MS BGD
MDI 720 : Statistiques

François Portier and Joseph Salmon
http://sites.google.com/site/fportierwebpage/
Télécom Paristech, Institut Mines-Télécom

Bootstrap : general principle

Purpose

To measure the accuracy of a statistic $\hat{\theta}$

Algorithm

Real world	Bootstrap world
$(X_1, \dots X_n) \sim \mathbb{P}$ **(unknown)**	$(X_1^*, \dots X_n^*) \sim \hat{\mathbb{P}}$ **(known)**
$\downarrow$	$\downarrow$
$\hat{\theta}$	$\hat{\theta}^*$

The estimator $\hat{\theta}$ is obtained from the data under $\mathbb{P}$. The bootstrap estimator $\hat{\theta}^*$ comes from data under $\widehat{\mathbb{P}}$ which estimates $\mathbb{P}$

Basic idea

$\hat{\theta}^*$ (**known**) mimics the behavior of $\hat{\theta}$ (**unknown**)

Bootstrap algorithm

- $X_1, \ldots, X_n$ are i.i.d. observations
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ a statistic of interest
  Examples : empirical mean $\bar{X}_n$ and median $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, number of bootstrap iterations B
Output : Bootstrap estimators $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$

Bootstrap algorithm

- $X_1, \dots, X_n$ are i.i.d. observations
- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ a statistic of interest
  Examples : empirical mean $\bar{X}_n$ and median $\mathrm{Med}_n(X_1, \dots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \dots, X_n$, number of bootstrap iterations B
Output : Bootstrap estimators $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$
pour $b = 1, \dots, B$ **faire**

Bootstrap algorithm

- $X_1, \ldots, X_n$ are i.i.d. observations
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ a statistic of interest
  Examples : empirical mean $\bar{X}_n$ and median $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, number of bootstrap iterations B
Output : Bootstrap estimators $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$
pour $b = 1, \ldots, B$ **faire**

  Draw (uniformly) with replacement among $X_1, \ldots, X_n$, and provide the new **random** sample :

  Bootstrap sample :$X_1^*, \ldots, X_n^*$

Bootstrap algorithm

- $X_1, \dots, X_n$ are i.i.d. observations
- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ a statistic of interest
  Examples : empirical mean $\bar{X}_n$ and median $\mathrm{Med}_n(X_1, \dots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \dots, X_n$, number of bootstrap iterations B
Output : Bootstrap estimators $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$
pour $b = 1, \dots, B$ **faire**

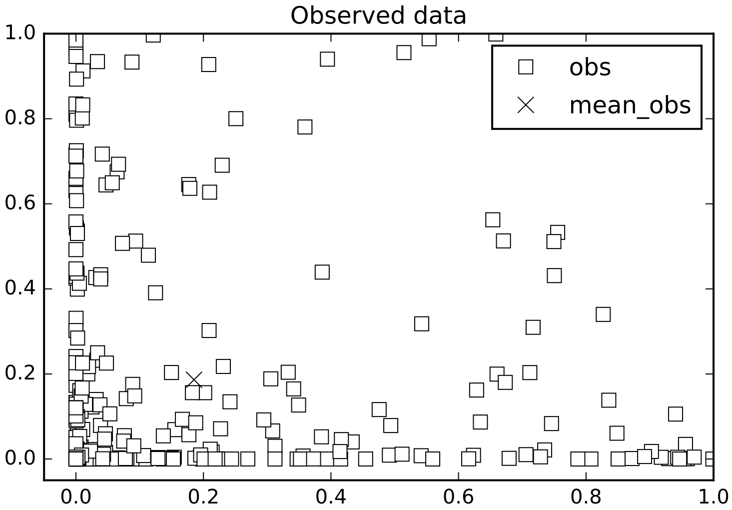
Draw (uniformly) with replacement among $X_1, \dots, X_n$, and provide the new **random** sample :

Bootstrap sample :$X_1^*, \dots, X_n^*$

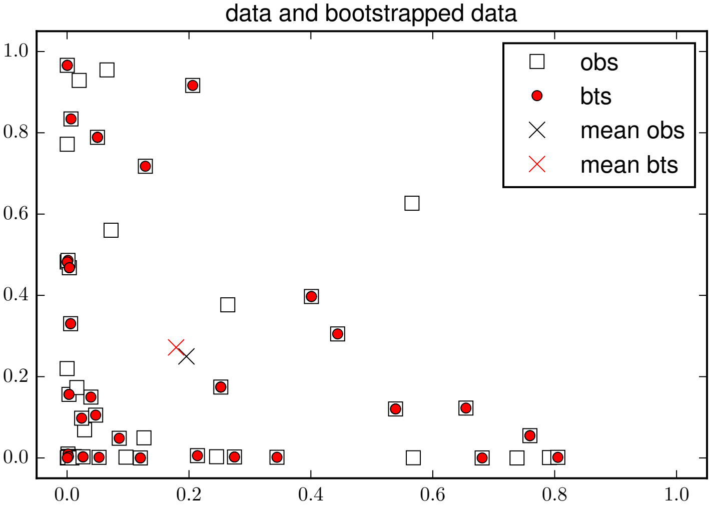
Apply the estimation over the bootstrap sample :

$$\hat{\theta}_b^* = \hat{\theta}(X_1^*, \dots, X_n^*)$$

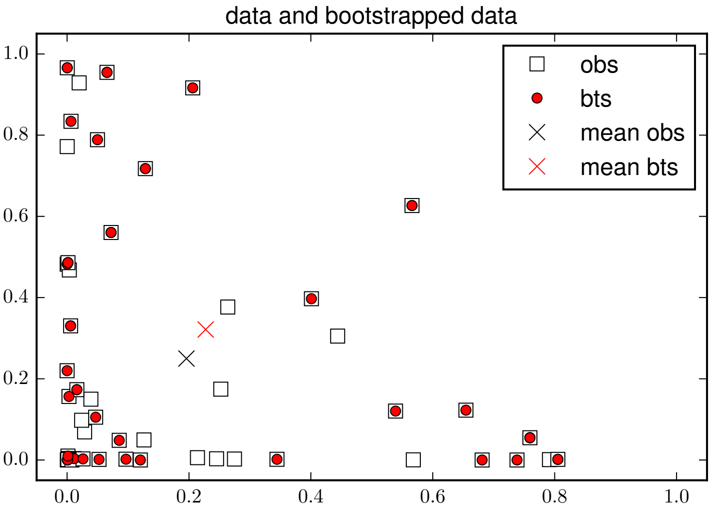
The original bootstrap for the mean



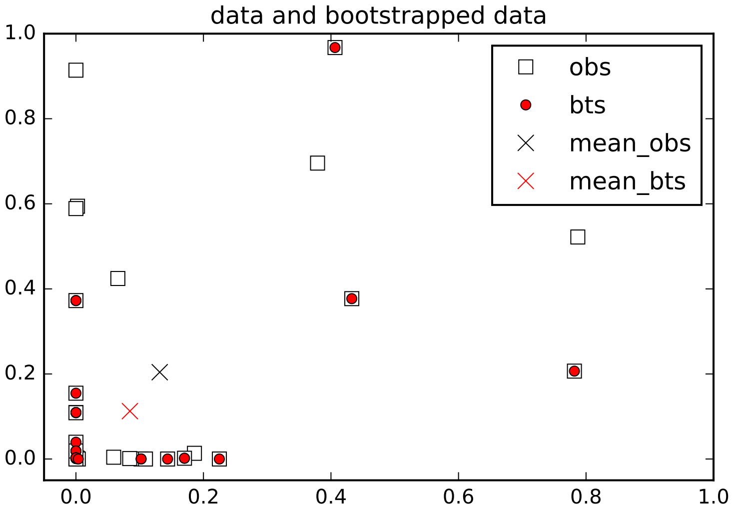
The original bootstrap for the mean



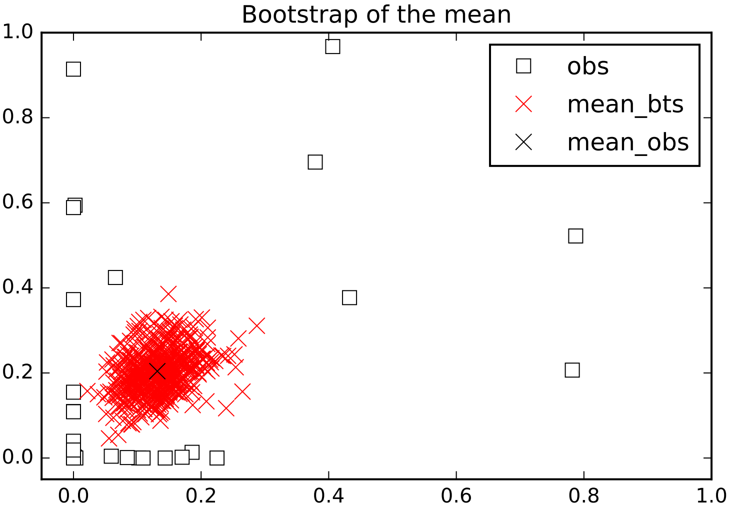
The original bootstrap for the mean



The original bootstrap for the mean



The original bootstrap for the mean

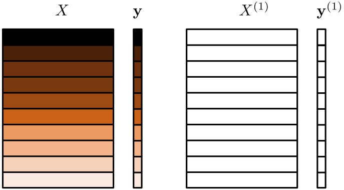


Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$

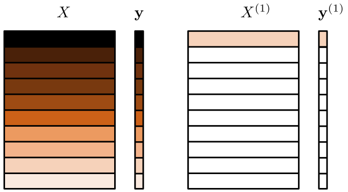
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



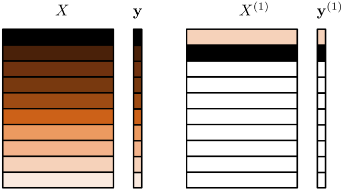
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



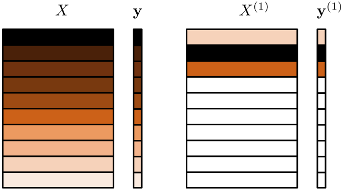
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



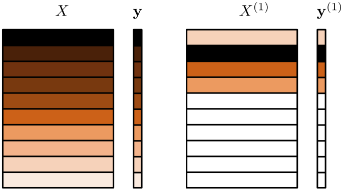
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



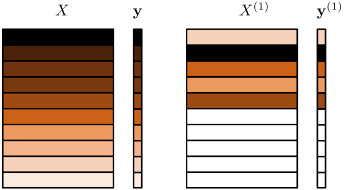
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



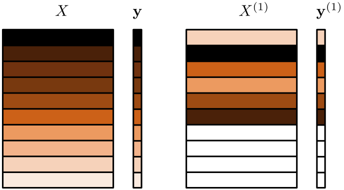
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



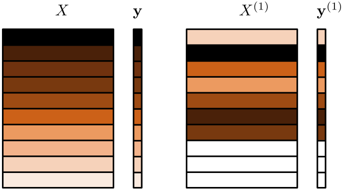
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



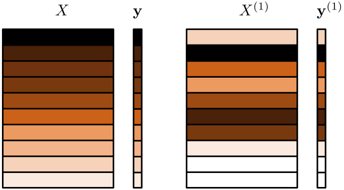
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



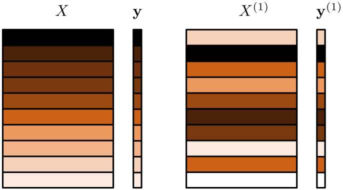
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



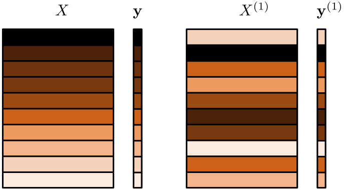
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n\times p}$ and $y \in \mathbb{R}^n$



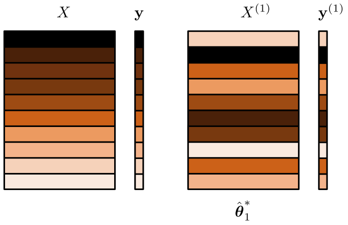
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



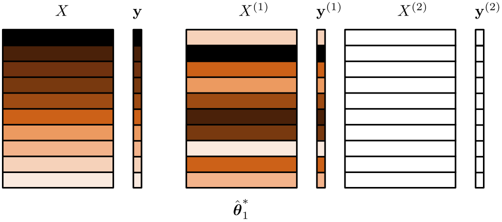
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



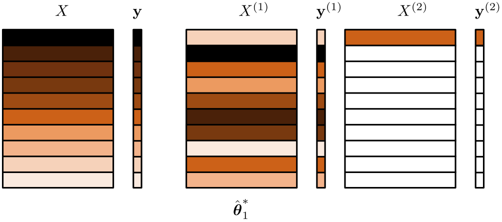
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



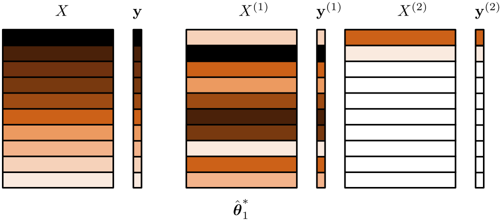
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



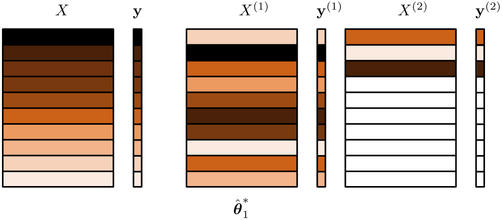
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



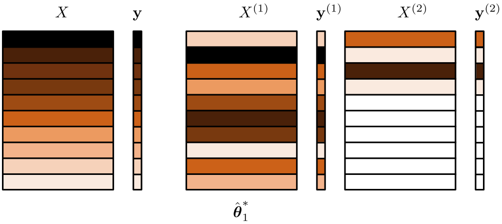
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



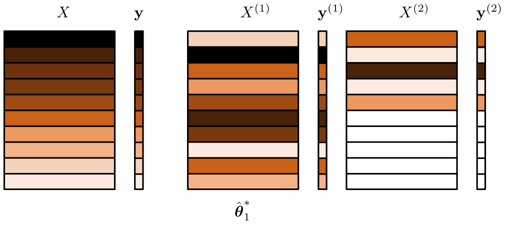
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



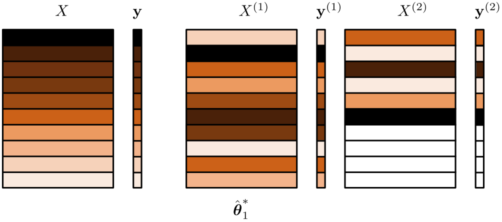
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



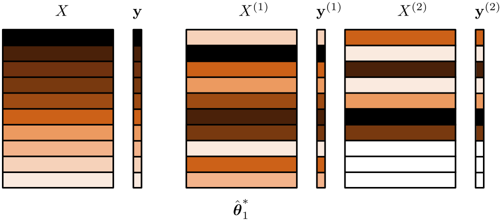
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



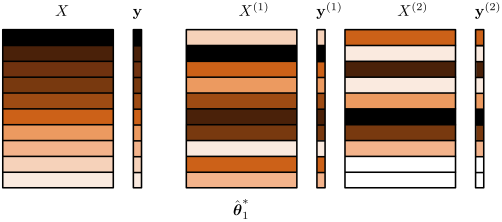
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



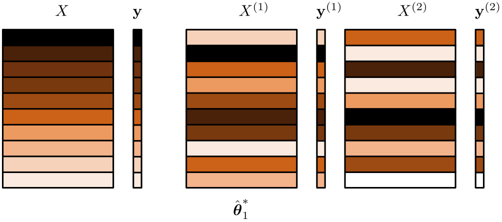
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



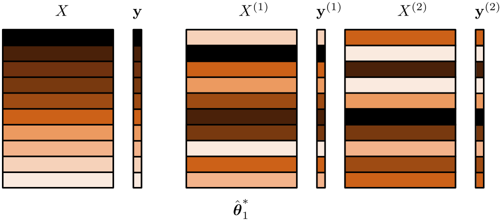
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



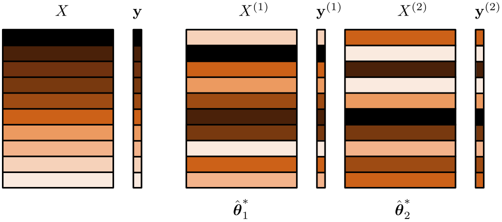
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



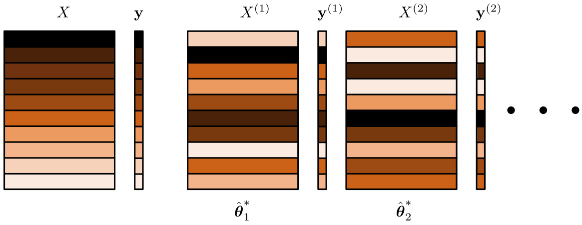
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



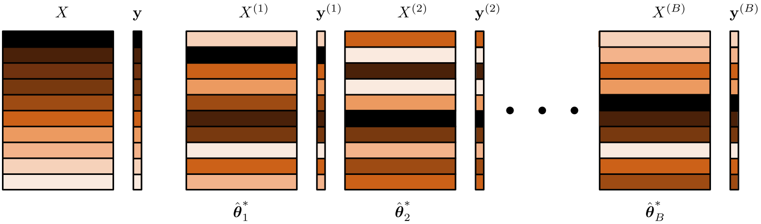
Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



Bootstrap (of the pairs) in regression

Let $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$



Some basic bootstrap estimates

Let θ_0 be the parameter of interest (unknown)

	the true (unknown)	the bootstrap
bias	$\mathbb{E}[\hat{\theta}] - \theta_0$	$B^{-1}\sum_{b=1}^{B} \hat{\theta}_b^* - \hat{\theta}$
variance	$\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$	$B^{-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*)^2$
mean-square error	$\mathbb{E}[(\hat{\theta} - \theta_0)^2]$	$B^{-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - \hat{\theta})^2$
quantiles	$\dots$	$\dots$
density	$\dots$	$\dots$

The statistics $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ are bootstrap"versions" of the statistic $\hat{\theta}$

How to use them ?

Exercise: For the sample mean, compute each quantity of the table (for the right-hand side column, replace sums by true expectations)

Origin Efron et Tibshirani (1993)

The term "bootstrap" comes from the sentence :

"to pull oneself up by one's own bootstrap" (réussir par soi-même)

taken from "The Surprising Adventures of Baron Munchausen" by *R. E. Raspe* (18th century).

Idea Efron (1979)

Based on the observed data, estimate the sampling distribution of some statistics, *e.g.*, mean, standard error, correlation, etc.

No asymptotic theory !

This course includes :

- Resampling schemes for **independent data**
- Large class of estimators : Delta-methods, parametric bootstrap, regression
- Studentized-Bootstrap and choice of B
- Emphasis on confidence intervals

Roots

Definition

A **statistical root** $\hat{R}$ is a measurable function of $(X_1, \ldots, X_n)$ such that $\hat{R}$ converges in distribution to G *i.e.*, $\hat{R} \rightsquigarrow G$

Examples

Let $X_1, \ldots, X_n$ be *i.i.d.*with distribution $\mathcal{U}[0,1]$

- the mean, $n^{1/2}(n^{-1}\sum_{i=1}^{n} X_i - 1/2)$
- the cdf, $n^{1/2}(n^{-1}\sum_{i=1}^{n} 1_{\{X_i \leqslant x\}} - x)$
- the minimum, $n(\min_{1 \leqslant i \leqslant n} X_i)$

Regular case (our context)

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

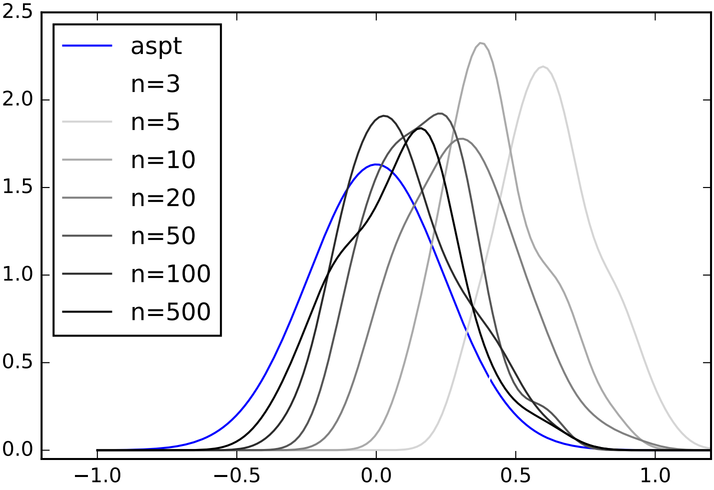


FIGURE: Example of a root with positive bias

Bootstrapping roots

The root $\hat{R}$ is usually given by the problem of interest

Aim of the bootstrap

To reproduce the "behavior" of a given root

Main steps

- (**definition step**)* Find a bootstrap root $\hat{R}^*$ that mimics the root of interest $\hat{R}$
- (**approximation step**)** For some B, compute $\hat{R}_1^*, \dots, \hat{R}_B^*$ and approximate the law of $\hat{R}$

*the definition step is often conducted with the help of asymptotic theory

**the approximation step follows from Monte Carlo simulation

Definition step in examples

Example 1 : The mean

Suppose that

$$\theta_0 = \int x dP(x) \qquad \hat{\theta} = \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \sigma^2 = \int (x-\theta_0)^2 dP(x)$$

From the central limit theorem, if $\mathbb{E}[X_1^2] < +\infty$, it holds that (root property)

$$\hat{R} = n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

The bootstrap version of $\hat{R}$ is

$$\hat{R}^* = n^{1/2}(\hat{\theta}^* - \hat{\theta}), \qquad \hat{\theta}^* = \bar{X}^*$$

Exercise: (Asymptotic validation of the bootstrap)
Show that $\hat{R}^* \rightsquigarrow \mathcal{N}(0, \sigma^2)$

Definition step in examples

Example 2 : The variance

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

if $\mathbb{E}[X_1^4] < +\infty$, we have that

$$\hat{R} = n^{1/2}(\hat{\sigma}^2 - \sigma^2) \rightsquigarrow \mathcal{N}(0, v), \qquad v = \mathrm{var}((X - \mathbb{E}[X])^2)$$

The bootstrap of this root is given by

$$\hat{R}^* = n^{1/2}(\hat{\sigma}^{*2} - \hat{\sigma}^2), \qquad \hat{\sigma}^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}^*)^2$$

Bootstrap vs asymptotics

Target

The **unknown** true distribution of

$$n^{1/2}(\hat{\theta} - \theta_0)$$

Two choices

The (estimated) **asymptotic** distribution, *i.e.*,

$$\mathcal{N}(0, \hat{\sigma}^2)$$

The **bootstrap** distribution, *i.e.*, the distribution of

$$n^{1/2}(\hat{\theta}^* - \hat{\theta})$$

Bootstrap vs asymptotics

Important difference 1

Whereas the validation of the bootstrap is asymptotic (in exercise), the construction of the confidence intervals does not rely on any central limit theorem but just on the bootstrap principle that says that

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \text{ mimics } n^{1/2}(\hat{\theta} - \theta_0).$$

Important difference 2

Simulation based method : Need to compute

$$n^{1/2}(\hat{\theta}_b^* - \hat{\theta}), \qquad b = 1, \dots B$$

to approximate the root's law

Bootstrap vs asymptotic

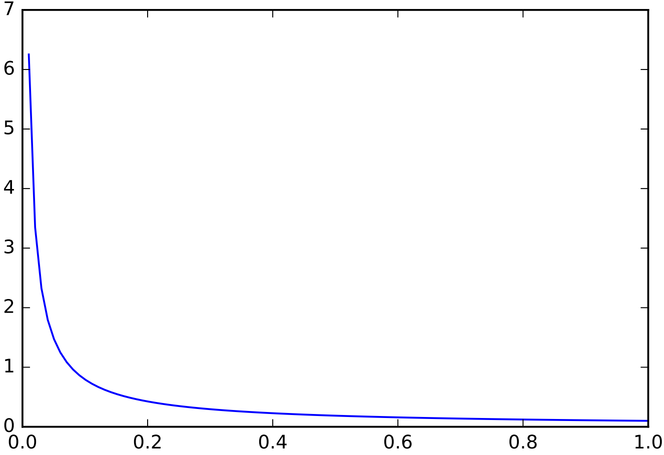


FIGURE: Plot of the density of the *beta*$(.1, 1)$ distribution

Bootstrap vs asymptotic

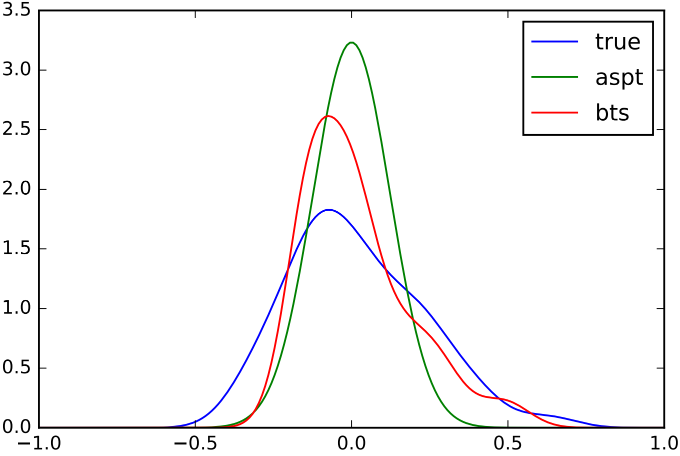


FIGURE: Plot of the true, the bootstrap and the asymptotic distribution of the root in the case of the mean with beta$(.1, 1)$ observations

Bootstrap vs asymptotic

```python
import numpy as np
from scipy.stats import gaussian_kde
from scipy.stats import norm
import matplotlib.pyplot as plt
```

```python
# Generation of the data
np.random.seed(1)
n = 20
a = .1
b = 1
X = np.random.beta(a, b, n)
```

Bootstrap vs asymptotic

```python
# Asymptotic
sigma = np.std(X)
x = .56
print(norm.pdf(x, loc=0, scale=sigma))
```

```python
# Bootstrap
B = 50
Xstarbarme = np.zeros([1, B])

for i in range(B):
    Xstar = X[np.random.randint(n, size=n)]
    Xstarbarme[:, i] = np.mean(Xstar)
Xstarbarme = np.sqrt(n) * (Xstarbarme - np.mean(X))
density_boot = gaussian_kde(Xstarbarme)
```

Quantiles of root

Let ξ_α denote the α-quantile of $n^{1/2}(\hat{\theta} - \theta_0)$

Quantiles are useful in ...

... building **confidence intervals**, *i.e.*,

$$\mathbb{P}\left(\theta_0 \in [\hat{\theta} - \xi_{1-\alpha/2}/n^{1/2}, \hat{\theta} - \xi_{\alpha/2}/n^{1/2}]\right) = 1 - \alpha.$$

...**testing**, *i.e.*, under $H_0 : \theta_0 = 1$

$$\mathbb{P}\left(n^{1/2}(\hat{\theta} - 1) \leqslant \xi_{\alpha/2} \textbf{ or } n^{1/2}(\hat{\theta} - 1) \geqslant \xi_{1-\alpha/2}\right) = \alpha$$

Exercise: Derive the previous equalities

Confidence intervals : Bootstrap vs asymptotic

Asymptotic :

$$\left[\hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}}\xi^{(\infty)}_{1-\frac{\alpha}{2}}, \hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}}\xi^{(\infty)}_{\frac{\alpha}{2}}\right]$$

where $\xi^{(\infty)}_{\alpha}$ is the α-quantile of the standard normal distribution and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Bootstrap :

$$\left[\hat{\theta}_n - \frac{1}{\sqrt{n}}\hat{\xi}_{B,1-\frac{\alpha}{2}}, \hat{\theta}_n - \frac{1}{\sqrt{n}}\hat{\xi}_{B,\frac{\alpha}{2}}\right]$$

where $\hat{\xi}_{B,\alpha}$ is a bootstrap estimator of the α-quantile of $n^{1/2}(\hat{\theta} - \theta_0)$ based on B-bootstrap samples

---

Exercise: Propose an algorithm to compute $\hat{\xi}_{B,\alpha}$

First conclusions

- The bootstrap is sample-based (no asymptotics)
- Easy to use :
  (i) no (mathematically involved) asymptotic theory
  (ii) embarrassingly parallel (but might need data copy)
  (iii) no need to estimate σ

Other examples

- Covariance
- Correlation coefficient
- Regression coefficient
- Testing the rank of a matrix
- etc.

Teaser

- Bootstrap is more accurate than asymptotics

The jackknife (the ancestor)

Origins : Quenouille (1949), Tukey (1958), Review : Miller (1974)

A leave-one-out procedure

1. Drop-off the i-th observation from the sample,

$$X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

   Compute : $\hat{\theta}_{-i} = \hat{\theta}(X_{-i}) = \hat{\theta}_{-i}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$

2. The Jackknife estimate of the bias is

$$\widehat{\mathrm{Bias}}_{\mathsf{jack}} = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta})$$

   The Jackknife estimate of the variance is

$$\widehat{\sigma^2}_{\mathsf{jack}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{-i} - \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{-i} \right)^2$$

Exercise

1. Give the Jackknife estimates of the bias and the variance in the case of the mean.
2. Suppose that $\mathbb{E}(\hat{\theta}) = \theta_0 + \frac{\theta_1}{n} + \frac{\theta_2}{n^2} + \ldots$, Show that the Jackknife improves the bias up to the order n^{-2}.

Other facts

1. Sometimes expressed in terms of the pseudo-values $n\hat{\theta} - (n-1)\hat{\theta}_{-i}$
2. The Jackknife is an approximation of the original bootstrap Beran (1984)
3. It works well for smooth transformations of the distribution
4. It fails for non-smooth transforms such as quantiles

Theorem (the Jackknife failure for the median)

Let $(X_1, \ldots, X_n)$ be *i.i.d.*random variables with law F and positive density f. Then

$$n\widehat{\sigma^2}_{\mathsf{jack}} \rightsquigarrow \frac{Y^2}{4f(F^{-1}(1/2))^2},$$

where $Y \sim \exp(1)$, whereas the asymptotic variance of $n^{1/2}(\hat{q}_{1/2} - q_{1/2})$ is

$$\frac{1}{4f(F^{-1}(1/2))^2}$$

Hint for the proof when $n = 2m$:
(i) $\widehat{\sigma^2}_{\mathsf{jack}} = \frac{(n-1)}{4}(X_{(m+1)} - X_{(m)})^2$,
(ii) use uniform variables and invariance Pyke (1965).

Delete d Jackknife

Notation : for $s \subset [\![1, n]\!]$, X_{-s} contains only the coordinates not in s. For a singleton $s = \{i\}$, we recover the previous notation.

Algorithm [SW89]

1. Drop-off d observations from the sample : compute the statistic $\hat{\theta}_{-s} = \hat{\theta}(X_{-s})$. Do this for all the possible d-uplet, *i.e.*, $j = 1, \dots, \binom{n}{d}$
2. The Jackknife estimate of the variance is

$$\hat{v}_{\text{jack-d}} = \frac{n-d}{d\binom{n}{d}} \sum_{s \subset [\![1,n]\!]} \left(\hat{\theta}_{-s} - \frac{1}{\binom{n}{d}} \sum_{s' \subset [\![1,n]\!]} \hat{\theta}_{-s'} \right)^2$$

Generalizing Efron's resampling scheme

Important remark (exercise)

Let $(X_1^*, \ldots, X_n^*)$ be a bootstrap sample. Then

$$\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^* = \sum_{i=1}^{n} w_{i,n} X_i$$

where $(w_{1,n}, \ldots, w_{n,n})$ is a random vector with multinomial distribution with parameter $1/n$

Natural question

Still working with other weights ?

The independent bootstrap

Consistency

Let $(w_1^*, \ldots, w_n^*)$ be *i.i.d.*random variables with mean 1 and variance equal to 1. If $\sigma^2 = \mathrm{var}(X_1) < +\infty$, then

$$n^{-1/2} \left(\sum_{i=1}^{n} (w_i - 1)(X_i - \bar{X}_n) \rightsquigarrow \mathcal{N}(0, \sigma^2) \right)$$

Hint : (1) replace $\bar{X}_n$ by EX (2) Apply Lindeberg's clt

Be careful !

The natural bootstrap estimator

$$n^{-1/2} \left(\sum_{i=1}^{n} w_i X_i - \bar{X}_n \right)$$

is not consistent

The Bayesian bootstrap

Consistency [Rub81]

Let $\xi_1, \dots, \xi_n$ be *i.i.d.*random variables with exponential distribution and mean 1. Let $\bar{\xi}_n = n^{-1} \sum_{i=1}^n \xi_i$ and define

$$X_{ni}^* = w_{ni} X_i, \qquad w_{ni} = \xi_i / \bar{\xi}$$

then

$$n^{1/2}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

Hint : (1) The previous equals $\frac{n^{-1/2}}{\bar{w_n}} \left(\sum_{i=1}^n (\xi_i - 1)(X_i - \bar{X}_n) \right)$ (2) Apply the previous result with the Delta-method

The exchangeably weighted bootstrap

Exchangeability

A random vector $(w_1, \ldots, w_n)$ is exchangeable when for any permutation $\sigma : [\![1, n]\!] \to [\![1, n]\!]$, $(w_1, \ldots, w_n)$ and $(w_{\sigma(1)}, \ldots, w_{\sigma(n)})$ have the same distribution.

Bootstrap consistency when [MN92] and [PW93]

1. for every $n \geqslant 1$, $(w_{1n}, \ldots, w_{nn})$ is exchangeable
2. $w_{n,i} \geqslant 0$, $i = 1, \ldots, n$, and $\sum_{i=1}^n w_{in} = n$, for every $n \geqslant 1$
3. as $n \to +\infty$,

$$\max_{1 \leqslant i \leqslant n} n^{-1} w_{in}^2 \to 0$$

$$n^{-1} \sum_{i=1}^n (w_{in} - 1)^2 \to 1$$

in probability

Parametric bootstrap

Be careful

Works only when the distribution of X_1 belongs to the model $\{\mathbb{P}_\theta : \theta \in \Theta\}$, *e.g.*, $\mathbb{P}_\theta = \mathcal{N}(\theta, 1)$

Algorithm

Estimate $\hat{\theta}_n$. Fix B the number of bootstrap iterations and initialize $b = 1$.

1. Draw independently $X_1^*, \ldots, X_n^*$ from $\mathbb{P}_{\hat{\theta}}$
2. Apply the same transformation

$$\hat{\theta}_b^* = \theta(X_1^*, \ldots, X_n^*)$$

3. Stop if $b = B$ else iterate.

Then

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \textbf{ mimics } n^{1/2}(\hat{\theta} - \theta_0)$$

Delta-method

Most of the estimators are not empirical means over the observed values $(X_1, \ldots, X_n)$ but transformations of empirical means, e.g., the covariance, the correlation, the estimated regression vector in linear regression. For the covariance :

$$\overline{xy} - \overline{x}\,\overline{y} = g(\overline{xy}, \overline{x}, \overline{y}), \qquad g(a, b, c) = a - bc$$

Informal statement

Whenever we are able to bootstrap the empirical mean, we shall also be able to bootstrap "smooth" transformations

Delta-method

If g is differentiable at $\mu_0 = E[X_1]$ and $\Sigma = Var(X_1) < +\infty$

$$n^{1/2}\left(g(\bar{X}_n) - g(\mu_0)\right) \rightsquigarrow \mathcal{N}\left(0,\, V\right)$$

with $V = \nabla g(\mu_0)^T\, \Sigma\, \nabla g(\mu_0)$

Delta-method (bootstrap version)

If g is differentiable at μ_0 and $n^{1/2}(\overline{X^*} - \overline{X}) \rightsquigarrow \mathcal{N}\left(0, \Sigma\right)$

$$n^{1/2}\left(g(\overline{X^*}) - g(\bar{X}_n)\right) \rightsquigarrow \mathcal{N}\left(0,\, V\right)$$

with $\overline{X^*} = n^{-1}\sum_{i=1}^n w_{ni} X_i$

M-estimators

Another interesting class of estimators is when $\hat{\theta}$ is defined by*

$$\hat{\theta} \in \arg\min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta)$$

Often** we have $n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}(0, v)$ for some θ_0

M-estimation bootstrap

If $\quad \hat{\theta}^* \in \arg\min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} w_{ni} m(X_i, \theta)$

then** $\quad n^{1/2}(\hat{\theta}^* - \hat{\theta}) \rightsquigarrow \mathcal{N}(0, V)$

*e.g. med-LS, OLS, WLS, MLE

** technical conditions in [WZ96]

Computational cost

Definition

A statistic is pivotal when the limiting distribution does not depend on $\mathbb{P}$

Examples

- the mean, $n^{1/2}\left(\frac{\bar{X}-EX}{\hat{\sigma}}\right)$ with $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- the cdf, $n^{1/2}\left(\frac{\hat{F}(x)-F(x)}{\hat{F}(x)^{1/2}(1-\hat{F}(x))^{1/2}}\right)$ with $\hat{F}(x) = n^{-1}\sum_{i=1}^{n} 1_{\{X_i \leqslant x\}}$

Requirement

To obtain a pivotal root, one needs an **estimate of the variance**

The t-bootstrap

Idea

Basic bootstrap : $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ mimics $n^{1/2}(\hat{\theta}^* - \theta_0)$

t-bootstrap* : $n^{1/2}\left(\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}\right)$ mimics $n^{1/2}\left(\frac{\hat{\theta} - \theta_0}{\hat{\sigma}}\right)$

Approximation**

When $n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma)$ with cdf Φ

- **Asymptotic :** $|\Phi(y) - \mathbb{P}(n^{1/2}(\hat{\theta} - \theta_0) \leqslant y)| \simeq \frac{C}{\sqrt{n}}$
- **Basic bootstrap :**
  $|\mathbb{P}_*(n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leqslant y) - \mathbb{P}(n^{1/2}(\hat{\theta} - \theta_0) \leqslant y)| \simeq \frac{C}{\sqrt{n}}$
- **t-bootstrap :**
  $|\mathbb{P}_*(n^{1/2}\left(\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}\right) \leqslant y) - \mathbb{P}(n^{1/2}\left(\frac{\hat{\theta} - \theta_0}{\hat{\sigma}}\right) \leqslant y)| \simeq \frac{C}{n}$

*t is for studentization

**Based on Edgeworth expansion [Hal92]

Confidence interval

$\xi_\alpha^{(\infty)}$: α-quantile of $\mathcal{N}(0,1)$

$\hat{\xi}_{B,\alpha}^{(tb)}$: α-quantile of $\sqrt{n}\left(\frac{\hat{\theta}*-\hat{\theta}}{\hat{\sigma}*}\right)$

$\hat{\xi}_{B,\alpha}^{(bb)}$: α-quantile of $\sqrt{n}(\hat{\theta}*-\hat{\theta})$

$\hat{q}_\alpha$: α-quantile of $\hat{\theta}*$

	formulas	accuracy
asymp.	$\left[\hat{\theta}-\frac{\hat{\sigma}}{\sqrt{n}}\xi_{1-\alpha/2}^{(\infty)},\ \hat{\theta}-\frac{\hat{\sigma}}{\sqrt{n}}\xi_{\alpha/2}^{(\infty)}\right]$	$n^{-1/2}$
basic boot.	$\left[\hat{\theta}-\frac{1}{\sqrt{n}}\hat{\xi}_{1-\alpha/2}^{(bb)},\ \hat{\theta}-\frac{1}{\sqrt{n}}\hat{\xi}_{\alpha/2}^{(bb)}\right]$	$n^{-1/2}$
t-boot.	$\left[\hat{\theta}-\frac{\hat{\sigma}}{\sqrt{n}}\hat{\xi}_{1-\alpha/2}^{(tb)},\ \hat{\theta}-\frac{\hat{\sigma}}{\sqrt{n}}\hat{\xi}_{\alpha/2}^{(tb)}\right]$	n^{-1}
percentile boot.	$\left[\hat{q}_{\alpha/2},\ \hat{q}_{1-\alpha/2}\right]$	$n^{-1/2}$

Remarks

- no variance estimation for the basic and the percentile
- The more accurate is the t-bootstrap
- the percentile is simple (invariance) and gives intervals in the range of θ

Computational cost

Bootstrap is computationally intensive :

- (**approximation step**) For some B, compute $\hat{R}_1^*, \ldots, \hat{R}_B^*$ and approximate the law of $\hat{R}$

Choice of B

- For procedures with accuracy $1/\sqrt{n}$, B should be at least equal to n
- For procedures with accuracy $1/n$, B should be at least equal to n^2

Regression model

$$Y = g(X) + \sigma(X)\epsilon$$

- X is random, *i.e.*, random design (ϵ and X independent)
- X is non-random, *i.e.*, deterministic design

goal : To estimate g

particular semiparametric problem $\Rightarrow$ particular bootstrap

2 bootstrap strategies

- Classical bootstrap : bootstrap of the pairs or M-estimation bootstrap

  $\Rightarrow$ OK for **random design**
- Bootstrap of the residuals

  $\Rightarrow$ OK for **random and deterministic design**

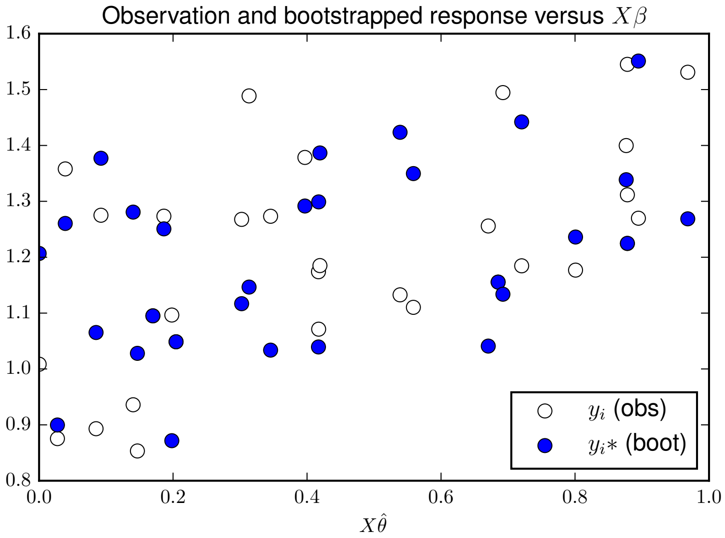
Bootstrap of the residuals

Algorithm

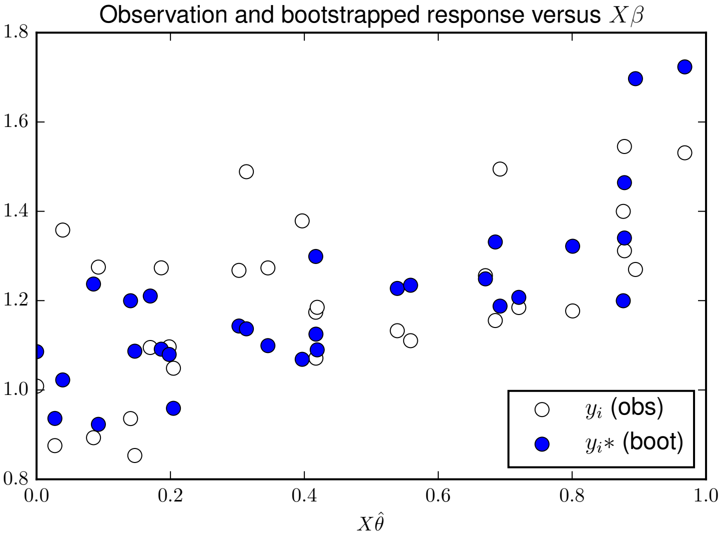
From the sample $(Y_1, X_1, \ldots, Y_n, X_n)$ compute $\hat{g}$ and the estimated residuals $\hat{\epsilon}_i = Y_i - \hat{g}(X_i)$. Initialize $b = 1$

1. Draw uniformly with replacement among $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$. It gives
$$(\hat{\epsilon}_1^*, \ldots, \hat{\epsilon}_n^*)$$
2. For $i = 1, \ldots, n$, compute the bootstrap response
$$Y_i^* = \hat{g}(X_i) + \hat{\epsilon}_i^*$$
3. From the sample $(Y_1^*, X_1, \ldots, Y_n^*, X_n)$ compute $\hat{g}_b^*$
4. Stop if $b = B$ else iterate

Bootstrap of the residuals



Bootstrap of the residuals



syllabus I

P. Bertail, Université Paris Ouest (see webpage)
L. Simard Université catholique de Louvain
J. Wellner, University of Washington (see webpage)

References I

- Rudolf Beran.
  Jackknife approximations to bootstrap estimates.
  Ann. Statist., 12(1) :101–118, 1984.
- B. Efron.
  Bootstrap methods : another look at the jackknife.
  Ann. Statist., 7(1) :1–26, 1979.
- Bradley Efron and Robert J. Tibshirani.
  An introduction to the bootstrap, volume 57 of *Monographs on Statistics and Applied Probability*.
  Chapman and Hall, New York, 1993.
- Peter Hall.
  The bootstrap and Edgeworth expansion.
  Springer Series in Statistics. Springer-Verlag, New York, 1992.

References II

- Rupert G. Miller.
  The jackknife—a review.
  Biometrika, 61 :1–15, 1974.
- David M. Mason and Michael A. Newton.
  A rank statistics approach to the consistency of a general bootstrap.
  Ann. Statist., 20(3) :1611–1624, 1992.
- Jens Præstgaard and Jon A. Wellner.
  Exchangeably weighted bootstraps of the general empirical process.
  Ann. Probab., 21(4) :2053–2086, 1993.
- R. Pyke.
  Spacings. (With discussion.).
  J. Roy. Statist. Soc. Ser. B, 27 :395–449, 1965.

References III

- M. H. Quenouille.
  Approximate tests of correlation in time-series.
  J. Roy. Statist. Soc. Ser. B., 11 :68–84, 1949.
- Donald B. Rubin.
  The Bayesian bootstrap.
  Ann. Statist., 9(1) :130–134, 1981.
- Jun Shao and C.F.J Wu.
  A general theory for jackknife variance estimation.
  Ann. Statist., 17(3) :1176–1197, 1989.
- John W. Tukey.
  Bias and confidence in not quite large samples (abstract).
  Annals of Mathematical Statistics, 29 :614, 1958.

References IV

- Jon A Wellner and Yihui Zhan.
  Bootstrapping z-estimators.
  University of Washington Department of Statistics Technical Report, 308, 1996.