The Bootstrap

Another tool to measure uncertainty

Travis McArthur

February 18, 2015

Motivation

Say you face one of the following situations:

- You have a two-step estimator where the results of the first estimator are treated as fixed quantities for the second estimator
- Your sample is small and the asymptotic approximation is suspect
- You would like to know the bias of an estimator without working out the (potentially complicated) math of it
- It is difficult to obtain the uncertainty of your estimator for a given quantity of interest, such as the median

Motivation

The bootstrap is a procedure used to simulate the properties of the population by using only the sample of data that we have actually collected.

Procedure: Step 1

- Estimate your model using the full sample as normal
- Store the results

Procedure: Step 2

- Choose a number B, the number of bootstrap iterations. Larger is better; 1000 is usually good enough.
- Draw, with replacement, B independent samples of size N from your dataset
- You will inevitably have duplicate observations in each bootstrap sample

Procedure: Step 3

- With each bootstrap sample you have drawn, estimate your model
- Store the results for each sample

Procedure: Step 4

- Your bootstrapped estimates are a window into the performance of your estimator
- Extract the quantity of interest from your bootstrapped samples
- The extraction procedure varies according to what you are interested in, e.g.:
  - Confidence interval of a parameter or function of parameters
  - Test statistic for a hypothesis test
  - Standard error of a parameter
  - Estimate of bias of an estimator

Example: Estimate of standard error

- Generate bootstrap estimates of your parameter:

$$\left\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\right\}$$

- Compute the variance of your bootstrap estimates:

$$\hat{V}_n^* = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \overline{\hat{\theta}^*}\right)^2$$

- Use the usual standard deviation formula to get S.E.:

$$s^*\left(\hat{\theta}\right) = \sqrt{\hat{V}_n^*}$$

Example: Bias estimation

- Obtain parameter estimate with full sample: $\hat{\theta}$
- Generate bootstrap estimates of your parameter:

$$\left\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\right\}$$

- Estimate of the bias:

$$\widehat{bias}\left(\hat{\theta}\right) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* - \hat{\theta}$$

Example: Confidence intervals

- Generate bootstrap estimates of your parameter: $\left\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\right\}$
- $q_n(x)$ is the $x \cdot 100$ percentile of some set
- Then the bootstrap estimate of the $(1-\alpha) \cdot 100\%$ confidence interval for $\hat{\theta}$ is: $[\hat{q}_n^*(\alpha/2)\,,\ \hat{q}_n^*(1-\alpha/2)]$
- For 95%:

$$[\hat{q}_n^*(0.025)\,,\ \hat{q}_n^*(0.975)]$$

- This is the percentile confidence interval.
- Useful when sampling distribution of estimator is not symmetric

Different flavors of bootstrap

Above is the nonparametric bootstrap

- Resampling the residuals and not the full data – parametric
- Clustered or autocorrelated data: use block bootstrap

Theory supporting the bootstrap

Bradley Efron (1979)

- An estimator for the CDF of any statistic is the Empirical Cumulative Distribution Function (ECDF)
- To compute the full estimate, we need every combination of every row of data:

$$\binom{2n-1}{n}$$

- To make this feasible, we take a sample of this huge set of combinations

Further reading

- Cameron & Trivedi, Microeconometrics, 2005, Ch. 11.
- Hansen, Econometrics, 2015, Ch. 10. <http://www.ssc.wisc.edu/ bhansen/econometrics/>
- MacKinnon, "Bootstrap Methods in Econometrics", Economic Record, 2006, 82: S2-S18.