

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: $\hat{\theta} = T(\hat{F})$

How large do we need B ?

Intuitively we understand that the \widehat{SE}_B has larger standard deviation than \widehat{SE}_∞ .

Theory, not to be discussed here, gives the following rules of thumb:

1. Even a small B is informative, say $B = 25$ or $B = 50$ is often enough to get a good estimate of $SE_F(\hat{\theta})$.
2. Very seldomly more than $B = 200$ is necessary to estimate $SE_F(\hat{\theta})$.

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: $\hat{F} \rightarrow (x_1^*, \dots, x_n^*) = x^*$
- 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.

The parametric bootstrap

When data are modeled to originate from a parametric distribution, so

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} F(x, \xi),$$

another estimate of F may be employed.

Suppose that the observed data are used to estimate ξ by $\hat{\xi}$. Then each **parametric bootstrap** pseudo-dataset \mathcal{X}^* can be generated by drawing $X_1^*, \dots, X_n^* \stackrel{\text{iid}}{\sim} F(x, \hat{\xi}) = \hat{F}_{\text{par}}$.

Again ...

... we can/must estimate $SD_{\hat{F}_{\text{par}}}(\hat{\theta}^*)$ by simulation:

1. Generate B **bootstrap samples** x^{1*}, \dots, x^{B*} , where

$$x^{b*} = (x_1^{b*}, \dots, x_n^{b*})$$

with $x_1^{b*}, \dots, x_n^{b*} \stackrel{\text{iid}}{\sim} \hat{F}_{\text{par}}$.

2. Evaluate the corresponding parameter estimates

$$\hat{\theta}^*(b) = s(x^{b*}), \quad b = 1, 2, \dots, B$$

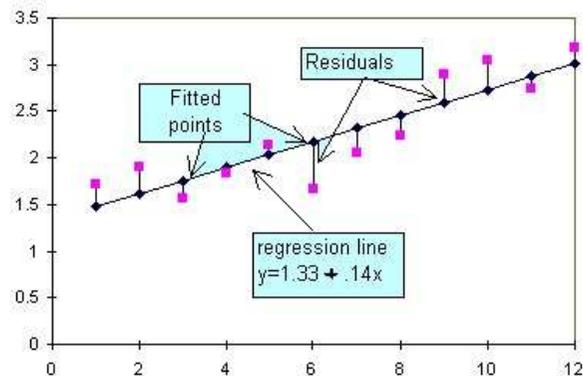
3. Estimate $SD_{\hat{F}_{\text{par}}}(\hat{\theta}^*)$ by

$$\widehat{SE}_B = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2}{B-1}}$$

where

$$\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b)$$

Review: Residuals



Bootstrapping regression

Consider the ordinary multiple regression model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, \dots, n,$$

where ϵ_i are iid mean zero random variables with constant variance.

- Naive: Bootstrapping by resampling from response variables to get distribution of $\hat{\boldsymbol{\beta}}^*$. However $Y_i | \mathbf{x}_i$ are not iid!
- Correct: **Bootstrap the residuals.**

Bootstrap the residuals

1. Fit the regression model to the observed data and obtain the fitted responses \hat{y}_i and residuals $\hat{\epsilon}_i$.
2. Sample a bootstrap set of residuals $\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*$ from the set of fitted residuals **completely at random and with replacement**.
3. Generate a bootstrap set of pseudo responses

$$Y_i^* = \hat{y}_i + \hat{\epsilon}_i^*, \quad \text{for } i = 1, \dots, n.$$

4. Regress Y^* on \mathbf{x} to obtain a bootstrap estimate $\hat{\boldsymbol{\beta}}^*$.

Repeat this process to get an empirical distribution of $\hat{\boldsymbol{\beta}}^*$.

Bootstrapping residuals: Remarks

This approach is also used for **autoregressive models**, for example.

Note: Bootstrapping the residuals is reliant on

- The model provides an appropriate fit
- The residuals have a constant variance

Otherwise, a different scheme is recommended.

Comment: No need to bootstrap for linear regression model and least squares estimation, as analytical results are then available.

Copper-nickel alloy

Data: 13 measurements of corrosion loss (y_i) in copper-nickel alloys, each with a specific iron content (x_i).

Question: Change in corrosion loss in the alloys as the iron content increases, relative to corrosion loss where there is no iron, i.e.

$$\theta = \beta_1 / \beta_0.$$

x_i	0.01	0.48	0.71	0.95	1.19	0.01	0.48
y_i	127.6	124.0	110.8	103.9	101.5	130.1	122.0
x_i	1.44	0.71	1.96	0.01	1.44	1.96	
y_i	92.3	113.1	83.7	128.0	91.4	86.2	

The observed data yield $\hat{\theta} = \hat{\beta}_1 / \hat{\beta}_0 = -0.185$.

Paired bootstrap

Suppose response and predictors are measured from a collection of individuals selected at random

⇒ Data pairs $z_i = (x_i, y_i)$ can be regarded as iid realisation from $Z_i = (X_i, Y_i)$ drawn from a **joint response-predictor distribution**.

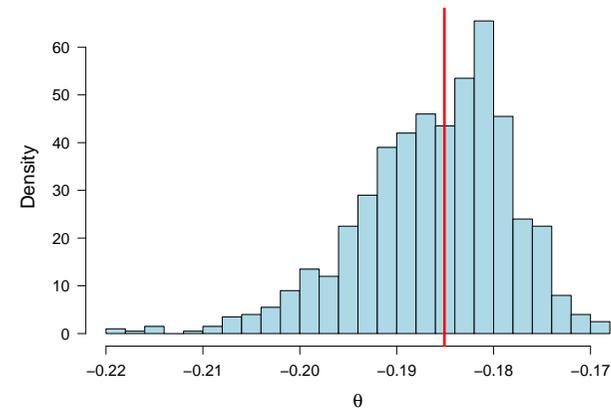
Bootstrap:

- Sample Z_1^*, \dots, Z_n^* completely at random with replacement from z_1, \dots, z_n .
- Apply regression model on pseudo dataset to get $\hat{\beta}^*$.

Repeat this approach many times.

Note: Paired bootstrap is less sensitive to violation of assumptions, e.g. adequacy of regression model, than bootstrapping the residuals.

Histogram of 10 000 bootstrap estimates



Show R-code demo-pairedBootstrap.R

Bootstrap bias correction

The mean value of

$$\hat{\theta}^* - \hat{\theta}$$

among the pseudo datasets is about -0.00125 .

The **bias-corrected bootstrap estimate** of β_1/β_0 is $-0.18507 - (-0.00125) = -0.184$.

AR(1) model: A model based approach

1. Use a standard method to estimate α
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 $\hat{\alpha}^*$

Bootstrapping dependent data

Critical requirement: Bootstrapped quantities are iid.

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

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.

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 (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.

Non-moving blocks bootstrap

Illustration and example:

See blackboard

Moving blocks bootstrap

Illustration:

See blackboard

- **Idea:** With **moving 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.

Confidence intervals

A “simple-minded” two-sided confidence interval with coverage $(1 - \alpha)$ for a parameter α is given by

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

where q_{α}^* is the α -bootstrap quantile in the distribution of $\hat{\theta}^*$.

Experience: Often good, but often **too low coverage**, i.e the true α for the interval is lower than the specified value.

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.

Confidence intervals

Based on consistency considerations a better confidence interval is

$$[\hat{\theta} - \hat{q}_{1-\alpha/2}, \hat{\theta} - \hat{q}_{\alpha/2}]$$

where \hat{q}_{α} is the α -bootstrap quantile in the distribution of $\hat{\theta}^* - \hat{\theta}$.

Due to the invariance of the quantile we have

$$\hat{q}_{\alpha} = q_{\alpha}^* - \hat{\theta}$$

where q_{α}^* is the α -bootstrap quantile in the distribution of $\hat{\theta}^*$.

Therefore, the bootstrap confidence interval becomes

$$[2\hat{\theta} - q_{1-\alpha/2}^*, 2\hat{\theta} - q_{\alpha/2}^*]$$

Note: Better bootstrap confidence intervals exist and often have better coverage accuracy — at the price of being somewhat more difficult to implement

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: R-code

see `demo-permTest.R`

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**

$$T = \left| \frac{(\frac{1}{m} \sum_{i=1}^m y_i)^2}{\frac{1}{m} \sum_{i=1}^m y_i^2} - \frac{(\frac{1}{n} \sum_{j=1}^n z_j)^2}{\frac{1}{n} \sum_{j=1}^n z_j^2} \right|$$