ECON-GA 2100 Notes on the Bootstrap

Konrad Menzel
km125@nyu.edu

October 16, 2025

The bootstrap is a strategy for estimating the sampling distribution of some statistic or estimator. It is frequently used for various tasks in estimation and inference, most importantly to obtain standard errors, critical values for tests or confidence intervals, and bias correction. A great initial reference for the bootstrap is the handbook chapter by Horowitz (2000).

1 Problem Description

Suppose we have a sample $Z_1, \dots, Z_n$ of observations with marginal distribution F_0, e.g. for regression $Z_i = (Y_i, X_i')'$ would consist of the dependent variable and regressors for the ith observation. For now we will focus on the case of i.i.d. observations

$$Z_i \overset{iid}{\sim} F_0(z) \tag{1}$$

There exist adaptations of the bootstrap for various settings in which observations may be dependent or non-identically distributed.

We are considering inference based on a statistic

$$T_n := T_n(z_1, \dots, z_n)$$

which could be e.g. a test statistic, estimator, or any other function of the data. The (exact) finite-sample distribution is given by the c.d.f.

$$G_n(t, F_0) := \mathbb{P}_{F_0}(T_n \leq t)$$

We are often interested in estimating percentiles of that distribution which could serve as critical values for hypothesis tests or confidence intervals based on the statistic T_n. If T_n is an estimator of a parameter, we may also want to estimate the first two moments of $G_n(t, F_0)$ to determine the bias and standard error of that estimator.

Clearly, if the true distribution F_0 of the data was known, we could in principle compute $G_n(t, F_0)$ since the exact form of the statistic as a function of the data is also known. In practice, there are two problems in working directly with $G_n(t, F_0)$,

- It is in general difficult to evaluate the exact form of $G_n(t, F)$, which may be a complicated function of F.
- More importantly, the sampling distribution of T_n depends on the unknown population distribution F.

So far we have seen two approaches:

1. Obtain the finite-sample distribution under strong auxiliary assumptions, e.g. $t-$ and $F-$distributions with the appropriate degrees of freedom for $t-$ and $F-$ratios in the Gaussian homoskedastic LRM.
2. Approximate $G_n(t, F_0)$ with the asymptotic (large sample) distribution of T_n,

$$G_\infty(t, \hat{F}_n) := \lim_n G_n(t, \hat{F}_n)$$

for an estimate $\hat{F}_n$ of the distribution or its relevant moments. E.g. in many settings we can apply a CLT to obtain Gaussian or chi-square limiting distributions with an appropriate variance or degrees of freedom for estimators or quadratic forms of sample moments.

The assumptions needed for the first approach are usually too restrictive to be plausible in practice. The asymptotic approach relies on an approximation, but may work quite well under relatively broad assumptions as long as n is sufficiently large. The main advantages of using $G_\infty(t, F)$ as an approximation to the exact distribution of T_n are that limiting distributions from central limit or extreme value theory usually

- have a very simple structure and are easy to evaluate,
- and only depend on a small number of features of the population distribution F, e.g. a Gaussian limiting distribution under the CLT depends only on the first two moments of F, which can usually be estimated consistently. I.e. taking limits is a strategy to eliminate other nuisance parameters from the sampling distribution of T_n.

The bootstrap is an alternative approach which relies on *resampling* from the original data.

2 The Bootstrap Principle

The *bootstrap principle* consists in estimating $G_n(t, F_0)$ by $G_n(t, \hat{F}_n)$, where $\hat{F}_n$ is a suitable estimator for F_0.

There are two common strategies for estimating F:

- The *parametric bootstrap* assumes a parametric model for the distribution

  $$F_0(z) = F(z; \theta_0)$$

  where the c.d.f. of z is known up to a finite-dimensional parameter $\theta \in \mathbb{R}^k$. Here we can find an estimator $\hat{\theta}$ for θ_0 and evaluate

  $$\hat{G}_n^{par}(t, F) := G_n(t, F(\cdot; \hat{\theta})$$

- The *nonparametric bootstrap* estimates the c.d.f. F by the empirical c.d.f.

  $$\hat{F}_n^{emp}(z) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{z_i \leq z\}$$

  and evaluate the distribution

  $$\hat{G}_n^{nonp}(t, F) := G_n(t, \hat{F}_n^{emp})$$

The justification of the parametric bootstrap typically comes from consistency of the estimator $\hat{\theta}$ together with some continuity assumptions. For the nonparametric bootstrap, notice that by a standard law of large numbers,

$$\hat{F}_n^{emp}(z) \xrightarrow{p} \mathbb{E}_{F_0}[\mathbb{1}\{z_i \leq z\}] \equiv F_0(z)$$

pointwise for each value of z. The Glivenko-Cantelli establishes the stronger conclusion that convergence of the empirical distribution is in fact uniform,

$$\sup_z |\hat{F}_n^{emp}(z) - F_0(z)| \to 0$$

3 Monte Carlo Algorithm for the Bootstrap

It is in most cases impractical to evaluate $G_n(t, \hat{F}_n)$ analytically. Instead, it is common practice to obtain percentiles or moments by simulating from that distribution. A generic algorithm for implementing the nonparametric bootstrap via Monte Carlo simulation is as follows:

1. For the bth bootstrap sample, generate n i.i.d. draws $z_{1,b}^*, \dots, z_{n,b}^*$ from the estimate $\hat{F}$ for the distribution.

2. Evaluate $T_{n,b}^* := T_n(z_{1,b}^*, \dots, z_{n,b}^*)$.

3. Repeat the first two steps B times to obtain B bootstrap draws of the statistic, $T_{n,1}^*, \dots, T_{n,B}^*$.

4. Approximate the distribution of T_n with

$$\hat{G}_n(t, \hat{F}) := \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{T_{n,b}^* \leq t\}$$

In the last step, we can make the simulation error in approximating $G_n(t, \hat{F})$ small by choosing B sufficiently large. In practice $B = 400$ or $B = 1000$ is often sufficient. This procedure is easy to parallelize, where each bootstrap replication is typically not very costly.

For the nonparametric bootstrap,

- the estimator $\hat{F}_n$ is the empirical distribution of the sample, i.e. a uniform discrete distribution with n mass points.

- Drawing $z_{1,b}^*, \dots, z_{n,b}^*$ from the empirical distribution therefore amounts to drawing from the sample $\{z_1, \dots, z_n\}$ uniformly at random and *with replacement.*

- In principle we could enumerate all n^n (equally likely) possibilities of distinct samples of size n to obtain the exact form of $G_n(t, \hat{F}_n^{emp})$, but this is obviously infeasible for typical sample sizes.

When we are interested in moments of the distribution of T_n, we can also directly simulate these using the same algorithm. E.g. to estimate the variance of T_n, we compute

$$\widehat{\mathrm{Var}}(T_n^*) := \frac{1}{B-1} \sum_{b=1}^{B} \left(T_{n,b}^* - \frac{1}{B} \sum_{r=1}^{B} T_{n,r}^* \right)^2$$

4 Applications

4.1 Critical Values for Tests

Suppose we want to conduct a hypothesis test based on the statistic

$$T_n := T_n(z_1, \dots, z_n)$$

Supposed also that the decision rule for that test is to reject the null hypothesis if $T_n > c$ (we can easily adapt the logic of what follows to two-sided tests), where c is the critical value.

Hence for a nominal confidence size $\alpha \in (0, 1)$, we need to determine a value of c such that under the null hypothesis,

$$P(T_n \leq c) = 1 - \alpha$$

That is, the exact critical value $c \equiv c(1-\alpha)$ would be the $(1-\alpha)$ quantile of the null distribution of T_n. That distribution is usually not known, but we can use the bootstrap to estimate $c(1-\alpha) = q_n(1-\alpha) := G_n^{-1}(1-\alpha; F)$.

One caveat for implementing the bootstrap is that we need to ensure that the null hypothesis in fact holds for the estimated distribution $\hat{F}_n$. This may require recentering of residuals or moment function before drawing from the sample, and then drawing from an appropriately adjusted estimate $\tilde{F}_n$ of the distribution. For least-squares regression, residuals are mechanically uncorrelated with regressors and the constant, so such a recentering is not necessary in that particular case.

We then proceed as before, generating B bootstrap replications $T^*_{n,1}, \ldots, T^*_{n,B}$, and estimate c by the $(1-\alpha)$ quantile of the bootstrap c.d.f.

$$\hat{c}_n(1-\alpha) := \hat{G}_n^{-1}(1-\alpha, \tilde{F})$$

This amounts to choosing the critical value $\hat{c}_n$ equal to the $(1-\alpha)B$ lowest value among the B draws $T^*_{n,b}$.

4.2 Confidence Intervals

Construction of a confidence interval for a scalar parameter θ_0 at confidence size $1-\alpha$ also requires finding critical values corresponding to the lower and upper bounds of that interval. It is most common to construct such a confidence interval by inverting a t-test based on the studentized estimator,

$$T_n(\theta) := \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n(\hat{\theta}_n)}$$

where $\hat{\sigma}_n(\hat{\theta}_n)$ is a short-hand for the (estimated) standard error of the estimator $\hat{\theta}_n$.

For an equal-tailed confidence interval we estimate the $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of $T_n(\theta_0)$, denoted by $q_n\left(\frac{\alpha}{2}\right)$ and $q_n\left(1-\frac{\alpha}{2}\right)$. We can then construct a confidence interval by inverting a two-sided t-test with critical values $q_n\left(\frac{\alpha}{2}\right)$ and $q_n\left(1-\frac{\alpha}{2}\right)$,

$$\begin{aligned}
CI &:= \left\{\theta \in \mathbb{R} : q_n\left(\frac{\alpha}{2}\right) \le T_n(\theta) \le q_n\left(1-\frac{\alpha}{2}\right)\right\} \\
&= \left\{\theta \in \mathbb{R} : q_n\left(\frac{\alpha}{2}\right) \le \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n(\hat{\theta}_n)} \le q_n\left(1-\frac{\alpha}{2}\right)\right\} \\
&= \left\{\theta \in \mathbb{R} : \hat{\theta}_n - \hat{\sigma}_n(\hat{\theta}_n) q_n\left(1-\frac{\alpha}{2}\right) \le \theta \le \hat{\theta} - \hat{\sigma}_n(\hat{\theta}_n) q_n\left(\frac{\alpha}{2}\right)\right\} \\
&= \left[\hat{\theta}_n - \hat{\sigma}_n(\hat{\theta}_n) q_n\left(1-\frac{\alpha}{2}\right), \hat{\theta} - \hat{\sigma}_n(\hat{\theta}_n) q_n\left(\frac{\alpha}{2}\right)\right]
\end{aligned}$$

Note that in general the distribution of $T_n(\theta_0)$ need not be symmetric about zero, so that the *lower* bound of the CI is determined by the *upper* tail of the distribution of $T_n(\theta_0)$, and the *upper* bound of the CI is determined by the *lower* tail of that distribution.

By construction,

$$\mathbb{P}(\theta_0 \in CI) = 1 - \mathbb{P}(\max_{\hat{\theta} \in CI} \hat{\theta} < \theta_0) - \mathbb{P}(\min_{\hat{\theta} \in CI} \hat{\theta} > \theta_0) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

As for hypothesis tests, we can then again replace the unknown quantiles of T_n with bootstrap estimates, where we generate bootstrap samples

$$T^*_{n,b} := \frac{\hat{\theta}^*_{n,b} - \hat{\theta}_n}{\hat{\sigma}^*_n(\hat{\theta}_n)}$$

and we also re-estimate the standard error of the estimator for each bootstrap sample $z^*_{1,b}, \dots, z^*_{n,b}$. We then obtain the bootstrap estimate of the c.d.f. of $T_n(\theta_0$,

$$\hat{G}_n(t, \tilde{F}) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{T^*_{n,b} \le t\}$$

from which we can obtain the τ quantile as

$$\hat{q}_n(\tau) := \hat{G}_n^{-1}(\tau, \hat{F}_n)$$

for $\tau = \frac{\alpha}{2}, 1 - \frac{\alpha}{2}$.

Alternatively it is also possible to obtain confidence intervals using the distribution of the unstudentized estimator, which turns out to have inferior statistical properties to confidence intervals based on the t-ratio. Also it is possible to construct symmetric confidence interval by inverting a one-sided test based on the absolute value of the t-ratio, where for an appropriately chosen critical value c the confidence interval includes all values of θ such that $|T_n(\theta)| \le c$.

4.3 Bias Correction

Suppose we have an estimator

$$\hat{\theta}_n := \hat{\theta}_n(z_1, \ldots, z_n)$$

which may be a nonlinear function of the data, and therefore potentially biased in finite sample even when it is consistent at the parametric root-n rate. We can use the bootstrap to estimate the bias

$$B_n := \mathbb{E}_{F_0}[\hat{\theta}_n - \theta_0]$$

that is, we are interested in estimating the expectation of the statistic

$$T_n := \hat{\theta}_n(z_1, \ldots, z_n) - \theta_0$$

under the distribution $G_n(t, F_0)$.

For the bootstrap counterpart, the estimator $\hat{\theta}_n$ is the sample analog for the population quantity θ_0. That is if we think of $\theta_0 = \theta(F_0)$ as a functional of the unknown population distribution, the bootstrap replaces it with $\hat{\theta}_n = \theta(\hat{F}_n)$, that same functional evaluated at the empirical distribution.

We then simulate the distribution of the bootstrap statistic

$$T^*_{n,b} := \hat{\theta}_n(z^*_{1,b}, \ldots, z^*_{n,b}) - \hat{\theta}_n =: \hat{\theta}^*_{n,b} - \hat{\theta}_n$$

which we denote by $G_n(t, \hat{F}_n)$.

The simulated bootstrap estimate of the bias given B replications is

$$\hat{B}_n := \frac{1}{B}\sum_{b=1}^{B} T^*_{n,b} = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^*_{n,b} - \hat{\theta}_n$$

We can then bias-correct the initial estimator by subtracting off the bootstrap estimate of the bias,

$$\hat{\theta}^{BC}_n := \hat{\theta}_n - \hat{B}_n = 2\hat{\theta}_n - \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^*_{n,b}$$

Under regularity conditions for parameteric problems, this bias correction removes estimator bias of the order up to n^{-1}, so that in particular in small samples, the bias-corrected estimator may have substantially lower bias than the standard estimator $\hat{\theta}_n$.

4.4 Variance Estimation

We often have an analytical formula for the asymptotic variance, but it may sometimes be tedious to implement or code up. An alternative is to use the bootstrap, where

$$T_n := \hat{\theta}_n, \quad T^*_{n,b} := \hat{\theta}^*_b = \hat{\theta}_n(z^*_{1,b}, \ldots, z^*_{n,b})$$

For a root-n consistent estimator, the asymptotic variance is

$$V := \mathbb{E}_{F_0}\left[n(T_n - \mathbb{E}[T_n])^2\right]$$

where the expectation is taken with respect to the distribution $G_n(t, F_0)$ The bootstrap analog of the variance is

$$V_n^* := \mathbb{E}_{\hat{F}_n}\left[n(T_{n,b}^* - \mathbb{E}_{\hat{F}_n}[T_{n,b}^*])^2\right]$$

where the expectation is taken with respect to the bootstrap distribution $G_n(t, \hat{F}_n)$.

We can then estimate the asymptotic variance using the simulated bootstrap estimate

$$\hat{V}_n^* := \frac{1}{B}\sum_{b=1}^{B} n\left(T_{n,b}^* - \frac{1}{B}\sum_{r=1}^{B} T_{n,r}^*\right)^2$$

5 Bootstrap Consistency

To justify the validity of the bootstrap, we generally have to rely again on asymptotic (large-sample) arguments.

Definition 1. *The bootstrap is* ***consistent*** *if for each $\varepsilon > 0$ and population distribution F_0,*

$$\lim_n \mathbb{P}_{F_0}\left(\sup_t |G_n(t, \hat{F}_n) - G_n(t, F_0)| > \varepsilon\right) = 0$$

The main idea for a consistency proof for the bootstrap is as follows: For large n, the distribution of T_n under F converges to some limiting distribution

$$G_\infty(t, F) := \lim_n G_n(t, F)$$

Hence, we may want to show that the following approximations hold,

$$G_n(t, \hat{F}_n) \approx G_\infty(t, \hat{F}_n) \approx G_\infty(t, F_0) \approx G_n(t, F_0)$$

where the first and the third approximation rely on some limiting argument for the asymptotic distribution of T_n, and the second approximation requires that $\hat{F}_n$ is a consistent estimator for F_0 (in an appropriate sense), and that the limiting distribution $G_\infty(t, F)$ varies continuously in F (again in an appropriate sense).

Specifically we need the following three requirements:

1. $\varrho(\hat{F}_n, F_0) \xrightarrow{p} 0$ for some metric $\varrho(\cdot)$.

2. $G_\infty(t, F)$ is continuous in t for each F, and

3. for any t and sequence $F_1, F_2, \ldots$ such that $\varrho(F_n, F_0) \to 0$, we have $|G_\infty(t, F_n) - G_\infty(t, F_0)| \to 0$.

Under these conditions, the bootstrap is consistent:

Theorem 1. *(Beran and Ducharme) If $G_\infty(t, F) := \lim_n G_n(t, F)$ exists and (1)-(3) hold, the bootstrap is consistent.*

Proof: We can see immediately that (1) and (3) together imply that $|G_\infty(t, \hat{F}_n) - G_\infty(t, F_0)$ converges to zero in probability. Hence, by the definition of $G_\infty(t, F)$ as the limit of $G_n(t, F)$, it follows that at each $t \in \mathbb{R}$,

$$|G_n(t, \hat{F}_n) - G_n(t, F_0)| \xrightarrow{p} 0$$

Finally, Polya's Lemma implies that pointwise convergence of a sequence of monotone functions to a bounded monotone continuous functions implies uniform convergence, establishing the claim □

This theorem is a generic consistency result for the bootstrap, where the high-level conditions (1)-(3) are fairly abstract and not always straightforward to verify. For special cases, results establishing bootstrap consistency from more primitive conditions are available. However if any of the conditions (1)-(3) doesn't hold, we can find counterexamples in which the nonparametric bootstrap is known to fail, see also examples below.

5.1 Refinements

One theoretical reason for choosing the bootstrap over standard asymptotic inference even when both are known to work is that under certain conditions the bootstrap is known to provide a more precise estimate of the asymptotic distribution of T_n. Specifically, the bootstrap may achieve refinements in approximating $G_n(t, F_0)$ when the statistic T_n is *asymptotically pivotal*:

Definition 2. *The statistic T_n is* ***asymptotically pivotal*** *if its limiting distribution does not depend on F, that is there exists $G_\infty(t)$ such that*

$$\lim_n G_n(t, F) = G_\infty(t) \quad \text{for all } F$$

To give a few examples,

- Under regularity conditions, the LS coefficient

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V),$$

where the asymptotic variance $V \equiv V(F_0)$ depends on the population distribution F_0 of the data. Hence the LS coefficient is not asymptotically pivotal.

- However, the studentized regression coefficient

$$\sqrt{n}V^{-1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I_k)$$

for any population distribution F_0. In particular, the t-ratio for any component of $\hat{\beta}$ is asymptotically pivotal.

- The F-statistic

$$F_n := (n-q)\frac{\hat{\mathbf{e}}_R'\hat{\mathbf{e}}_R - \hat{\mathbf{e}}_U'\hat{\mathbf{e}}_U}{\hat{\mathbf{e}}_U'\hat{\mathbf{e}}_U} \xrightarrow{d} \chi^2_{(q)}$$

which does not depend on F_0. So the F-statistic is also asymptotically pivotal.

It is possible to use Edgeworth (higher-order) expansions of the c.d.f. $G_n(t, F)$ to show that if T_n is asymptotically pivotal and satisfies a CLT (and technical conditions for the Edgeworth expansions), then the bootstrap estimates the finite-sample distribution of T_n with an error

$$G_n(t, \hat{F}_n) = G_n(t, F_0) + O_P(n^{-1})$$

whereas the asymptotic approximation only satisfies

$$G_\infty(t, \hat{F}_n) = G_n(t, F_0) + O_P(n^{-1/2})$$

That is, the approximation error for the bootstrap vanishes at the faster $\frac{1}{n}$ rate, rather than only at $\frac{1}{\sqrt{n}}$, and can therefore be reasonably accurate in smaller samples.

However if T_n is not asymptotically pivotal, then we generally have

$$G_n(t, \hat{F}_n) = G_n(t, F_0) + O_P(n^{-1/2})$$

so that no refinements are obtained from the bootstrap in that case. An accessible, semi-formal explanation of this argument is given in Horowitz (2000)'s handbook chapter.

One interpretation of this result is that Gaussian asymptotic inference only accounts for estimates of the first two moments of the population distribution F_0, whereas the bootstrap also uses information in the sample about higher moments (skewness, kurtosis, etc.) for the approximation. In the non-pivotal case, any remaining nuisance parameters in the limiting distribution are still only estimated at the $n^{-1/2}$ rate, and therefore that estimation error dominates other contributions in the Edgeworth expansion.

6 Failure of the Nonparametric Bootstrap

There are a number of practically relevant cases in which the conditions for the consistency result fail for the nonparametric bootstrap in a way that renders it inconsistent:

- **Fat tailed distributions:** The nonparametric bootstrap is inconsistent for sample averages of random variables for which a central limit theorem fails, so that the limiting distribution of the sample average is not Gaussian. The reason for the failure of the bootstrap is that the empirical distribution for the sample $z_1, \dots, z_n$ supports at most n distinct values and has therefore necessarily a bounded support. Hence the average of the bootstrap sample always satisfies a CLT, so that the bootstrap estimate of the distribution is always Gaussian, and therefore different from the limit of the sampling distribution, see also Mammen (1992b) for formal results.

- **Extrema:** The nonparametric bootstrap usually fails for approximating extreme order statistics of the sample, e.x. the maximum $T_n := \max\{z_1, \dots, z_n\}$. Here the failure stems from approximating the possibly continuous population distribution of Z_i with the discrete empirical distribution. This introduces the possibility of ties where the maximum of the bootstrap sample, $T^*_{n,b} := \max\{z^*_{1,b}, \dots, z^*_{n,b}\}$ may coincide with the sample maximum $T_n := \max\{z_1, \dots, z_n\}$ with probability converging to a strictly positive limit. This generates a point mass of the bootstrap distribution at T_n, whereas the distribution of the sample maximum for a continuous distribution of Z_i possesses no such mass points. See a discussion of this example in Horowitz (2000).

- **Non-differentiable functions of sample means:** We may often be interested in approximating the distribution of a statistic $T_n := h(\bar{z}_n)$ is a function of the sample mean $\bar{z}_n := \frac{1}{n}\sum_{i=1}^n z_i$. By the delta-rule, the bootstrap for $\sqrt{n}(h(\bar{z}_n) - h(\mathbb{E}[z_i])$ works under regularity conditions if $h : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable, however the bootstrap is generally inconsistent when the function is not differentiable, e.g. $h(z) := \max\{0, z\}$ or $h(z) := |z|$. The main reason is that the asymptotic distribution changes discontinuously for values of $\mathbb{E}[z_i]$ at which the first derivative of $h(z)$ is discontinuous or undefined. Since the bootstrap centers the distribution of the sample mean at $\bar{z}_n$ rather than $\mathbb{E}[z_i]$, it fails to converge to a well-defined limit when the population mean is at such a discontinuity. For a formal description of the problem, see Andrews (2000) and Fang and Santos (2015).

Broadly speaking, these examples highlight two deeper reasons why the nonparametric bootstrap may fail: for one it is approximating a population distribution that may be continuous and have unbounded support with a discrete distribution, resulting in bounded moments or ties for the bootstrap when those should not exist under the population distribution. The second problem comes from nuisance parameters in the asymptotic (sampling) distribution that can't be estimated consistently by any method, in particular not by the bootstrap.

Note that these examples concern only the nonparametric bootstrap; in some of these cases, the parametric bootstrap may still be consistent under the appropriate auxiliary assumptions. Some of these problems of the nonparametric bootstrap can in principle be resolved by m out of n bootstrap or subsampling (see Politis, Romano, and Wolf (1999)), however these methods are usually only pointwise consistent and may be substantially less precise, so the nonparametric bootstrap should be preferred in situations in which it is known to work.

7 Bootstrap for Regression Analysis

When we apply the bootstrap principle to regression, we may have to modify some aspect of its implementation. Consider the linear regression model (LRM)

$$y_i = x_i'\beta + e_i, \quad \mathbb{E}[e_i|x_i] = 0$$

The discussion so far treated $z_i = (y_i, x_i')'$ as jointly random. However for regression, we are typically interested in inference *conditional* on regressors (fixed-design approach). That is we want to treat $x_1, \ldots, x_n$ as fixed, and regard $y_1, \ldots, y_n$ as draws from the conditional distribution of y_i given x_i.

7.1 Residual Bootstrap

In a first approach, we consider the **homoskedastic** LRM which imposes the stronger assumption that e_i is independent of x_i, that is

$$y_i = x_i'\beta + e_i, \qquad e_i|x_i \overset{iid}{\sim} F_e$$

where the distribution $F_e(e)$ has mean zero.

The **residual bootstrap** estimates β via least squares and obtains LS residuals

$$\hat{e}_i := y_i - x_i'\hat{\beta}_{LS}$$

We then estimate F_e with the empirical distribution of $\hat{e}_i$,

$$\hat{F}_e(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{e}_i \leq t\}$$

We then generate the bth bootstrap samples $\left\{(x_{1,b}^*, y_{1,b}^*), \ldots, (x_{n,b}^*, y_{n,b}^*)\right\}$ where we hold fixed $x_{i,b}^* := x_i$, and draw $\hat{e}_{i,b}^*$ independently and uniformly at random from $\hat{e}_1, \ldots, \hat{e}_n$ (with replacement) to form

$$y_{i,b}^* := x_i'\hat{\beta} + \hat{e}_{i,b}^*$$

We then obtain the statistic of interest $T^*_{n,b} := T_n((x^*_{1,b}, y^*_{1,b}), \ldots, (x^*_{n,b}, y^*_{n,b}))$ for the bootstrap sample. Given B bootstrap replications, we then estimate the distribution of the statistic with

$$\hat{G}_n(t, \hat{F}_n) := \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{T^*_{n,b} \leq t\}$$

The main problem with this approach is that independence of e_i and x_i is a strong assumption, in particular we have to assume that residuals are homoskedastic.

7.2 Wild Bootstrap

The Wild bootstrap provides an alternative for implementing the bootstrap and does not require full independence of e_i and x_i. We do maintain the LRM

$$y_i = x_i'\beta + e_i, \quad \mathbb{E}[e_i|x_i] = 0$$

In the absence of a parametric model, it would obviously be challenging to estimate the full conditional distribution $F_{e|x}(e|x)$ of e_i given x_i consistently. The Wild Bootstrap (Liu (1988), Mammen (1992a)) builds on the insight that it is generally sufficient to match the first three conditional moments of $F_{e|x}(e|x)$. We can therefore work with a simpler distribution that matches

$$\begin{aligned} \mathbb{E}[e^*_{i,b}|x_i] &= 0 \\ \mathbb{E}[(e^*_{i,b})^2|x_i] &= \hat{e}_i^2 \\ \mathbb{E}[(e^*_{i,b})^3|x_i] &= \hat{e}_i^3 \end{aligned}$$

We can implement this by forming

$$y^*_{i,b} := x_i'\hat{\beta} + e^*_{i,b}$$

with residuals $e^*_{i,b} := v_i\hat{e}_i$, where the multiplier v_i is a random variable with $\mathbb{E}[v_i] = 0$ and $\mathbb{E}[v_i^2] = \mathbb{E}[v_i^3] = 1$. A popular choice suggested by Mammen (1993) is the two-point distribution

$$P(v_i = v) = \begin{cases} \frac{\sqrt{5}-1}{2\sqrt{5}} & \text{if } v = \frac{1+\sqrt{5}}{2} \\ \frac{\sqrt{5}+1}{2\sqrt{5}} & \text{if } v = \frac{1-\sqrt{5}}{2} \\ 0 & \text{otherwise} \end{cases}$$

Otherwise, the Wild Bootstrap is implemented in a manner that is completely parallel to the residual bootstrap. The Wild bootstrap is consistent for the (potentially heteroskedastic) LRM and also achieves refinements for pivotal, asymptotically normal statistics.

8 Bootstrap with Dependent Data

So far we have only discussed the bootstrap for independent observations. When the data is dependent, knowledge of the marginal distribution of z_i is no longer sufficient to obtain the distribution of a sample statistic $T_n = T_n(z_1, \dots, z_n)$. Rather we need to estimate all relevant features of the *joint* distribution of the sample $z_1, \dots, z_n$. This generally requires that we make some assumptions on the structure of dependence. Without going into too much detail, we will just mention a few popular approaches:

- **Cluster-dependent data:** Suppose the observations can be grouped into G groups of respective sizes $n_1, \dots, n_G$, $n = \sum_{g=1}^{G} n_g$ such that whenever $g \neq g'$, observations z_{gi} and $z_{g'j}$ are independent for each $i = 1, \dots, n_g$ and $j = 1, \dots, n_{g'}$. Then we say that the observations exhibit cluster-dependence, where we do not restrict the dependence between z_{gi} and z_{gj} for $i, j = 1, \dots, n_g$.

  In that case, we can generate bootstrap samples by drawing clusters from $\{z_{11}, \dots, z_{1n_1}\}$, $\dots, \{z_{G1}, \dots, z_{Gn_G}\}$ at random and with replacement, and otherwise proceed as in the i.i.d. case. If G is sufficiently large, and cluster size n_g does not vary too much across clusters, this bootstrap inherits most of the desirable features of the i.i.d. bootstrap.

- **Stationary dependence:** When the sequence $z_1, \dots, z_n$ exhibits time-series dependence that is stationary and ergodic, pairs of observations that are a sufficiently large number of lags apart are close to independent. As an approximation, we can therefore restrict our attention to the joint distribution within "blocks" of a finite number, say m, adjacent units $z_j, z_{j+1}, \dots, z_{j+m-1}$. The bootstrap can then draw blocks at random to obtain bootstrap samples that exhibit a similar dependence structure as the original series, where the block lenght m should be wide enough to include all observations that exhibit meaningful dependence with units at the center of that block, but also short enough relative to n so that the original series can produce a sufficiently large number of - nearly independent - blocks of that length. This approach is known as the *block bootstrap,* see H.Künsch (1989), Hall and Horowitz (1996) for details.

- **Markov processes:** In the time series context, the block bootstrap may not work very well when n is not very large, or dependence in the series is strong. However if the series is Markovian, it can be easier to estimate its law of motion (transition matrix) rather than the distribution of sufficiently wide blocks. Given the estimated law of motion, we can generate bootstrap samples from the resulting Markov process, see Horowitz (2003) for details and formal results.

References

Andrews, D. (2000): "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68(2), 399–405.

Fang, Z., and A. Santos (2015): "Inference on Directionallly Differentiable Functions," working paper, Kansas State and UCLA.

Hall, P., and J. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64(4), 891–916.

H.Künsch (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, 17, 1217–1261.

Horowitz, J. (2000): "The Bootstrap," Handbook of Econometrics, Vol V Chapter 52.

——— (2003): "Bootstrap Methods for Markov Processes," *Econometrica*, 71(4), 1049–1082.

Liu, R. (1988): "Bootstrap Procedures Under Some Non-i.i.d. Models," *Annals of Statistics*, 16(4), 1696–1708.

Mammen, E. (1992a): "Higher Order Accuracy of the Bootstrap Method for Smooth Functionals," *Scandinavian Journal of Statistics*, 19, 255–269.

——— (1992b): *When does the Bootstrap Work: Asymptotic Results and Simulations*, vol. 77 of *Lecture Notes in Statistics*. Springer, Berlin.

——— (1993): "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics*, pp. 255—-285.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*. Springer, New York.