Lehigh University
Department of Economics

Muzhe Yang
Spring 2013

Bootstrap Methods[1]

- Bootstrap due to Efron (1979):
  - Provides an alternative approximation for the asymptotic distribution of a statistic.
  - Does so by viewing the sample as the *population* and obtaining B *resamples* leading to B realizations of the statistic.
  - There are many, many ways to bootstrap.
- A bootstrap without asymptotic refinement:
  - is no better than regular asymptotic theory;
  - though is popular as it may be simpler to implement;
  - leading example is bootstrap estimate of standard errors.
- A bootstrap with asymptotic refinement:
  - is asymptotically better than regular asymptotic theory;
  - hopefully then does better in finite samples;
  - leading example is the bootstrap-t method.

1 Bootstrap Dos and Don'ts

Four commandments from Horowitz (2001):

> 1) **Do** use the bootstrap to estimate the probability distribution of an asymptotically pivotal statistic or the critical value of a test based on an asymptotically pivotal statistic whenever such a statistic is available;
> 2) **Don't** use the bootstrap to estimate the probability distribution of a non-asymptotically-pivotal statistic such as a regression slope coefficient or standard error if an asymptotically pivotal statistic is available;
> 3) **Do** recenter the residuals of an over-identified model before applying the bootstrap to the model;
> 4) **Don't** apply the bootstrap to models for dependent data, semi- or nonparametric estimators, or non-smooth estimators without first reading Section 4 of Horowitz (2001).

2 Bootstrap Summary

Follows MMA closely.

i.i.d. sample of size N: $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N\}$
the i-th observation: $\mathbf{w}_i = (y_i, \mathbf{x}_i')'$

$$\widehat{\theta} \text{ satisfies}^2\text{:} \begin{cases} \widehat{\theta} \xrightarrow{p} \theta_0 \\ \sqrt{N}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, V) \end{cases}$$

$$\text{statistics of inference interest:} \begin{cases} \text{standard error: } se(\widehat{\theta}) \Rightarrow \text{confidence interval for } \theta_0 \\ t\text{-statistic: } t = \frac{\widehat{\theta}-\theta_0}{se(\widehat{\theta})} \Rightarrow \text{critical value or } p\text{-value} \end{cases}$$

[1] This handout is supplemental to MMA chapter 11. To simplify notation, all "$AsyVar$" is written as "Var" throughout this handout.

[2] For simplicity, we present results for scalar θ. For vector $\boldsymbol{\theta}$ in most instances replace $\boldsymbol{\theta}$ by θ_j, the j-th component of $\boldsymbol{\theta}$.

Muzhe Yang
Spring 2013

2.1 Bootstrap without asymptotic refinement

2.1.1 Nonparametric bootstrap

If we just know that y ∼i.i.d.(μ, σ^2), then the sample mean $\widehat{\mu} = \overline{y}$ satisfies $\widehat{\mu} \xrightarrow{p} \mu$ and $\sqrt{N}(\widehat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2)$. There are two ways to get the asymptotic variance of the estimate $\widehat{\mu}$, which is crucial for inferences.

(a) Analytical result: $Var(\widehat{\mu}) = \sigma^2/N$ if σ^2 is known and $\widehat{Var}(\widehat{\mu}) = \widehat{\sigma}^2/N$ if σ^2 is unknown.

(b) Bootstrap (nonparametric bootstrap): extremely useful when analytical result is not available or computationally intractable.

(b-1) Pretend that you have the population, although what you actually have is just a sample from the population. The magic of bootstrap comes from your imagination (or willingness to assume) that the sample at your hand were the "population."

(b-2) Now the "population" becomes $\{y_1, y_2, \cdots, y_N\}$.

(b-3) To get the distribution of $\widehat{\mu}$, we need as many realizations of $\widehat{\mu}$ as possible, say B realizations.

(b-4) Then we need resampling from the "population" *with replacement* B times and in each resampling ($b = 1, \cdots, B$) we obtain a sample from the "population" of the same size N, which is just as large as the "population" and is called a *bootstrap sample.* Notice that in each bootstrap sample so obtained, some of the original data points will appear multiple times whereas others will not appear at all. This method is *an empirical distribution function (EDF) bootstrap* or *nonparametric bootstrap.* In single-equation regression models, it is also called a *paired bootstrap* since $\mathbf{w}_i = (y_i, \mathbf{x}_i')'$ both y_i and $\mathbf{x}_i$ are resampled jointly.

(b-5) For each bootstrap sample, we can get $\widehat{\mu}_b^* = \overline{y}_b$ ($b = 1, \cdots, B$), and now we have B estimates $\{\widehat{\mu}_1^*, \widehat{\mu}_2^*, \cdots, \widehat{\mu}_B^*\}$.

(b-6) The realizations of $\widehat{\mu}$ $\{\widehat{\mu}_1^*, \widehat{\mu}_2^*, \cdots, \widehat{\mu}_B^*\}$ actually gives us an empirical distribution of $\widehat{\mu}$, and therefore we can directly calculate any moments from this empirical distribution of $\widehat{\mu}$:

$$\begin{aligned}\overline{\widehat{\mu}} &= \frac{1}{B}\sum_{b=1}^{B}\widehat{\mu}_b^* \ \text{ (where } \widehat{\mu}_b^* = \overline{y}_b);\\ \widehat{Var}(\widehat{\mu}) &= \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\mu}_b^* - \overline{\widehat{\mu}})^2 \rightarrow \text{good estimate for } Var(\widehat{\mu}) = \sigma^2/N.\end{aligned}$$

(b-7) It turns out that $\sum_{b=1}^{B}(\widehat{\mu}_b^* - \overline{\widehat{\mu}})^2/(B-1)$ is a very good estimate for $Var(\widehat{\mu}) = \sigma^2/N$ for observations $\mathbf{w}_i$ that are i.i.d. over i.

2.1.2 Parametric bootstrap

If we know that y ∼i.i.d. $N(\mu, \sigma^2)$, then we can obtain B bootstrap samples, in a parametric way, of size N by drawing from:

$$\underset{N\times 1}{\mathbf{y}_b} \sim N\left(\boldsymbol{\iota}_N \cdot \widehat{\mu}, \widehat{\sigma}^2 \cdot I_N\right) \text{ where } b = 1, \cdots, B;\ \boldsymbol{\iota}_N = (1, \cdots, 1)'_{N\times 1};\ I_N \text{ is an } (N \times N) \text{ identity matrix.}$$

$$\Rightarrow \widehat{\mu}_b^* = \overline{y}_b, \overline{\widehat{\mu}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mu}_b^*.$$

$$\Rightarrow \widehat{Var}(\widehat{\mu}) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\mu}_b^* - \overline{\widehat{\mu}})^2.$$

For a regression model with additive i.i.d. disturbances (errors), $y_i = g(\mathbf{x}_i, \beta) + u_i$, we can obtain residuals $\widehat{u}_1, \widehat{u}_2, \cdots, \widehat{u}_N$, where $\widehat{u}_i = y_i - g(\mathbf{x}_i, \widehat{\beta})$. We then resample $\{\widehat{u}_1, \widehat{u}_2, \cdots, \widehat{u}_N\}$ and get a new draw of residuals $\{\widehat{u}_1^*, \widehat{u}_2^*, \cdots, \widehat{u}_N^*\}$, leading to a bootstrap sample $\{(y_1^*, \mathbf{x}_1), (y_2^*, \mathbf{x}_2), \cdots, (y_N^*, \mathbf{x}_N)\}$, where $y_i^* = g(\mathbf{x}_i, \widehat{\beta}) + u_i^*$. This bootstrap is called a *residual bootstrap.* It uses information intermediate between

Muzhe Yang
Spring 2013

the nonparametric and parametric bootstrap. It can be applied if the error term has a distribution that does not depend on unknown parameters.

2.2 Bootstrap with asymptotic refinement

Read MMA 11.2.2.

2.3 Asymptotic pivotal statistic

Read MMA 11.2.3. For example, consider $y \sim$i.i.d.(μ, σ^2):

$$\textit{not} \text{ asymptotically pivotal:} \qquad \widehat{\mu} = \overline{y} \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{N}\right) \text{ if } \sigma^2 \text{ is unknown;}$$

$$\text{asymptotically pivotal:} \qquad t = \frac{\widehat{\mu} - \mu}{se(\widehat{\mu})} \overset{a}{\sim} N(0,1).$$

2.4 A general bootstrap algorithm

(1) Procedure

[P1] Given data $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N$, draw a bootstrap sample of size N using a method given in the following and denote this new sample $\mathbf{w}_1^*, \mathbf{w}_2^*, \cdots, \mathbf{w}_N^*$.

[P2] Calculate an appropriate statistic using the bootstrap sample. Examples of this statistic are:

- **(a)** $\widehat{\theta}^*$, a consistent estimate of θ based on the *bootstrap* sample
- **(b)** $se(\widehat{\theta}^*)$, the standard error of $\widehat{\theta}^*$ calculated by available analytical results based on the *bootstrap* sample
- **(c)** $t^* = (\widehat{\theta}^* - \widehat{\theta})/se(\widehat{\theta}^*)$, where $\widehat{\theta}$ is the estimate of θ based on the *original* sample.

[P3] Repeat step [P1] and [P2] B independent times, where B is a large number, obtaining B bootstrap replications of the statistic of interest, such as $\widehat{\theta}_1^*, \cdots, \widehat{\theta}_B^*$ or $t_1^*, \cdots, t_B^*$.

[P4] Use these B bootstrap replications to obtain a bootstrapped version of the statistic.

(2) Bootstrap sampling methods

Step [P1] is used to approximate the true unknown data-generating-process (DGP). There are mainly two types of bootstrap: nonparametric bootstrap (a.k.a. empirical distribution function bootstrap or paired bootstrap) and parametric bootstrap. We emphasize the nonparametric sampling methods on the grounds of its simplicity and robustness due to its reliance on weak distributional assumptions.

(3) Number of bootstraps

Read MMA 11.2.4.

2.5 Standard error estimation

Read MMA 11.2.5.

2.6 Hypothesis testing

Read MMA 11.2.6.

2.7 Confidence intervals (CI)

Read MMA 11.2.7.

2.8 Bias reduction

Read MMA 11.2.8.

3 Example

Read MMA 11.3.

References

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7(1): 1-26.

Horowitz, J. L. (2001). "The Bootstrap," in *Handbook of Econometrics*, J. J. Heckman and E. Leamer (Eds.), Volume 5, 3159-3228, Amsterdam: North-Holland.