addition, we can argue that we can estimate the distribution of $\sqrt{T}\left(\hat{\theta}-\theta\right)$ by the distribution of $\sqrt{T}(\hat{\theta}^{(b)}-\hat{\theta})$. This makes sense if we assume that $\hat{\theta}$ arises from sampling from a distribution with parameter θ, while $\hat{\theta}^{(b)}$ arises from sampling from a distribution with parameter $\hat{\theta}$

$$P\left(q_{(\alpha/2)}-\hat{\theta} \leq \hat{\theta}^{(b)}-\hat{\theta} \leq q_{(1-\alpha/2)}-\hat{\theta}\right)$$
$$= P\left(q_{(\alpha/2)}-\hat{\theta} \leq \hat{\theta}-\theta \leq q_{(1-\alpha/2)}-\hat{\theta}\right) = 1-\alpha$$

Rearranging:

$$P\left(2\hat{\theta}-\ q_{\left(1-\frac{\alpha}{2}\right)} \leq \theta \leq 2\hat{\theta}-q_{\left(\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

This means the bootstrap confidence interval for θ is given by:

$$CI(1-\alpha) = \left[2\hat{\theta}-q_{\left(1-\frac{\alpha}{2}\right)}; 2\hat{\theta}-q_{\left(\frac{\alpha}{2}\right)}\right]$$

Bootstrap - Remarks

The bootstrap procedure only works with independent data

If a dependency between neighboring observations is suspected (e.g. high frequency stock data), the bootstrap procedure will not work

➔ How to build confidence intervals when data are dependent?

The Block Bootstrap

Block bootstrap is the bootstrap method used when data are dependent. In the block bootstrap, groups (i.e., blocks) of consecutive observations are sampled instead of individual observations.

In its simplest form, the data is divided into k non-overlapping blocks of length l, where $\mathrm{T} = \mathrm{k} * l$. If l is large enough, then the generated samples should preserve most of the dependency between neighboring observations (rule of thumb: $l = T^{1/3}$).

Under a number of relatively mild conditions, it can be shown that this estimator is consistent, though its rate of convergence may not be as high as that for the *i.i.d.* bootstrap seen above.

Block Bootstrap Confidence Intervals - Remarks

Block Bootstrap intervals tend to be wider than the traditional bootstrap intervals

This means that the block bootstrap is usually not as accurate as the traditional bootstrap

However, block-bootstrap allows us to deal with a degree of dependency between observations not possible with the traditional bootstrap

Question 1: Bootstrap Confidence Interval

1. Simulate T = 100 observations of the law $\mathcal{N}(1{,}16)$. (We will consider these observations as a series of monthly percentage returns on a stock).

2. Compute the statistic $\widehat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}$ based on the total sample. Think of $\widehat{SR}$ as a Sharpe ratio, if we assume the risk-free rate is zero.

3. Form a 95% bootstrap confidence interval for SR (hint: use the function *boot()* from the package "*boot*" on R).

Question 1: Bootstrap Confidence Interval

Bootstrap confidence interval for SR:

$$[0.1173; 0.5130]$$

<u>Remark</u>: because we use random data in this exercise, you will not obtain the same confidence interval as the one above.

Question 2: Block Bootstrap

1. Generate again B = 2000 random samples of size T from the original data simulated in the 1st exercise. This time, however, do the sampling based on 10 blocks (i.e. k = 10). Use each of the generated samples to obtain new estimates of our statistic SR, denoted by $\widehat{SR}^{(BB)}, b = 1, \ldots, B$

2. Using a similar approach to the previous exercise, form a 95% bootstrap confidence interval for SR (hint: use function *tsboot()* from package "*boot*" on R).

3. How does the new confidence interval obtained from block sampling compare with the confidence interval computed in the previous section? Is it wider?

   Repeat questions (1)-(2) several times to get a good idea of how both confidence intervals compare. Comment!

Question 2: Block Bootstrap

1. Block bootstrap confidence interval*:

$$[0.1174; 0.5199]$$

**The confidence interval is built based on the same approach as in question 1.*

2. Comparing the bootstrap confidence interval and the block bootstrap confidence interval:

Confidence Interval at 95%	
Bootstrap	**Block Bootstrap**
$[0.1173; 0.5130]$	$[0.1174; 0.5199]$

The block bootstrap interval appear to be wider than the traditional bootstrap interval