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

Plug in estimate - example

$$\theta = t(F) = \mathsf{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathsf{E}(X^3) - 3\mu\sigma^2 - \mu^2}{\sigma^3}$$

where

$$\mu = \mathsf{E}(X) \text{ and } \sigma = \sqrt{\mathsf{Var}(X)}$$

What is the plug in estimate?

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 $\mathrm{SD}_F(\hat{\theta})$: $SD_{\hat{F}}(\hat{\theta}^\star)$.
- this estimate can in principle be computed in practice usually not (need to be approximated via MC).

Bootstrap estimate of standard error - example

$$\theta = t(F) = \frac{\mathsf{E}(X^3) - 3\mu\sigma^2 - \mu^2}{\sigma^3}$$

where

$$\mu = \mathsf{E}(X) \text{ and } \sigma = \sqrt{\mathsf{Var}(X)}$$

The plug in estimate is:

$$\hat{\theta} = s(\boldsymbol{x}) = \frac{\bar{(}x^3) - 3\bar{x}\ s^2 - \mu^2}{s^3}$$

What is the standard error of $\hat{\theta}$?
†

† Show code boot_example.R

Bootstrapping regression

Consider the ordinary multiple regression model

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

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

- Parameters of interest $\boldsymbol{\beta}$
- Want to estimate $SD(\hat{\beta})$

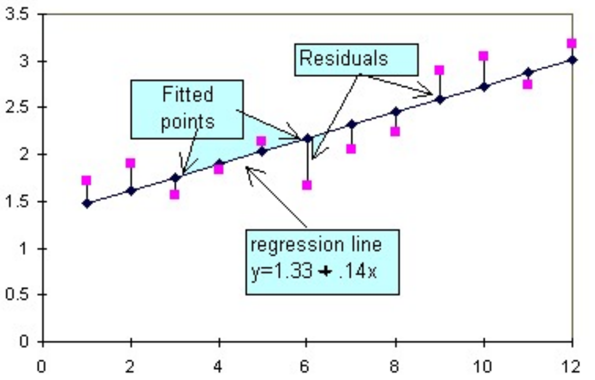
Review: Linear Regression

- Least square estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \text{argmin}\{\sum (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2\} \Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

- Residuals

$$e_i = Y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$$



Bootstrap regression

Alternative 1: Bootstrap the residuals $e_i = Y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$

Alternative 2: Bootstrap the pairs $\boldsymbol{Z}_i = (\boldsymbol{X}_i, Y_i)$

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 $\hat{\epsilon}_1^\star, \ldots, \hat{\epsilon}_n^\star$ from the set of fitted residuals completely at random and with replacement.
3. Generate a bootstrap set of pseudo responses

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

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

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

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.

Bootstrap the pair $\boldsymbol{Z}_i = (\boldsymbol{X}_i, Y_i)$

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

$\Rightarrow$ Data pairs $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)$ can be regarded as iid realisation from $\boldsymbol{Z}_i = (\boldsymbol{X}_i, Y_i)$ drawn from a joint response-predictor distribution.

Bootstrap:

- Sample $\boldsymbol{Z}_1^\star, \ldots, \boldsymbol{Z}_n^\star$ completely at random with replacement from $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$.
- Apply regression model on pseudo dataset to get $\hat{\boldsymbol{\beta}}^\star$.

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.

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 $\hat{\theta} = \hat{\beta}_1/\hat{\beta}_0 = -0.185$.

Show R-code demo-pairedBootstrap.R

Bias of an estimator

- We observe $X_1, X_2, \ldots, X_n \sim F$ iid
- Parameter of interest $\theta = t(F)$
- Estimator $\hat{\theta} = s(X)$

  (may or may not be based on the plug-in principle)
- Bias definition

$$\text{bias}_F(\hat{(}\theta), \theta) = \text{E}_F[\hat{\theta}] - \theta = \text{E}_F[s(\boldsymbol{x})] - t(F)$$

Bootstrap estimate of bias

We want to estimate

$$\text{bias}_F(\hat{(\theta)}, \theta) = \mathsf{E}_F[s(\boldsymbol{x})] - t(F)$$

Idea: Apply the plug-in principle and define the bootstrap estimate of bias as:

$$\text{bias}_{\hat{F}} = \mathsf{E}_{\hat{F}}[s(\boldsymbol{x}^\star)] - t(\hat{F})$$

where $\hat{F}$ is an estimate of F (for example the empirical distribution)

Bias estimate of the bias

1. Generate B bootstrap samples $x^{1\star}, \dots, x^{B\star}$.
2. Evaluate the corresponding parameter estimates

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

3. Approximate the bootstrap expectation $\mathrm{E}_{\hat{F}}[s(\boldsymbol{x}^\star)]$ as:

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

4. Approximate the ideal bootstrap estimate for bias as

$$\widehat{\mathrm{bias}}_B = \hat{\theta}^\star(\cdot) - t(\hat{F})$$

Bias corrected estimate

One we have estimated the bias we can compute the bias-corrected estimator

$$\hat{\theta}_c = \hat{\theta} - \widehat{\text{bias}}_B = \hat{\theta} - [\hat{\theta}^\star(\cdot) - t(\hat{F})]$$

Note: Bias correction will not always give an improved estimator.

We have that $\text{Var}(\hat{\theta}_c) \geq \text{Var}(\hat{\theta})$ so if the bias is small is better not to do bias correction.

Bootstrap bias correction

Copper-nickel alloy example

The mean value of

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

among the pseudo datasets is about -0.00125.

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

Confidence intervals (percentile method)

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

$$[q^{\star}_{\alpha/2}, q^{\star}_{1-\alpha/2}]$$

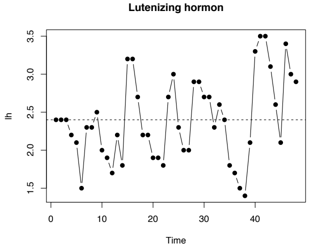
where $q^{\star}_{\alpha}$ is the α-bootstrap quantile in the distribution of $\hat{\theta}^{\star}$.

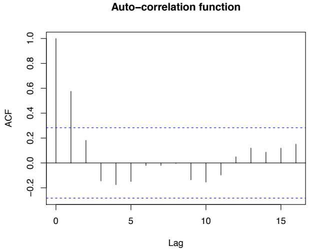
Experience: Often good, but often too low coverage, i.e the true α for the interval is lower than the specified value.

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

Bootstrapping dependent data

Critical requirement: Boostrapped quantities are iid.





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 $x_1, \ldots, x_n$ into b non-overlapping blocks of length l, where ideally $n = l \cdot b$.
- Sample $\mathcal{B}_1^\star, \ldots, \mathcal{B}_b^\star$ independently from $\{\mathcal{B}_1, \ldots, \mathcal{B}_b\}$ with replacement. Concatenate these blocks to form a pseudo dataset $\mathcal{X}^\star = (\mathcal{B}_1^\star, \ldots, \mathcal{B}_b^\star)$.
- 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 $T = |\overline{y} - \overline{z}|$, 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