

Stat 451 Lecture Notes 08¹²

Bootstrap

Ryan Martin
UIC

www.math.uic.edu/~rgmartin

¹Based on Chapter 9 in Givens & Hoeting, Chapter 24 in Lange

²Updated: April 4, 2016

- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure
- 7 Further remarks

- For hypothesis testing and confidence intervals, there is a “statistic” whose sampling distribution is required.
- For example, to test if $H_0 : \mu = \mu_0$ based on a sample from a $N(\mu, \sigma^2)$ population, we use the t-statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

whose null distribution (under stated conditions) is Student-t.

- But almost any deviation from this basic setup leads to a tremendously difficult distributional calculation.
- The goal of the *bootstrap* is to give a simple approximate solution based on simulations.

- For a distribution with cdf F , suppose we are interested in a parameter $\theta = \varphi(F)$, written as a functional of F .
- Examples:
 - Mean: $\varphi(F) = \int x dF(x)$;
 - Median: $\varphi(F) = \inf\{x : F(x) \geq 0.5\}$;
 - ...
- Given data $\mathcal{X} = \{X_1, \dots, X_n\}$ from F , the *empirical cdf* is

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}, \quad x \in \mathbb{R}.$$

- Then a natural estimate of θ is $\hat{\theta} = \varphi(\hat{F})$, the same functional of the empirical cdf.

- For inference, some statistic $T(\mathcal{X}, F)$ is used; e.g.,

$$T(\mathcal{X}, F) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

- Again, the sampling distribution of $T(\mathcal{X}, F)$ may be very complicated, unknown, or could depend on unknown F .
- *Bootstrap Idea:*
 - Replace the unknown cdf F with the empirical cdf $\hat{F}$.
 - Produce a numerical approximation of the sampling distribution of $T(\mathcal{X}, F)$ by repeated sampling from $\hat{F}$.

Notation (cont.)

- Let $\mathcal{X}^* = \{X_1^*, \dots, X_n^*\}$ be an iid sample from $\hat{F}$, i.e., a sample of size n (with replacement) from $\mathcal{X}$.
- Given $\mathcal{X}^*$, the statistic $T^* = T(\mathcal{X}^*, \hat{F})$ can be evaluated.
- Repeated sampling of $\mathcal{X}^*$ gives a sequence of T^* 's which can be used to approximate the distribution of $T(\mathcal{X}, F)$.
- For example, $\text{V}\{T(\mathcal{X}, F)\} \approx \text{Var}\{T_1^*, \dots, T_B^*\}$.
- *Why should bootstrap work?*
 - Glivenko–Cantelli theorem says that $\hat{F} \rightarrow F$ as $n \rightarrow \infty$.
 - So, iid sampling from $\hat{F}$ should be approximately the same as iid sampling from F when n is large.³

³Lots of difficult theoretical work has been done to determine what it means for this approximation to be good and in what kinds of problems does it fail.

- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure
- 7 Further remarks

- Above procedure is essentially the *nonparametric bootstrap*.
- Sampling distribution of $T(\mathcal{X}, F)$ is approximated directly by the empirical distribution of the bootstrap sample $T_1^*, \dots, T_B^*$.
- For example:
 - Quantiles of $T(\mathcal{X}, F)$ are approximated by sample quantiles of $T_1^*, \dots, T_B^*$.
 - Variance of $T(\mathcal{X}, F)$ is approximated by the sample variance of $T_1^*, \dots, T_B^*$.
- Bootstrap sample size is usually rather large, e.g., $B \sim 1000$, but computationally manageable.

Example: variance of a sample median

- Example 29.4 in DasGupta (2008).
- $X_1, \dots, X_n$ iid Cauchy with median μ .
- The sample mean $\bar{X}$ is a bad estimator of μ (*why?*), so use the sample median M_n instead.
- For odd n , say $n = 2k + 1$, there is an exact formula:

$$V(M_n) = \frac{2n!}{(k!)^2 \pi^n} \int_0^{\pi/2} x^k (\pi - x)^k (\cot x)^2 dx.$$

- A CLT-type of asymptotic approximation is

$$\widehat{V(M_n)} = \pi^2 / 4n.$$

- What about a bootstrap approximation?

Example: variance of sample median (cont.)

- With $n = 21$ we have

$$V(M_n) = 0.1367 \quad \text{and} \quad \widehat{V(M_n)} = 0.1175.$$

- A bootstrap⁴ estimate of $V(M_n)$ using $B = 5000$ is

$$\widehat{V(M_n)}_{\text{boot}} = 0.1102.$$

- Slight under-estimate of the variance...
- Main point is that we got a pretty good answer with essentially no effort — computer does all the hard work.

⁴Note that I used `set.seed(77)` in the code...

- What does it mean for the bootstrap to “work?”
 - $H_n(x)$ is the true distribution function for $\hat{\theta}_n$.
 - $H_n^*(x)$ is the true distribution function for $\hat{\theta}_n^*$.
 - Bootstrap is “consistent” if the *distance* between $H_n(x)$ and $H_n^*(x)$ converges to 0 (in probability) as $n \rightarrow \infty$.
- The bootstrap is successful in many problems, but there are known situations when it may fail:
 - support depends on parameter;
 - true parameter sits on boundary of parameter space;
 - estimator convergence rate $\neq n^{-1/2}$;
 - ...
- The bootstrap can detect skewness in the distribution of $\hat{\theta}_n$ while CLT-type of approximations will not — often has a “second-order accuracy” property.
- Bootstrap often underestimate variances.

Bootstrap confidence intervals (CIs)

- A primary application of bootstrap is to construct CIs.
- The simplest approach is the *percentile method*.
- Let $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ be a bootstrap sample of point estimators.
- A two-sided $100(1 - \alpha)\%$ bootstrap percentile CI is

$$[\xi_{\alpha/2}^*, \xi_{1-\alpha/2}^*],$$

where ξ_p^* is the $100p$ percentile in the bootstrap sample.

- Simple and intuitive, but there are “better” methods.

- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap**
- 4 Bootstrap in regression
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure
- 7 Further remarks

- Parametric bootstrap is a variation on the standard (nonparametric) bootstrap discussed previously.
- Let $F = F_\theta$ be a parametric model.
- The parametric bootstrap replaces sampling iid from $\hat{F}$ with sampling iid from $F_{\hat{\theta}}$, where $\hat{\theta}$ is some estimator of θ .
- Potentially more complicated than nonparametric bootstrap because sampling from $F_{\hat{\theta}}$ might be more difficult than sampling from $\hat{F}$.

Example: variance of sample median

- $X_1, \dots, X_n$ iid Cauchy with median μ .
- Write M_n for sample median.
- Parametric bootstrap samples $X_1^*, \dots, X_n^*$ from a Cauchy distribution with median M_n .
- Using $B = 5000$, parametric bootstrap gives

$$\widehat{V(M_n)}_{\text{p-boot}} = 0.1356.$$

- A bit closer to the true variance, $V(M_n) = 0.1367$, compared to nonparametric bootstrap.

Example: random effect model

- Hierarchical model:

$$(\mu_1, \dots, \mu_n) \stackrel{\text{iid}}{\sim} N(\lambda, \psi^2)$$
$$Y_i \mid \mu_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, \dots, n.$$

- Parameters (λ, ψ) unknown but σ_i 's known.
- Parameter of interest is $\psi \geq 0$, and values $\psi \approx 0$ are of interest because it suggests homogeneity.
- Non-hierarchical version: $Y_i \stackrel{\text{ind}}{\sim} N(\lambda, \sigma_i^2 + \psi^2)$, $i = 1, \dots, n$.
- Can estimate ψ via maximum likelihood.
- Use parametric bootstrap to get confidence intervals?

Example: random effect model (cont.)

- Want to see what happens when $\psi \approx 0$.
- Take $\psi = n^{-1/2}$, near the boundary of $\psi \geq 0$.
- Two-sided 95% parametric bootstrap percentile intervals have pretty low coverage in this case, even for large n .
- It is possible to get intervals with *exact* coverage...

n	Coverage	Length
50	0.758	0.183
100	0.767	0.138
250	0.795	0.079
500	0.874	0.039

- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression**
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure
- 7 Further remarks

- Consider an *observational study* where pairs $z_i = (x_i^\top, y_i)^\top$ are sampled from a joint predictor-response distribution.
- Let $\mathcal{Z} = \{z_1, \dots, z_n\}$.
- Following the basic bootstrap principle above, repeatedly sample $\mathcal{Z}^* = \{z_1^*, \dots, z_n^*\}$ with replacement from $\mathcal{Z}$.
- Then do the same approximation of sampling distributions based on the empirical distribution from the bootstrap sample.
- This is called the *paired bootstrap*.
- What about for a *fixed* design?
 - Complication is that the y_i 's are not iid.
 - In such cases, first resample the residuals $e_i = y_i - \hat{y}_i$ from the original LS fit, and then set $y_i^* = x_i^\top \hat{\beta} + e_i^*$.

Example: ratio of slope coefficients

- Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

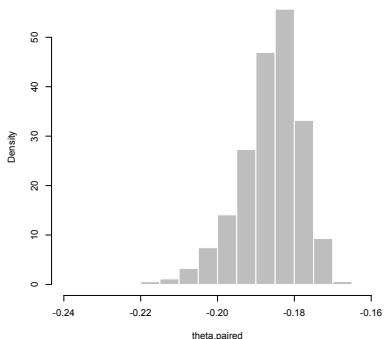
where $\varepsilon_1, \dots, \varepsilon_n$ are iid mean zero, not necessarily normal.

- Assume this is an *observational study*.
- Suppose the parameter of interest is $\theta = \beta_1/\beta_0$.
- A natural estimate of θ is $\hat{\theta} = \hat{\beta}_1/\hat{\beta}_0$.
- To get the (paired) bootstrap distribution of $\hat{\theta}$:
 - Sample $\mathcal{Z}^* = \{z_1^*, \dots, z_n^*\}$ with replacement from $\mathcal{Z}$.
 - Fit the regression model with data $\mathcal{Z}^*$ to obtain $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.
 - Evaluate $\hat{\theta}^* = \hat{\beta}_1^*/\hat{\beta}_0^*$.

Example: ratio of slope coefficients (cont.)

- Figure below shows bootstrap distribution of $\hat{\theta} = \hat{\beta}_1 / \hat{\beta}_0$.
- 95% bootstrap percentile confidence interval for θ is:

$$(-0.205, -0.173).$$



- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression
- 5 Better bootstrap CIs**
- 6 Remedies for bootstrap failure
- 7 Further remarks

Confidence intervals (CIs)

- Consider the problem of constructing a $100(1 - \alpha)\%$ CI for a parameter θ .
- When the parameter is a mean μ of a normal population, then the t CI, given below, is often used.

$$\bar{X} \pm t_{\alpha/2, n-1} \cdot S/\sqrt{n}.$$

- Since, in this case, $(\bar{X} - \mu)/(S/\sqrt{n})$ has an exact Student-t distribution, $\pm t_{\alpha/2, n-1} \cdot S/\sqrt{n}$ are appropriate quantiles of the sampling distribution of $\bar{X} - \mu$.
- There are a variety of approaches for this:
 - We discussed the *bootstrap percentile* method before.
 - Here we talk about *bootstrap t*.

- Again, $\hat{\theta}_n$ is a given estimate of θ .
- Suppose $\hat{\sigma}_n$ is an estimate of the standard deviation of $\hat{\theta}_n$:
 - $\hat{\sigma}_n = S/\sqrt{n}$ in the normal mean problem.
 - Can often use delta theorem approximation for $\hat{\sigma}_n$.
- Let $T_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$ be the “t-statistic” and

$$T_n^* = (\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^*$$

its bootstrap counterpart.

- Then the bootstrap t CI for θ is then given by

$$\left[\hat{\theta}_n + c_{\alpha/2}^* \hat{\sigma}_n, \hat{\theta}_n + c_{1-\alpha/2}^* \hat{\sigma}_n \right],$$

where c_p^* is the 100 p th percentile of T_n^* .

Example: regression example (cont.)

- Linear model: parameter of interest $\theta = \beta_1/\beta_0$.
- MLE is $\hat{\theta} = \hat{\beta}_1/\hat{\beta}_0$; any available approximation to the variance σ^2 of $\hat{\theta}$?
- Delta theorem gives:

$$\hat{\sigma}^2 = \hat{\theta}^2 \left(\frac{\widehat{V(\hat{\beta}_1)}}{\hat{\beta}_1^2} + \frac{\widehat{V(\hat{\beta}_0)}}{\hat{\beta}_0^2} - \frac{2\widehat{C(\hat{\beta}_0, \hat{\beta}_1)}}{\hat{\beta}_0\hat{\beta}_1} \right).$$

- Get (paired) bootstrap versions, $(\hat{\theta}^*, \hat{\sigma}^{2*})$, and compute the “t-statistics” $T^* = (\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$.
- Bootstrap CIs for θ :
 - Percentile: $[-0.205, -0.174]$;
 - t: $[-0.203, -0.167]$.

- In some cases, bootstrap percentile CIs are not so good.
- Can improve upon these by correcting the bootstrap bias.
- The *bias-corrected* (BC) bootstrap percentile CI is one idea.
- The two-sided $100(1 - \alpha)\%$ BC bootstrap CI boils down to picking different quantiles of the bootstrap distribution of $\hat{\theta}$.
 - Instead of $[\xi_{\alpha/2}^*, \xi_{1-\alpha/2}^*]$, use $[\xi_{\beta_1}^*, \xi_{\beta_2}^*]$, where β_1 and β_2 are quantities that depend user-specified constants (a, b) .
 - Text gives formulas for (β_1, β_2) and discusses choice of (a, b) .
- See G&H, Section 9.3.2.1, for details.

- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure**
- 7 Further remarks

Example: bootstrap failure

- Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$.
- Consider $T_n = n(\theta - \hat{\theta}_n)$, where $\hat{\theta}_n = X_{(n)} = \max X_i$.
- Let $T_n^* = n(\hat{\theta}_n - \hat{\theta}_n^*)$, where $\hat{\theta}_n^* = X_{(n)}^*$.
- Claim is that the distributions of T_n^* and T_n are not close when $n \rightarrow \infty$.
- Indeed, for any $t \geq 0$,

$$P(T_n^* \leq t) \geq P(T_n^* = 0) = \dots = 1 - \left(\frac{n-1}{n}\right)^n \rightarrow 1 - e^{-1}.$$

- If $\theta = 1$, then for small t we have $P(T_n \leq t) \rightarrow 1 - e^{-t} \approx 0$, while $P(T_n^* \leq t)$ is eventually larger than $1 - e^{-1} \approx 0.63$.

Remedy: m -out-of- n bootstrap

- We have seen some cases where bootstrap fails.
- Surprisingly, one way to fix bootstrap failure is to take bootstrap samples of size $m = o(n)$ instead of size n as was advocated in the basic bootstrap.
- This is called m -out-of- n bootstrap, and is consistent in cases, such as the $\text{Unif}(0, \theta)$ model, where regular bootstrap fails.
- For details of the theory, see DasGupta Chapter 29.7–29.8 and Bickel, Gotze, and van Zwet, *Statistica Sinica*, 1997.

Example: bootstrap failure (cont.)

- Reconsider the uniform distribution problem.
- Same setup as before but, for $m = o(n)$, let

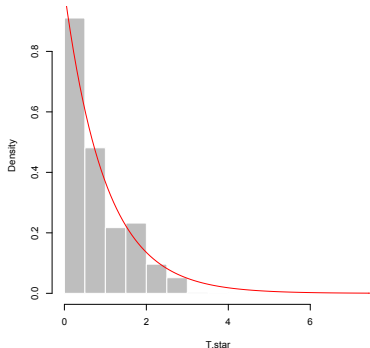
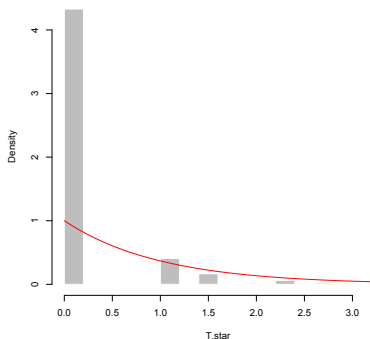
$$T_{n,m}^* = m(\hat{\theta}_n - \hat{\theta}_{n,m}^*),$$

where $\hat{\theta}_{n,m}^* = \max\{X_1^*, \dots, X_m^*\}$.

- Then theory guarantees that $P(T_{n,m}^* \leq t)$ and $P(T_n \leq t)$ are similar as $n \rightarrow \infty$, i.e., m -out-of- n bootstrap is consistent.
- Simulation details on next slide.
- *Rule of thumb:* Take $m \approx 2\sqrt{n}$...

Example – bootstrap failure (cont.)

- Same uniform problem, $n = 100$ and $m = 20$.
- Histograms of T_n^* (ordinary, left) and $T_{n,m}^*$ (modified, right) below, with true limiting distribution of T_n overlaid.



- 1 Introduction
- 2 Nonparametric bootstrap
- 3 Parametric bootstrap
- 4 Bootstrap in regression
- 5 Better bootstrap CIs
- 6 Remedies for bootstrap failure
- 7 Further remarks**

Hypothesis testing with bootstrap

- In hypothesis testing, whether it be for selecting cut-offs for rejection rules or p-values, the sampling distribution of interest is that *under the null*.
- The nonparametric bootstrap, therefore, is not appropriate.
- Simplest adjustment is to the parametric bootstrap, where one would sample from P_{θ_0} , the true distribution under the null; then this corresponds to our previously discussed Monte Carlo approximations.
- Section 9.3.3 in G&H discusses some adjustments to the nonparametric-type of bootstrap that are suitable for hypothesis testing problems.

Related idea: permutation tests

- There are some problems where the hypothesis of interest is that two groups are “identical.”
- For example, one might want to test if two groups have the same distribution.
- In this case, when the null hypothesis is true, there is nothing special about the particular grouping of the observed data.
- That is, if the observed data is $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$, under the null hypothesis that the groups are the same, all partitions of $\{X, Y\}$ into groups of size n and m , respectively, are equivalent to that observed.
- Can approximate sampling distributions of test statistics by computing them over all/some partitions of the data.

Bootstrapping dependent data

- We discussed bootstrap for iid and independent not-iid (regression) problems.
- A special feature of all of these kinds of problems is that the order in which the data is presented is not relevant.
- For dependent data problems, however, the order is important.
- This means some adjustments to the basic bootstrap strategies are needed.
- I don't really know anything about bootstrap for dependent data, but the adjustments basically boil down to working with blocks of data, rather than individual data points.
- Some details are given in Section 9.5 in G&H.

- Bootstrap is a simple and clever idea, one that allows for numerical approximation of sampling distributions without difficult analytical work.
- Be careful, however — that bootstrap is basically automatic does not mean that it will always give reasonable results:
 - quality of bootstrap approximation depends on the actual sample size, n , so if n is small, then it's not likely the bootstrap will do a good job;
 - there are cases where bootstrap actually will fail, so it cannot be used blindly without consideration of the underlying theory.