14.385
Nonlinear Econometrics

Lecture 7.

Theory: Consistency and Accuracy of Bootstrap.

Reference: Horowitz, Bootstrap.

Consistency of Bootstrap

The idea is that for large n

$$G_n(t, \widehat{F}) \approx G_\infty(t, \widehat{F}) \approx G_\infty(t, F_0) \approx G_n(t, F_0).$$

Consistency relies on asymptotics.

Def. $G_n(t, \widehat{F})$ is consistent if under each $F_0 \in \mathcal{F}$

$$\sup_t |G_n(t, \widehat{F}) - G_\infty(t, F_0)| \to_p 0,$$

Example 1. (Inference on Mean)

Statistic and parameter of interest:

$T_n = \sqrt{n}(\bar{X} - \theta), \quad \theta = \theta(F_0) = E_{F_0} X,$

Want to know:

$G_n(t, F_0) = P_{F_0}(T_n \leq t)$

Bootstrap DGP and Bootstrap "population" parameter:

$F_n =$ empirical df, $\theta(F_n) = E_{F_n} X = \bar{X}$

We create bootstrap samples $\{X_i^*\}$ by sampling from the original sample $\{X_i\}$ randomly with replacement.

Bootstrap version of T_n:

$T_n^* = \sqrt{n}(\bar{X}^* - \bar{X})$

Bootstrap gives us:

$G_n(t, F_n) = P_{F_n}(T_n \leq t)$.

Bootstrap should work as long as the limit distribution of T_n varies smoothly in F and if the (triangular) CLT holds with the same limit for any sequence in $\mathcal{F}$, and $\{\widehat{F}\} \in \mathcal{F}$ with probability one.

One formal theorem is as follows (this follows Horowitz):

Theorem (Bickel & Ducharme): $G_n(t, \widehat{F})$ is consistent if for each $F_0 \in \mathcal{F}$: (i) $\rho(\widehat{F}, F_0) \rightarrow_p 0$, (ii) $G_\infty(t, F)$ is continuous in t for each $F \in \mathcal{F}$, (iii) for any t and $F_n \in \mathcal{F}$ such that $\rho(F_n, F_0) \rightarrow 0$, we have

$$|G_n(t, F_n) - G_\infty(t, F_0)| \rightarrow 0,$$

for each t, where ρ is some metric.

The result follows from the extended continuous mapping theorem.

Proof: Clearly, $\rho(\widehat{F}, F_0) \rightarrow_p 0$ implies

$$|G_n(t, F_n) - G_\infty(t, F_0)| \rightarrow_p 0.$$

Next we apply Polya's Lemma: Pointwise convergence of a sequence of monotone functions to a bounded monotone continuous function implies that the sequence converges to this function uniformly. Thus, $\sup_t |G_n(t, F_n) - G_\infty(t, F_0)| \rightarrow_p 0$. $\square$

Remark: In way, the theorem is nearly at tautology, but it really helps organize thinking.

Remark*: Bickel and Friedman formally verify the conditions in Example 1 by checking the conditions of the above theorem using Kantarovich-Wasserstein-Mallows metric

$$\rho(P, Q) = \inf_{X,Y} \{E\|Y - X\|^2, Y \sim P, X \sim Q\}.$$

Theoretical Exercise. Supply the details of the proof. Hint: If you let $X_n = \Phi^{-1}(F_{Y_n}(Y_n)) =_d N(0, 1)$, where F_{Y_n} is the distribution function of Y_n and Φ^{-1} is the quantile function of the standard normal. Then $Y_n = \sqrt{n}(\bar{X} - \mu)/\sigma = F_{Y_n}^{-1}(\Phi(X_n)) \approx X_n + o(1)$, conditional on a set S that contains X_n with a probability close to one, which holds by the central limit theorem.

The following is a nice theorem due to Mammen:

Theorem (Mammen): Let $\{X_i, i \leq n\}$ be an iid sample from population. For a sequence of normalizing constants t_n and σ_n define:

$\bar{g}_n = \frac{1}{n}\sum g_n(X_i)$, $T_n = \frac{(\bar{g}_n - t_n)}{\sigma_n}$,
$\bar{g}_n^* = \frac{1}{n}\sum g_n(X_i^*)$, $T_n^* = \frac{(\bar{g}_n^* - \bar{g}_n)}{\sigma_n}$.

Nonparametric bootstrap is consistent if and only if T_n is asymptotically normal.

Proof: The sufficiency follows from Bickel and Friedman. For the necessity see Mammen's article. □

Example 2. (Inference in Regression)

Option (a) Bootstrap. (Asymptotically pivotal option (b) below is more accurate)

Data $X_i = (Y_i, W_i)$, where W_i is regressor, and Y_i dependent variable. Model $Y_i = W_i'\beta + \epsilon_i$.

Parameter of interest: $\theta(F_0) = \beta_j$.

Statistic of interest: $T_n = \sqrt{n}(\widehat{\beta} - \beta_j)$.

Want to know $G_n(t, F_0) = P_{F_0}(T_n \leq t)$.

$F_n =$ empirical df.

We create bootstrap samples $\{X_i^*\}$ by sampling from the original sample $\{X_i\}$ randomly with replacement.

Under bootstrap DGP the "population" parameter is $\theta(F_n) = \widehat{\beta}_j$.

For each bootstrap sample we compute $T_n^* = \sqrt{n}(\widehat{\beta}^* - \widehat{\beta}_j)$ (bootstrap realization of T_n).

Bootstrap gives us $G_n(t, F_n) = P_{F_n}(T_n \leq t)$.

Example 2. (Inference in Regression)

Option (b) This option bootstraps asymptotically pivotal statistic.

Data $X_i = (Y_i, W_i)$, where W_i is regressor, and Y_i dependent variable. Model $Y_i = W_i'\beta + \epsilon_i$.

Parameter of interest: $\theta(F_0) = \beta_j$, the j-th component of β.

Statistic of interest:

$$T_n = \sqrt{n}(\widehat{\beta}_j - \beta_j)/s.e.(\widehat{\beta}_j).$$

Want to know exact law $G_n(t, F_0) = P_{F_0}(T_n \leq t)$.

$F_n =$ empirical df.

We create bootstrap samples $\{X_i^*\}$ by sampling from the original sample $\{X_i\}$ randomly with replacement.

Under bootstrap DGP the "population" parameter is $\theta(F_n) = \widehat{\beta}_j$.

For each bootstrap sample we compute

$$T_n^* = \sqrt{n}(\widehat{\beta}_j^* - \widehat{\beta}_j)/s.e.^*(\widehat{\beta}_j)$$

where $s.e.^*(\widehat{\beta}_j)$ denotes the recomputed value of the standard error using bootstrap samples.

Bootstrap gives us $G_n(t, F_n) = P_{F_n}(T_n \leq t)$.

Option (b) is more accurate than the "natural" option (a).

Accuracy of Bootstrap.

This example has an interesting feature:

$$T_n \to_d N(0,1),$$

that is $G_n(t, F_0) \approx \Phi(t)$ in large samples, and the exact law is almost independent of DGP F_0. Therefore, $G_n(t, \hat{F}) \approx \Phi(t)$ too.

Def. A statistic T_n is called **asymptotically pivotal** relative to a class of DGPs $\mathcal{F}$, if its limit law does not depend on DGP F:

$$G_\infty(t, F) = G_\infty(t)$$

for all $F \in \mathcal{F}$.

Under asymptotic pivotality, the exact law is not very sensitive to the underlying DGP. Replacement of true DGP F_0 with $\hat{F}$ results in a good approximation of the exact law.

"Theorem": *The approximation error of bootstrap applied to asymptotically pivotal statistic is smaller than the approximation error of bootstrap applied to an asymptotically non-pivotal statistic.*

"Proof": A simple example is the case of the exact pivotality, where the bootstrap makes no error at all. For mean-like stat this claim as well as regularity conditions for it are made formal by appealing to Edgeworth expansion. See Horowitz for more details. □.

Theoretical Exercise. Make the claim formal for the case of bootstrapping the sample mean. Explain first the idea of the Edgeworth expansion: which is to first expand the characteristic function around the normal and then use the Fourier inversion to get the approximation to the distribution function. Then explain why in the case of asymptotically pivotal satistic (t-stat) the bootstrap makes smaller error than in the case of asymptotically non-pivotal statistic (sample mean).

The case is very close to the exact pivotality, which we used to tabulate the exact law *without* any approximation error.

Failure of Nonparametric Bootstrap:

I. When normal CLT does not apply for average-like statistics

Recall from 381, this happens e.g. when $X_i's$ have infinite variance. Under such conditions $\bar{X}$ is approximately distributed as a stable Pareto-Levy variable.

If you don't recall 381, the idea is conveyed by the following example:

Example: Cauchy Average If $X_i \sim Cauchy$, then $\bar{X} \sim Cauchy$, since sum of i.i.d. Cauchy is Cauchy. Therefore, bootstrap fails by Mammen's theorem.

Cauchy variable has no mean and has infinite variance. The behavior of the Cauchy mean is entirely driven by the extreme order statistics, whose behavior the nonparametric bootstrap fails to approximate entirely. Indeed, in any finite sample, the nonparametric bootstrap DGP has finite support and all finite moments, but the true DGP has no moments. Thus, true DGP and bootstrap DGP will predict very different behaviors for the sample mean.

Practical Relevance: Firm sizes, city sizes X_i have Cauchy-like tails (Zipf's law).

Foreign Exchange rates seem to have Cauchy-like tails.

II. When limit distribution $G_\infty(t, F)$ is not continuous with respect to perturbations in F

If so, $G_\infty(t, \hat{F})$ can deviate a great deal from $G_\infty(t, F_0)$.

Case **I** is actually a special case of this one.

Example: Distribution of Maximum of a Sample.

$X_i \sim U(0, \theta_0)$ i.i.d.,

$T_n = n(\theta_n - \theta_0)$, $\theta_n = \max\{X_1, ..., X_n\}$,

$F_n = EDF$ (nonparametric bootstrap)

$T_n^* = n(\theta_n^* - \theta_n)$, $\theta_n = \max\{X_1^*, ..., X_n^*\}$,

Then one can show that

$$T_n \to_d Exponential = \ln U(0, 1)$$

but

$$T_n^* \to_d \quad Something \quad Else,$$

since

$P_{F_n}[T_n^* = 0] = 1 - P_{F_n}[T_n^* > 0] = 1 - (1 - 1/n)^n \to 1 - e^{-1}$.

The nonparametric bootstrap tells us that there is a pointmass at 0, but the limit exponential variable we found above has no point masses.

Therefore nonparamatric bootstrap is inconsistent.

It is easy to show that the parametric bootstrap is consistent in this case, as well as in the case of Cauchy.

More Substantial Examples of Nonparametric Bootstrap Failure:

1. Extreme and Near-Extreme Quantile Regression

2. Non-regular Maximum Likelihood for Auction, Price Search Models, and Frontier Models.

In these case, one can use either (variants of) parametric bootstrap or subsampling.

Subsampling.

Idea: draw subsamples of size $m < n$ from the original data

(i) with replacement = m out of n bootstrap

(ii) without replacement = subsampling bootstrap.

Subsampling works when nonparametric bootstrap does not.

Typically less accurate than bootstrap, when the latter works.

For the precise details of the algorithms, see:

Reference: Horowitz, Bootstrap.

Reference: Romano, Politis, Wolf, Subsampling.

The m Out of n Bootstrap:

The estimate of $G_n(t, F_0)$ is $G_m(t, F_n)$.

Consistency arguments can be made similarly to the n out of n case.

However, drawing fewer observations often fixes cases of failure via a smoothing mechanism:

Example: Distribution of a Maximum. Recall our extremes example, then

$$P_{F_n}[T_m^* = 0] = 1 - (1 - 1/n)^m \approx 1 - e^{-m/n} \approx 0,$$

if $m/n \to 0$. Thus the point-mass problem goes away.

The Subsampling Bootstrap:

Provided the number of subsamples is large, i.e. $b/n \to 0$, then the estimate of $G_n(t, F_0)$ is given by

$$G_m(t, F_0) + o_p(1),$$

because the subsamples of size m are drawn from the true DGP F_0. As $m \to \infty$,

$$G_m(t, F_0) \to G_\infty(t, F_0).$$

Theoretical Exercise. Supply details of the proof of subsampling consistency for the case of the sample mean. See Horowitz for an outline or Politis and Romano article.

Reference. Politis, Dimitris N.; Romano, Joseph P. Large sample confidence regions based on subsamples under minimal assumptions. Ann. Statist. 22 (1994), no. 4, 2031–2050.

Empirical Example 1 (Bootstrap):

Bertrand, Duflo, Mullainathan, "Should we trust difference-in-difference estimators?"

Document success of bootstrap for estimation of policy effects. Works better than robust standard errors.

X_i is a vector of data on individual i.

Bootstrap "resamples" individuals to form new samples.

Statistic of interest is

$$T_n = (\widehat{\beta}_j - \beta_j)/s.e.(\widehat{\beta}_j).$$

Bootstrap statistic

$$T_n^* = (\widehat{\beta}_j^* - \widehat{\beta}_j)/s.e^*(\widehat{\beta}_j).$$

is recomputed for each bootstrap sample.

Then we use quantiles of simulated distribution of T_n^* as critical values for tests or use them to form confidence regions.

Why successful?

Empirical Example 2 (Subsampling)

Chernozhukov, V. "Inference for Extremal Conditional Quantiles, with an Application to Birthweights." (www.mit.edu/˜vchern)

Parameter of interest θ: very low percentiles of birthweights conditional on quality of prenatal care, smoking, and other characteristics.

A sample of black mothers.

The estimate $\widehat{\theta}$ is based on extremal regression quantiles.

Limit distribution: for some $\mathcal{A}_n$

$$\mathcal{A}_n(\widehat{\theta} - \theta) \rightarrow_d M,$$

where M is a functional of a marked Poisson process.

Using (a form of) subsampling on the statistic $\mathcal{A}_n(\widehat{\theta} - \theta)$ leads to straightforward inference that requires no knowledge of M.