STAT 641: BOOTSTRAPPING METHODS

Jiyoun Myung

Department of Statistics and Biostatistics
California State University, East Bay

Spring 2021, Day 13

Review

The Bootstrap Principle

Real world

Unknown probability distribution

$$F \to x = (x_1, x_2, \ldots, x_n)$$

Observed random sample

$$\widehat{\theta} = s(x)$$

Statistic of interest

Bootstrap world

Empirical distribution

$$\widehat{F} \to x^* = (x_1^*, x_2^*, \ldots, x_n^*)$$

Bootstrap sample

$$\widehat{\theta}^* = s(x^*)$$

Bootstrap replication

The Bootstrap Principle

Suppose

- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is an observed random sample from an unknown probability distribution F.
- $\theta = t(F)$ is a parameter of interest on the basis of $\mathbf{x}$.
- $\hat{\theta} = s(\mathbf{x})$ is an estimate for θ.

For an evaluation of the statistical properties such as bias and standard error for the estimate $\hat{\theta}$, we wish to estimate the sampling distribution of $\hat{\theta}$.

The bootstrapping method mimics the data-generating process by sampling from an estimate $\hat{F}$ of the unknown distribution F. Thus the role of the above real quantities is taken by their analogous quantities in the "bootstrap world":

- $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is a bootstrap from $\hat{F}$.
- $\theta^* = t(\hat{F})$ is the parameter in the bootstrap world.
- $\hat{\theta}^* = s(\mathbf{x}^*)$ is the bootstrap replication of θ.

- The bootstrap distribution of $\hat{\theta}_B - \hat{\theta}$ approximates (fairly well) the sampling distribution of $\hat{\theta} - \theta$.

Empirical distribution function (EDF)

Having observed a random sample of size n from a probability distribution F,

$$F \rightarrow (x_1, x_2, \cdots, x_n), \tag{4.1}$$

the *empirical distribution function* $\hat{F}$ is defined to be the discrete distribution that puts probability $1/n$ on each value x_i, $i = 1, 2, \cdots, n$. In other words, $\hat{F}$ assigns to a set A in the sample space of x its empirical probability

$$\widehat{\text{Prob}}\{A\} = \#\{x_i \in A\}/n, \tag{4.2}$$

the proportion of the observed sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ occurring in A. We will also write $\text{Prob}_{\hat{F}}\{A\}$ to indicate (4.2). The hat symbol "$\wedge$" always indicates quantities calculated from the observed data.

The Plug-In Principle

The bootstrap method is a direct application of the plug-in principle.

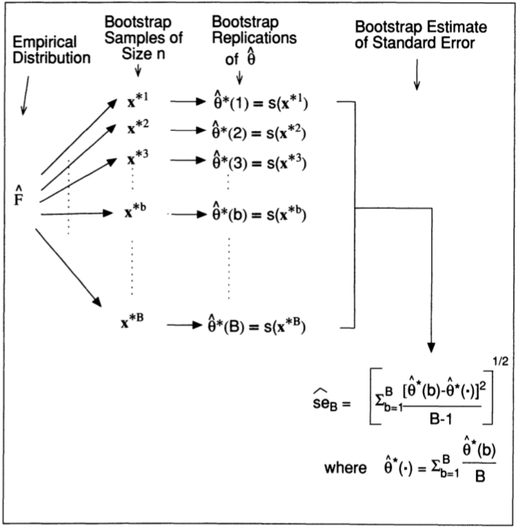
The plug-in principle is a simple method of estimating parameters from samples. The *plug-in estimate* of a parameter $\theta = t(F)$ is defined to be

$$\hat{\theta} = t(\hat{F}).$$

In other words, we estimate the function $\theta = t(F)$ of the probability distribution F by the same function of the empirical distribution $\hat{F}$, $\hat{\theta} = t(\hat{F})$. (Statistics like (4.13) that are used to estimate parameters are sometimes called *summary statistics*, as well as *estimates* and *estimators*.)

The Non-Parametric Bootstrap

The bootstrap algorithm for estimating the standard error of $\hat{\theta} = s(\mathbf{x})$.



Non-parametric vs Parametric Bootstrap

Non-parametric Bootstrap

$$
\begin{aligned}
X_1^{*1}, X_2^{*1}, \ldots, X_n^{*1} &\sim \hat{F} \\
X_1^{*2}, X_2^{*2}, \ldots, X_n^{*2} &\sim \hat{F} \\
&\vdots \\
X_1^{*B}, X_2^{*B}, \ldots, X_n^{*B} &\sim \hat{F}
\end{aligned}
$$

Parametric Bootstrap

$$
\begin{aligned}
X_1^{*1}, X_2^{*1}, \ldots, X_n^{*1} &\sim \hat{F}_{\mathsf{par}} \\
X_1^{*2}, X_2^{*2}, \ldots, X_n^{*2} &\sim \hat{F}_{\mathsf{par}} \\
&\vdots \\
X_1^{*B}, X_2^{*B}, \ldots, X_n^{*B} &\sim \hat{F}_{\mathsf{par}}
\end{aligned}
$$

The Parametric Bootstrap

Example: Let $X_1, X_2, \ldots, X_n \sim N(0, \sigma^2)$, where σ^2 is unknown number. A natural way to estimate σ^2 is via the sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. How do we estimate the variance of the sample variance?

Assume we generate B sets of samples:

$$
\begin{gathered}
X_1^{*(1)}, \cdots, X_n^{*(1)} \sim N(0, S_n^2) \\
X_1^{*(2)}, \cdots, X_n^{*(2)} \sim N(0, S_n^2) \\
\vdots \\
X_1^{*(B)}, \cdots, X_n^{*(B)} \sim N(0, S_n^2).
\end{gathered}
$$

Then, to estimate the variability of S_n^2,

$$
\widehat{\mathsf{Var}}_B(S_n^2) = \frac{1}{B-1}\sum_{b=1}^{B}\left(S_n^{2*}(b) - \bar{S}_n^{2*}\right)^2,
$$

where $S_n^{2*}(b)$, $b = 1, 2, \ldots, B$ are the sample variance of each bootstrap sample, and $\bar{S}_n^{2*} = \frac{1}{B}\sum_{b=1}^{B} S_n^{2*}(b)$.

Bias Correction

Why would we want to estimate the bias of $\hat{\theta}$?

The usual reason is to correct $\hat{\theta}$ so that it becomes less biased. If $\widehat{\text{bias}}$ is an estimate of $\text{bias}_F(\hat{\theta}, \theta)$, then the obvious **bias-corrected estimator** is

$$\bar{\theta} = \hat{\theta} - \widehat{\text{bias}}.$$

Taking $\widehat{\text{bias}}$ equal to $\widehat{\text{bias}}_B = \hat{\theta}^*(\cdot) - \hat{\theta}$ gives

$$\bar{\theta} = \hat{\theta} - \widehat{\text{bias}}_B = 2\hat{\theta} - \hat{\theta}^*(\cdot).$$

- Bias estimation is usually interesting and worthwhile, but the exact use of a bias estimate is often problematic. Authors recommend using:

a. $\hat{\theta}$ if $\widehat{\text{bias}}_B$ is small relative to the $\widehat{\text{se}}_B$.

b. $\bar{\theta}$ if $\widehat{\text{bias}}_B$ is large relative to the $\widehat{\text{se}}_B$.

c. They suggest and justify that if $\widehat{\text{bias}}_B < .25\widehat{\text{se}}_B$ then the bias can be ignored (unless the goal is precise confidence interval estimation using this standard error).

Bootstrap CIs

CI	Symmetric	Range Resp	Trans Resp	Accuracy	Normal Samp Dist?	Other
BS SE	Yes	No	No	1^{st} order	Yes	param assump $F(\hat{\theta})$
BS-t	No	No	No	2^{nd} order	Yes/No	computer intensive
perc	No	Yes	Yes	1^{st} order	No	small $n \to$ low accuracy
BCa	No	Yes	Yes	2^{nd} order	No	limited param assump

Regression Models with Bootstrap

- If the least squares estimation procedure is used to estimate the regression parameters and the model is reasonable and the noise term can be considered to be independent and identically distributed random variables with mean 0, and finite variance σ^2, bootstrapping will not add anything.
- That is because of the Gauss – Markov theorem that asserts that the least squares estimates of the regression parameters are unbiased and have minimum variance among all unbiased estimators.
- Moreover, if the residuals can be assumed to have a Gaussian distribution, the least squares estimates have the nice additional property of being maximum likelihood estimates and are therefore the most efficient (accurate) estimates.

Two types of Bootstrap regression

We can treat the predictors as random (potentially changing from sample to sample) or we can treat them as fixed.

- Random x resampling
  - it is also called **observation resampling** or **case resampling**.
  - Resample observations as with correlation example.
- Fixed x resampling
  - it is also called is called **model based resampling** or **residual resampling**.
  - Resample residuals as follows
    - Fit a model and compute residuals
    - Generate the bootstrap data by $Y* =$ (Fit) $+$ Bootstrap sample of OLS residuals.

Observation Resampling vs Residual Resampling

- **Observation resampling** is a good choice when we are modeling observational data in which the explanatory variables are observed randomly from a population.
- **Residual resampling** is a good choice if we are analyzing data from a designed experiment in which the explanatory variables have a small number of specified values.
- Residual resampling requires a "true" model in order to obtain the residuals which are resampled. Observation (or random) resampling does not. Residual resampling keeps the same X's in every bootstrap sample.
- As the sample size grows (with other conditions), two methods become similar, assuming the model is correctly identified.
- Random resampling usually leads to a larger estimate of standard error (with enough bootstrap replications) since it allows for more sources of variation (from randomness in X's).
- Bootstrap SE of residual resampling will be close to classical SE (OLS formula) as $B \to \infty$. But, Bootstrap SE of observation resampling does not always agree with classical SE.

Permutation Test

Fisher's permutation test

Algorithm 15.1

Computation of the two-sample permutation test statistic

1. Choose B independent vectors $\mathbf{g}^*(1), \mathbf{g}^*(2), \cdots, \mathbf{g}^*(B)$, each consisting of n z's and m y's and each being randomly selected from the set of all $\binom{N}{n}$ possible such vectors. [B will usually be at least 1000; see Table (15.3).]

2. Evaluate the permutation replications of $\hat{\theta}$ corresponding to each permutation vector,

$$\hat{\theta}^*(b) = S(\mathbf{g}^*(b), \mathbf{v}), \qquad b = 1, 2, \cdots, B. \tag{15.17}$$

3. Approximate ASL_{perm} by

$$\widehat{\text{ASL}}_{\text{perm}} = \#\{\hat{\theta}^*(b) \geq \hat{\theta}\}/B. \tag{15.18}$$

HT with the bootstrap

Computing the bootstrap test statistic for $H_0 : F = G$

Algorithm 16.1

Computation of the bootstrap test statistic for testing $F = G$

1. Draw B samples of size $n + m$ with replacement from $\mathbf{x}$. Call the first n observations $\mathbf{z}^*$ and the remaining m observations $\mathbf{y}^*$.
2. Evaluate $t(\cdot)$ on each sample,

$$t(\mathbf{x}^{*b}) = \bar{\mathbf{z}}^* - \bar{\mathbf{y}}^*, \quad b = 1, 2, \cdots B. \tag{16.2}$$

3. Approximate $\mathrm{ASL}_{\mathrm{boot}}$ by

$$\widehat{\mathrm{ASL}}_{\mathrm{boot}} = \#\{t(\mathbf{x}^{*b}) \geq t_{obs}\}/B, \tag{16.3}$$

where $t_{obs} = t(\mathbf{x})$ the observed value of the statistic.

HT with the bootstrap

Algorithm 16.2

Computation of the bootstrap test statistic for testing equality of means

1. Let $\hat{F}$ put equal probability on the points $\tilde{z}_i = z_i - \bar{z} + \bar{x}, i = 1, 2, \ldots n$, and $\hat{G}$ put equal probability on the points $\tilde{y}_i = y_i - \bar{y} + \bar{x}$, $i = 1, 2, \ldots m$, where $\bar{z}$ and $\bar{y}$ are the group means and $\bar{x}$ is the mean of the combined sample.
2. Form B bootstrap data sets $(\mathbf{z}^*, \mathbf{y}^*)$ where $\mathbf{z}^*$ is sampled with replacement from $\tilde{z}_1, \tilde{z}_2, \cdots \tilde{z}_n$ and $\mathbf{y}^*$ is sampled with replacement from $\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_m$.
3. Evaluate $t(\cdot)$ defined by (16.5) on each data set,

$$t(\mathbf{x}^{*b}) = \frac{\bar{z}^* - \bar{y}^*}{\sqrt{\bar{\sigma}_1^{2*}/n + \bar{\sigma}_2^{2*}/m}}, \quad b = 1, 2, \cdots B. \tag{16.6}$$

4. Approximate $\mathrm{ASL_{boot}}$ by

$$\widehat{\mathrm{ASL}}_{\mathrm{boot}} = \#\{t(\mathbf{x}^{*b}) \geq t_{obs}\}/B, \tag{16.7}$$

where $t_{obs} = t(\mathbf{x})$ is the observed value of the statistic.

Permutation test vs Bootstrap

- ASL_{perm} is the exact probability of obtaining a test statistic as extreme as the one observed, having fixed the data values of the combined sample. In contrast, the bootstrap explicitly estimates the probability mechanism under the null hypothesis.

- A permutation test can only test the null hypothesis $F = G$, while the bootstrap can test equal means and equal variances, or equal means with possibly unequal variances.