Bootstrapping in statistics

Bootstrap is a computer-based technique for doing statistical inference (usually with a minimum of assumptions). It is not Bayesian.

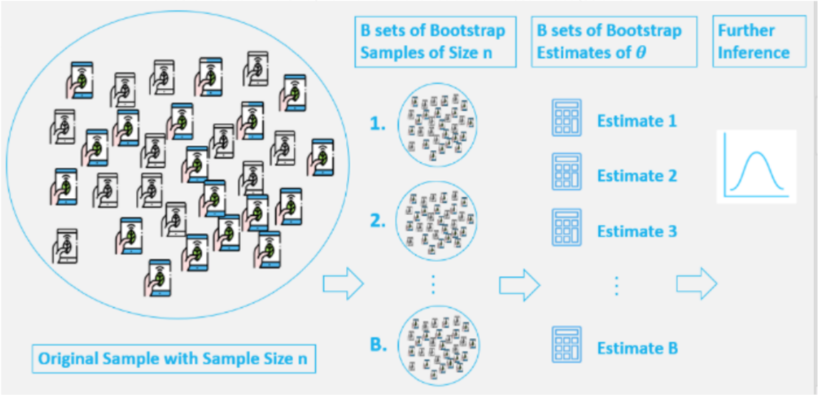
What have we learned

- Empirical distribution function
- Plug-in estimate
- Bootstrap sample
- Bootstrap estimate of the standard error

Brief reminder: Empirical distribution and plug-in principle

- assume iid observations $F \to (x_1, \ldots, x_n)$
- empirical distribution $\hat{F}$ puts prob. $1/n$ to each observed value.
- parameter of interest: $\theta = t(F)$
- plug-in estimator: $\hat{\theta} = t(\hat{F})$

The bootstrap idea



Brief reminder: Bootstrap estimator for standard error

- assume

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

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

Brief reminder: Bootstrap estimator for standard error

- assume

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

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

- want to estimate $\mathrm{SD}_F(\hat{\theta})$

Brief reminder: Bootstrap estimator for standard error

- assume

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

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

- want to estimate $\mathrm{SD}_F(\hat{\theta})$
- bootstrap sample: $\hat{F} \to (x_1^\star, \ldots, x_n^\star) = x^\star$

Brief reminder: Bootstrap estimator for standard error

- assume

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

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

- want to estimate $\mathrm{SD}_F(\hat{\theta})$
- bootstrap sample: $\hat{F} \to (x_1^\star, \ldots, x_n^\star) = x^\star$
- bootstrap replication of $\hat{\theta}$: $\hat{\theta}^\star = s(x^\star)$

Brief reminder: Bootstrap estimator for standard error

- assume

$$F \to (x_1, \dots, x_n) = x$$

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

- want to estimate $\mathrm{SD}_F(\hat{\theta})$
- bootstrap sample: $\hat{F} \to (x_1^\star, \dots, x_n^\star) = x^\star$
- bootstrap replication of $\hat{\theta}$: $\hat{\theta}^\star = s(x^\star)$
- ideal bootstrap estimate of $\mathrm{SD}_F(\hat{\theta})$: $SD_{\hat{F}}(\hat{\theta}^\star)$.

Brief reminder: Bootstrap estimator for standard error

- assume

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

$$\hat{F} : \text{empirical distribution}$$

$$\theta = t(F)$$

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

- want to estimate $\mathrm{SD}_F(\hat{\theta})$
- bootstrap sample: $\hat{F} \to (x_1^\star, \ldots, x_n^\star) = x^\star$
- bootstrap replication of $\hat{\theta}$: $\hat{\theta}^\star = s(x^\star)$
- ideal bootstrap estimate of $\mathrm{SD}_F(\hat{\theta})$: $SD_{\hat{F}}(\hat{\theta}^\star)$.
- this estimate can in principle be computed in practice usually not (need to be approximated via MC).

Bootstrapping dependent data

- Critical requirement: Boostrapped quantities are iid
- How can we use the bootstrap in other contexts?

Bootstrapping dependent data

Consider a first-order stationary autoregressive process, the AR(1) model:

$$X_t = \alpha X_{t-1} + \epsilon_t$$

where $|\alpha| < 1$ and ϵ_t are iid with mean zero and constant variance.

Here, a method akin to bootstrapping the residuals for linear regression can be applied.

AR(1) model: A model based approach

1. Use a standard method to estimate α
2. Define the estimated innovations $\hat{e}_t = X_t - \hat{\alpha} X_{t-1}$ for $t = 2, \ldots, n$ and let $\bar{\epsilon}$ be the mean of these.
3. Recenter $\hat{e}_t$ to have mean zero by defining $\hat{\epsilon}_t = \hat{e}_t - \bar{e}$.
4. Resample $n+1$ values from the set $\{\hat{\epsilon}_2, \ldots, \hat{\epsilon}_n\}$ with replacement to yield pseudo innovations $\{\epsilon_0^\star, \ldots, \epsilon_n^\star\}$.
5. Generate pseudo data as $X_0^\star = \epsilon_0^\star$ and $X_t^\star = \hat{\alpha} X_{t-1}^\star + \epsilon_t^\star$ for $t = 1, \ldots, n$.
6. From each bootstrap sample compute $\hat{\alpha}^\star$

AR(1) model: A model based approach

Issue: Pseudo-data series is not stationary.

Remedy: Sample larger number of pseudo innovations and generate data series earlier, i.e. $X_k^\star$ for k much less than zero. The first portion of the data can be discarded as burn-in.

Show `Lutenizing_boot.R` code

Block bootstrap

An alternative bootstrap procedure for time series data is to draw blocks from the observed series.

- Issue: We cannot simply sample from the individual observations, as this would destroy the correlation that we try to capture.
- Idea: Block data to preserve covariance structure within each block, even though structure is lost between blocks.

Here, we consider

- Non-moving blocks bootstrap
- Moving blocks bootstrap

Non-moving blocks bootstrap

Illustration and example:

See blackboard

Non-moving blocks bootstrap (II)

- Split $x_1, \dots, x_n$ into b non-overlapping blocks of length l, where ideally $n = l \cdot b$.
- Sample $\mathcal{B}_1^\star, \dots, \mathcal{B}_b^\star$ independently from $\{\mathcal{B}_1, \dots, \mathcal{B}_b\}$ with replacement. Concatenate these blocks to form a pseudo dataset $\mathcal{X}^\star = (\mathcal{B}_1^\star, \dots, \mathcal{B}_b^\star)$.
- Replicate this process B times and estimate for each bootstrap sample $\hat{\theta}_i^\star$.
- Approximate the distribution of $\hat{\theta}$ by the distribution of these B pseudo values.

Moving blocks bootstrap

Illustration:

See blackboard

Show `Lutenizing_boot.R` code

Block bootstrap

- Idea: With blocks bootstrap, choose block size l large enough so that observations more than l units apart will be nearly independent.
- Advantage: Less model dependent than residuals approach. However, choice of block size l can be quite important, and effective methods to choose l are still laking.

Permutation test

(related to idea of bootstrapping.)

Consider a medical experiment where rats are randomly assigned to treatment and control groups. Under the null hypothesis the outcome measured does not depend on the group assignment.

Idea: Shuffling the labels randomly among rates will not change the joint null distribution of the data.

Recall: P-value

- Let t_1 denote the original test statistic, e.g. difference of group mean outcomes, and $t_2, \ldots, t_B$ the test statistics computed from the datasets resulting from B permutations of labels.
- Under the null hypothesis $t_2, \ldots, t_B$ are from the same distribution that yielded $t_1 \Rightarrow$ We can compare them.

We can use the P-value:

> *P-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.*

Permutation test: Example

The simple model for independent data from two sources:

$$y_i \sim F_1, \quad i = 1, \ldots, m$$

$$z_j \sim F_2, \quad j = 1, \ldots, n$$

$$\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z}) = (y_1, \ldots, y_m, z_1, \ldots, z_n)$$

The permutation method for hypothesis testing is based on resampling under the null hypothesis $H_0 : F_1 = F_2$, by permuting the order of the original data to generate B bootstrap samples $\boldsymbol{x}^*$, valid given that the null hypothesis is true.

The p-value for a test based on a test quantity $t(\boldsymbol{x})$ can be estimated as $\#\{t(\boldsymbol{x}^*) \geq t(\boldsymbol{x})\}/B$. H_0 is rejected if the p-value is smaller than a given threshold (typically 0.05 or 0.01)

Permutation test: Example

1. We test the hypothesis

$$H_0 : F_1 = F_2 \qquad \text{against} \qquad H_1 : F_1 \neq F_2$$

   using the test quantity $T = |\bar{y} - \bar{z}|$, by means of the permutation method to compute an estimate tof the p-value for the test.

2. The test only tests for differences that can be detected by the test quantity. Consider an alternative test quantity $T = \left| \frac{\left(\frac{1}{m}\sum_{i=1}^{m} y_i\right)^2}{\frac{1}{m}\sum_{i=1}^{m} y_i^2} - \frac{\left(\frac{1}{n}\sum_{j=1}^{n} z_j\right)^2}{\frac{1}{n}\sum_{j=1}^{n} z_j^2} \right|$

Permutation test: R-code

see demo-permTest.R