

## Bootstrapping in statistics

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

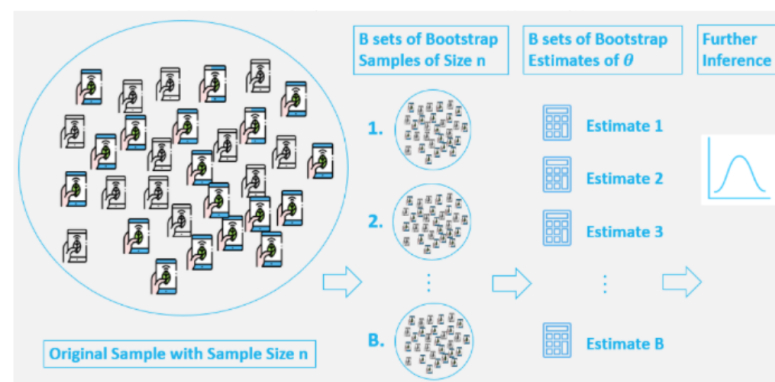
## What have we learned

- 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  $F \rightarrow (x_1, \dots, x_n)$
- 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

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

## Brief reminder: Bootstrap estimator for standard error

- assume

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

- want to estimate  $SD_F(\hat{\theta})$

## Brief reminder: Bootstrap estimator for standard error

- assume

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

- want to estimate  $SD_F(\hat{\theta})$
- bootstrap sample:  $\hat{F} \rightarrow (x_1^*, \dots, x_n^*) = x^*$

## Brief reminder: Bootstrap estimator for standard error

- assume

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

- want to estimate  $SD_F(\hat{\theta})$
- bootstrap sample:  $\hat{F} \rightarrow (x_1^*, \dots, x_n^*) = x^*$
- bootstrap replication of  $\hat{\theta}$ :  $\hat{\theta}^* = s(x^*)$

## Brief reminder: Bootstrap estimator for standard error

- assume

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

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

## Bootstrapping dependent data

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

## Brief reminder: Bootstrap estimator for standard error

- assume

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

$\hat{F}$  : empirical distribution

$$\theta = t(F)$$

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

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

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

## 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.  $X_k^*$  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  $\mathcal{B}_1^*, \dots, \mathcal{B}_b^*$  independently from  $\{\mathcal{B}_1, \dots, \mathcal{B}_b\}$  with replacement. Concatenate these blocks to form a pseudo dataset  $\mathcal{X}^* = (\mathcal{B}_1^*, \dots, \mathcal{B}_b^*)$ .
- Replicate this process  $B$  times and estimate for each bootstrap sample  $\hat{\theta}_i^*$ .
- Approximate the distribution of  $\hat{\theta}$  by the distribution of these  $B$  pseudo values.

## 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 lacking.

## Moving blocks bootstrap

Illustration:

See blackboard

Show Lutenizing\_boot.R code

## 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 rats 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  $t_2, \dots, t_B$  the test statistics computed from the datasets resulting from  $B$  permutations of labels.
- Under the null hypothesis  $t_2, \dots, t_B$  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:

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

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

$$\mathbf{x} = (\mathbf{y}, \mathbf{z}) = (y_1, \dots, y_m, z_1, \dots, z_n)$$

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  $\mathbf{x}^*$ , valid given that the null hypothesis is true.

The p-value for a test based on a test quantity  $t(\mathbf{x})$  can be estimated as  $\#\{t(\mathbf{x}^*) \geq t(\mathbf{x})\}/B$ .  $H_0$  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

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

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

