MS BGD
MDI 720 : Statistiques

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

Bootstrap : le principe général

But

mesurer le degré de précision d'un estimateur $\hat{\theta}$

Algorithme

monde réel	monde Bootstrap
$(X_1, \dots X_n) \sim \mathbb{P}$ **(inconnue)**	$(X_1^*, \dots X_n^*) \sim \hat{\mathbb{P}}$ **(connue)**
$\downarrow$	$\downarrow$
$\hat{\theta}$	$\hat{\theta}^*$
$\hat{\theta}$ provient de données sous $\mathbb{P}$	$\hat{\theta}^*$ provient de données sous $\widehat{\mathbb{P}}$ (qui approche $\mathbb{P}$)

Idée de base

$\hat{\theta}^*$ (**connu**) reproduit le comportement de $\hat{\theta}$ (**inconnu**)

L'algorithme

- $X_1, \ldots, X_n$ i.i.d.
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ une statistique/estimateur d'intérêt
  Examples : moyenne empirique $\bar{X}_n$ ou médiane $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, nombre d'itérations B
Output : Estimateur Bootstrap $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$

L'algorithme

- $X_1, \ldots, X_n$ i.i.d.
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ une statistique/estimateur d'intérêt
  Examples : moyenne empirique $\bar{X}_n$ ou médiane $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, nombre d'itérations B
Output : Estimateur Bootstrap $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$
pour $b = 1, \ldots, B$ **faire**

L'algorithme

- $X_1, \ldots, X_n$ i.i.d.
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ une statistique/estimateur d'intérêt
  Examples : moyenne empirique $\bar{X}_n$ ou médiane $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, nombre d'itérations B
Output : Estimateur Bootstrap $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$
pour $b = 1, \ldots, B$ **faire**
(i) Tirer (uniformément) avec remise dans $X_1, \ldots, X_n$ afin d'obtenir un nouvelle échantillon (**aléatoire**) :

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

L'algorithme

- $X_1, \ldots, X_n$ i.i.d.
- $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ une statistique/estimateur d'intérêt
  Examples : moyenne empirique $\bar{X}_n$ ou médiane $\mathrm{Med}_n(X_1, \ldots, X_n)$

Algorithme : Bootstrap

Input : $X_1, \ldots, X_n$, nombre d'itérations B
Output : Estimateur Bootstrap $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$
pour $b = 1, \ldots, B$ **faire**

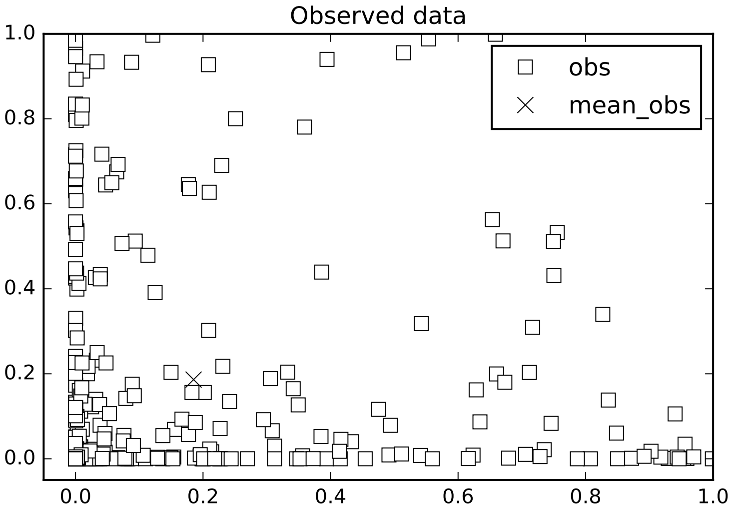
(i) Tirer (uniformément) avec remise dans $X_1, \ldots, X_n$ afin d'obtenir un nouvelle échantillon (**aléatoire**) :

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

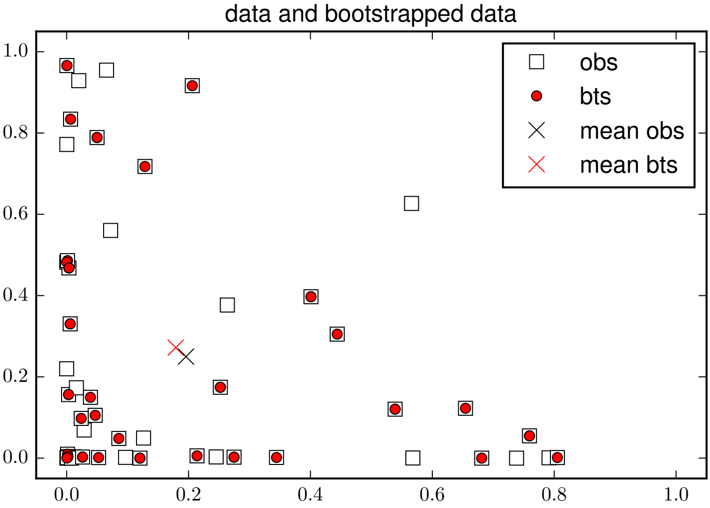
(ii) Calculer l'estimateur/la statistique sur cet échantillon

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

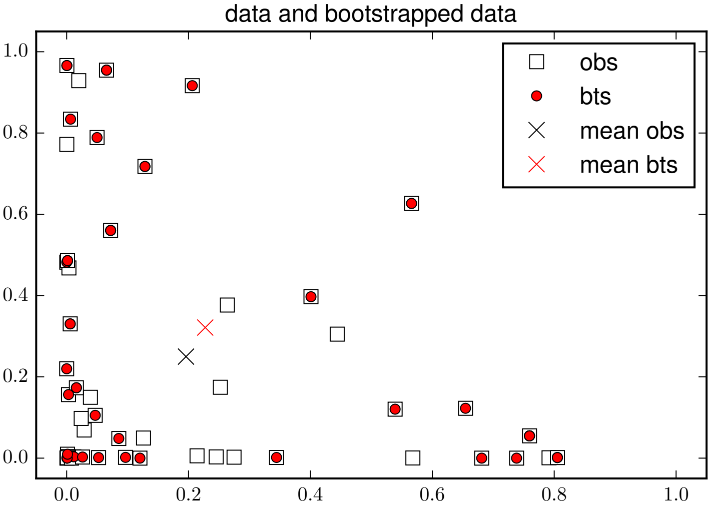
Le bootstrap pour la moyenne



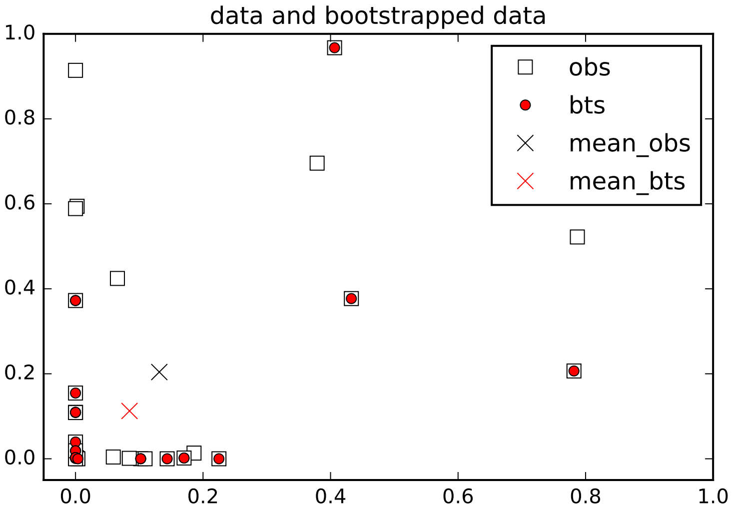
Le bootstrap pour la moyenne



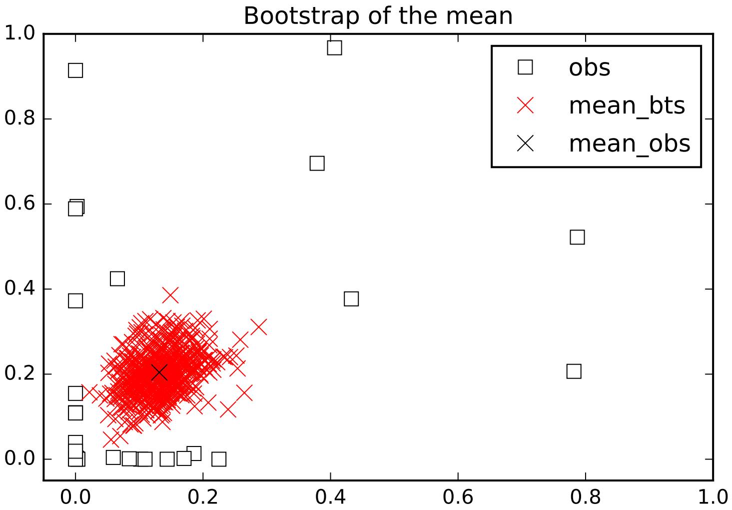
Le bootstrap pour la moyenne



Le bootstrap pour la moyenne



Le bootstrap pour la moyenne

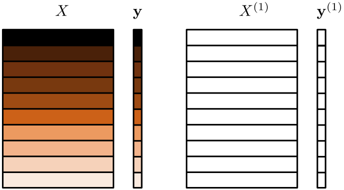


Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$

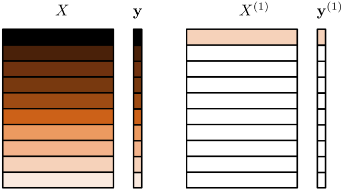
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



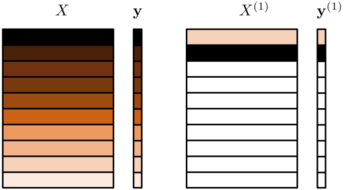
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



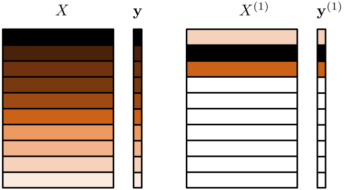
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



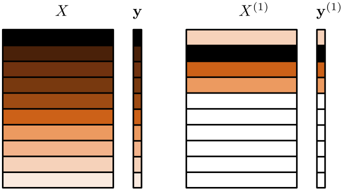
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



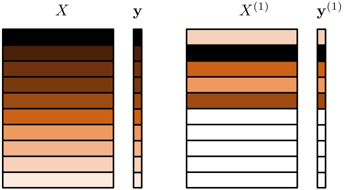
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



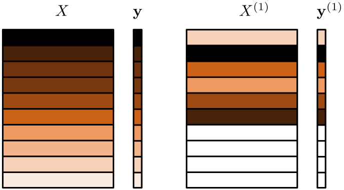
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



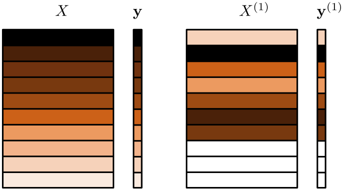
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



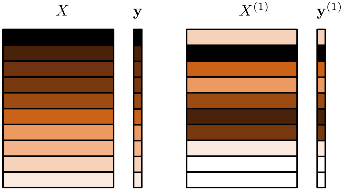
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



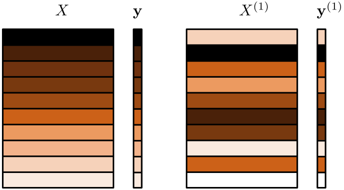
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



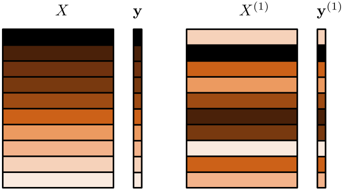
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



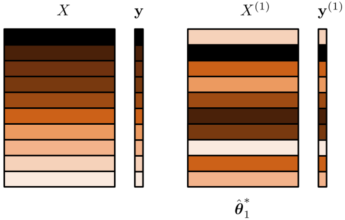
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



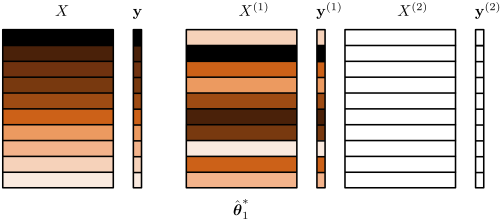
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



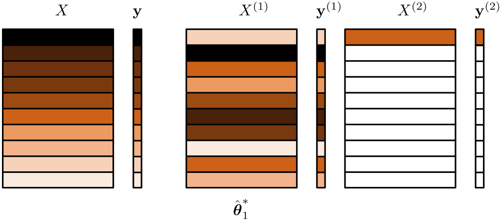
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



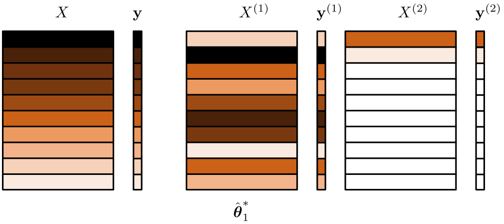
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



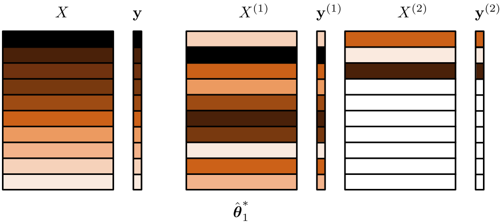
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



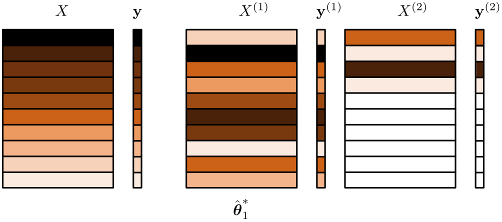
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



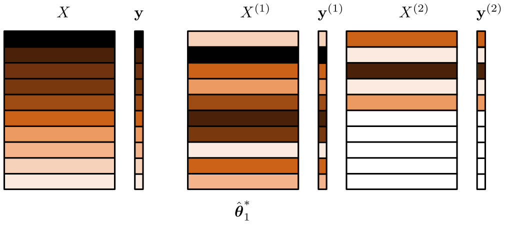
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



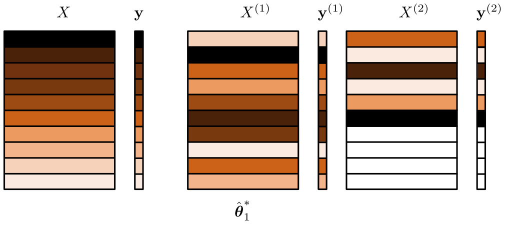
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



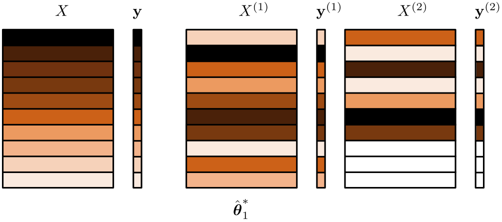
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



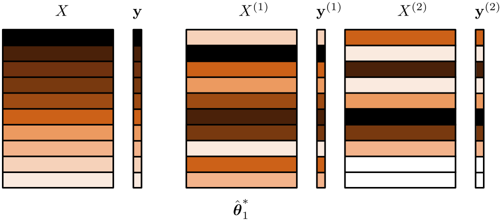
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



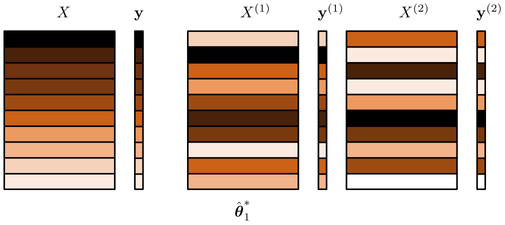
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



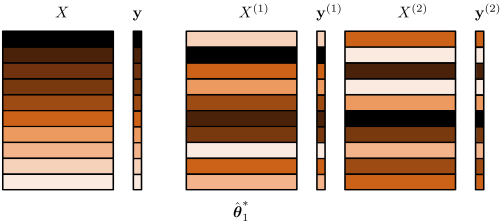
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



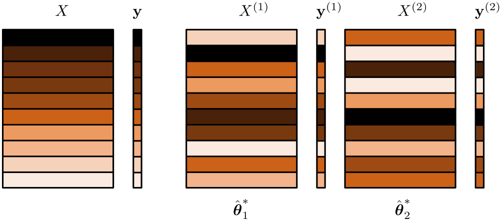
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



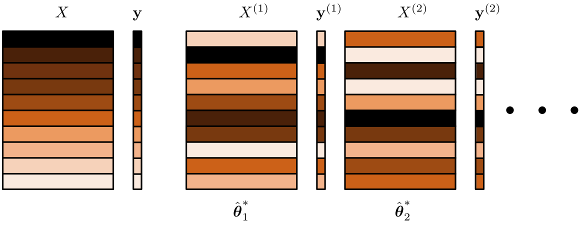
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



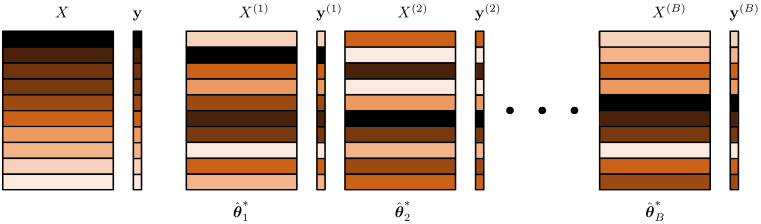
Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



Bootstrap (des paires) en régression

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$



Premiers estimateurs bootstrap

Soit θ_0 le paramètre d'intérêt (inconnu)

	le vrai (inconnu)	le bootstrap
biais	$\mathbb{E}[\hat{\theta}] - \theta_0$	$B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^* - \hat{\theta}$
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	...	...
density	...	...

Les statistiques $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ sont des "versions" bootstrap de $\hat{\theta}$

Comment les utiliser ?

Origine Efron et Tibshirani (1993)

Le terme “bootstrap" provient de la phrase :

“to pull oneself up by one’s own bootstrap” (réussir par soi-même)

issue de “The Surprising Adventures of Baron Munchausen” by *R. E. Raspe* (18th century).

Idée Efron (1979)

En se basant uniquement sur les données observées, reproduire la distribution d’une statistique, *e.g.*, moyennes, écart-type, corrélation, etc...

Pas de théorie asymptotique !

Racine statistique

Definition

Une **racine statistique** $\hat{R}$ est une fonction de $(X_1, \ldots, X_n)$ qui converge en distribution vers G *i.e.*, $\hat{R} \rightsquigarrow G$

Examples

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

- la moyenne, $n^{1/2}(n^{-1} \sum_{i=1}^{n} X_i - 1/2)$
- la cdf, $n^{1/2}(n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leqslant x\}} - x)$
- le minimum, $n(\min_{1 \leqslant i \leqslant n} X_i)$
- la variance, ...

Notre contexte : le cas régulier

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

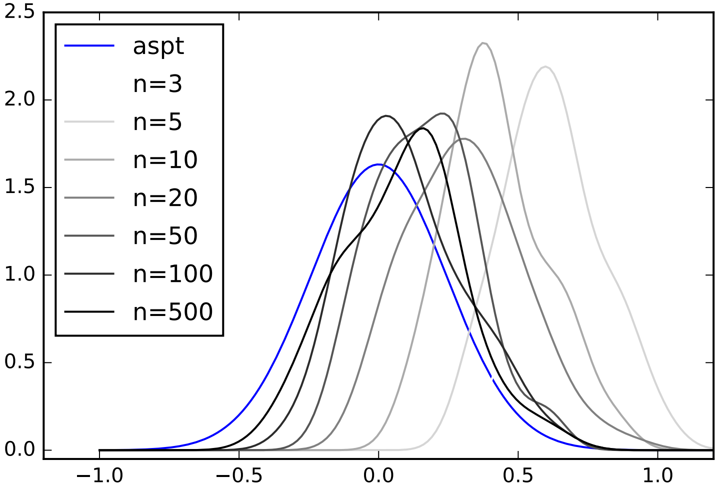


FIGURE – Exemple d'une racine dont le biais est positif

Bootstrapper une racine

La racine choisie $\hat{R}$ est donnée par le problème considéré

But du bootstrap

reproduire le "comportement" d'une racine statistique

2 étapes :

- (**étape de définition**)* Trouver la racine bootstrap $\hat{R}^*$ qui reproduit la racine d'intérêt $\hat{R}$
- (**étape d'approximation**)** Pour B (grand), calculer $\hat{R}^*_1, \ldots, \hat{R}^*_B$ et approcher la loi de $\hat{R}$

*pour l'étape de définition : bon sens et théorie asymptotique

**pour l'étape d'approximation : simulation de Monte Carlo

Exemples

Example 1 : La moyenne

Supposons

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

D'après le TCL, si $\mathbb{E}[X_1^2] < +\infty$, alors

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

Une version bootstrap pour $\hat{R}$ est

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

Exemples

Exemple 2 : la variance

Soient

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

Si $\mathbb{E}[X_1^4] < +\infty$, alors

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

avec $v = \mathrm{var}((X - \mathbb{E}[X])^2)$. Une version bootstrap de $\hat{R}$ est

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

Quantiles d'une racine

Soit ξ_α le α-quantile de $n^{1/2}(\hat{\theta} - \theta_0)$

Les quantiles sont très utiles pour...

... créer des **intervalles de confiance**, *i.e.*,

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

...faire des **tests**, *i.e.*, sous $H_0 : \theta_0 = 1$

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

Exercise: Obtenir les égalités précédentes pour la moyenne

Intervalle de confiance : Bootstrap vs asymptotique

Asymptotique :

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

ou $\xi^{(\infty)}_{\alpha}$ est le α-quantile d'une loi normale standardisée et $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Bootstrap :

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

ou $\hat{\xi}_{B,\alpha}$ est un estimateur bootstrap du α-quantile de $n^{1/2}(\hat{\theta} - \theta_0)$ basé sur B échantillon bootstrap

---

Exercise: Proposer un algorithme pour calculer $\hat{\xi}_{B,\alpha}$

Bootstrap vs asymptotique

But

la distribution **inconnue** de

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

2 possibilités

La distribution **asymptotique** (estimée), *i.e.,*

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

La distribution **bootstrap**, *i.e.,* la distribution de

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

Bootstrap vs asymptotique

Différence importante 1

L'utilisation du bootstrap ne requiert aucune considération théorique comme le calcul, parfois difficile de la loi asymptotique (e.g., intervalle de confiance pour la variance).

Différence importante 2

Le bootstrap est basé sur la simulation de

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

pour approcher la loi cible de $\hat{R}$.

Bootstrap vs asymptotic

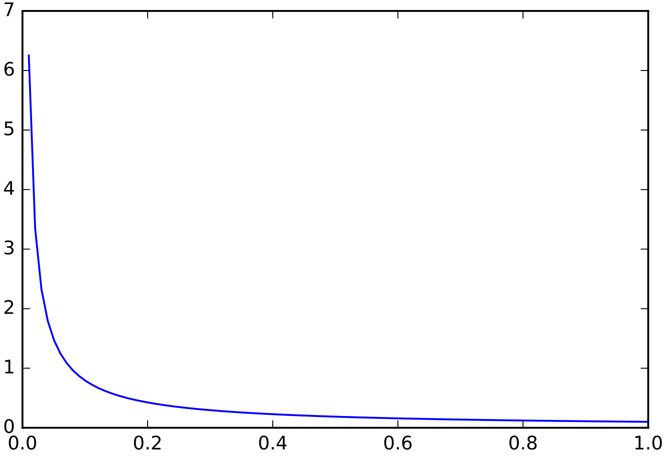


FIGURE – Graphe de la densité d'une *beta*$(.1, 1)$

Bootstrap vs asymptotic

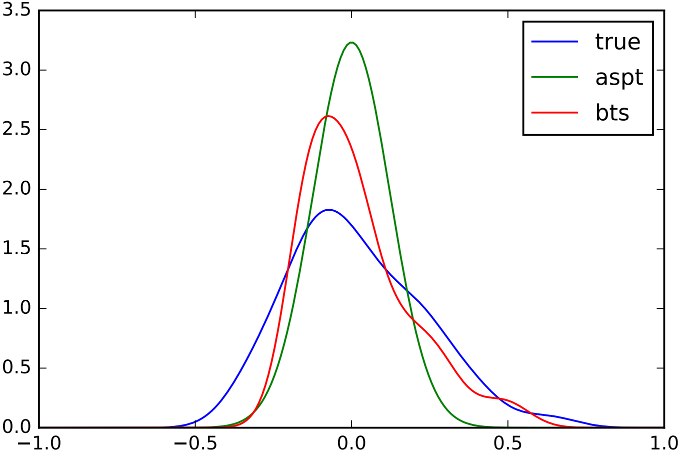


FIGURE – Graphe de la distribution théorique (la vraie), bootstrap et asymptotique de la racine dans le cas de la moyenne d'une beta$(.1, 1)$

Bootstrap vs asymptotique

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

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

Bootstrap vs asymptotique

```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)
```

1er conclusions

- Le bootstrap est “sample-based” (pas de théorie asymptotique)
- Facile à utiliser :
  (i) pas de théorie asymptotique
  (ii) parallélisable
  (iii) pas besoin d’estimer σ

Autres exemples

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

Racine pivotale

Definition

Une statistique est pivotale lorsque la distribution limite ne dépend pas de $\mathbb{P}$

Exemples

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

Requirement

Pour obtnir une statistique pivotale, on doit **estimer la variance**

Le t-bootstrap

Idée

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

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

Approximation**

Supposons que $n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma)$ dont la cdf est noté Φ

- **Asymptotique :** $|\Phi(y) - \mathbb{P}(n^{1/2}(\hat{\theta} - \theta_0) \leqslant y)| \simeq \frac{C}{\sqrt{n}}$
- **bootstrap basique :**
  $|\mathbb{P}_*(n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leqslant y) - \mathbb{P}(n^{1/2}(\hat{\theta} - \theta_0) \leqslant y)| \simeq \frac{C}{\sqrt{n}}$
- **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 pour studentization

**Basée sur des expansions d'Edgeworth [Hal92]

Confidence interval

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

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

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

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

	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.	$\left[\hat{q}_{\alpha/2},\ \hat{q}_{1-\alpha/2}\right]$	$n^{-1/2}$

Remarques

- pas d'estimation de la variance pour le bootstrap basique et le percentile
- le plus précis est le t-bootstrap
- le percentile est simple et donne des intervalles dans l'image de θ

Le Bootstrap est très demandeur en temps de calcul

- (**l'étape d'approximation**) Calculer $\hat{R}_1^*, \ldots, \hat{R}_B^*$ et approcher la loi de $\hat{R}$

Choice of B

- Les procédures dont la précision est $1/\sqrt{n}$: B devra être de l'ordre de n
- Les procédures dont la précision est $1/n$: B devra être de l'ordre de n^2

Model de régression

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

- X est aléatoire, *i.e.*, "random design" (ϵ et X indépendants)
- X n'est pas aléatoire, *i.e.*, "deterministic design"

But : estimer g

particular semiparametric problem $\Rightarrow$ particular bootstrap

2 stratégies pour le bootstrap

- bootstrap classique : bootstrap des paires
  $\Rightarrow$ OK pour **"random design"**
- Bootstrap des résidus
  $\Rightarrow$ OK pour **"random" et "deterministic design"**

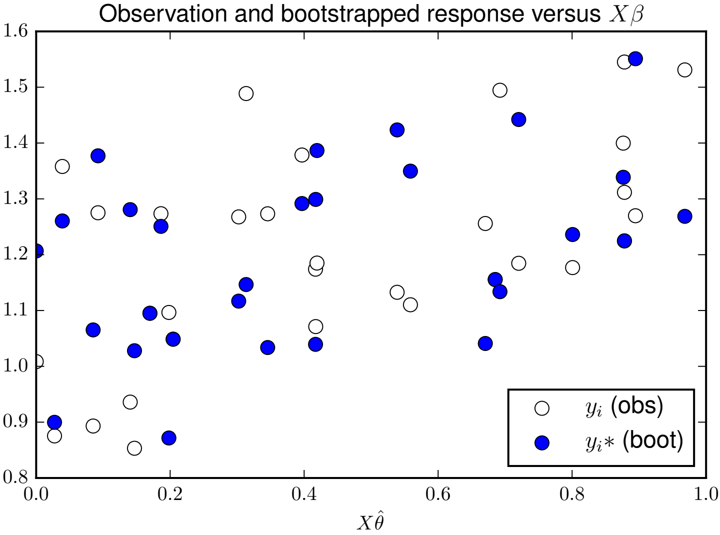
Bootstrap des résidus

Algorithme

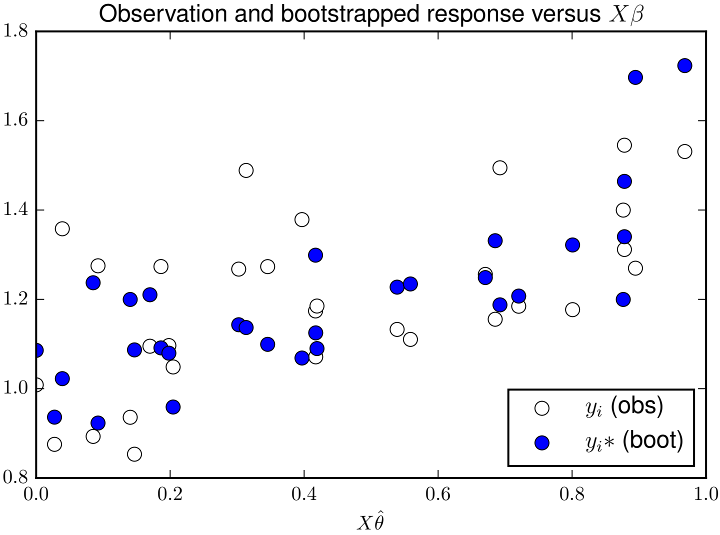
Soient $(Y_1, X_1, \dots, Y_n, X_n)$. Calculer $\hat{g}$ et les résidus estimés $\hat{\epsilon}_i = Y_i - \hat{g}(X_i)$. Initialiser $b = 1$

1. Tirer uniformément avec remise dans $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$. Cela donne
$$(\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*)$$
2. Pour $i = 1, \dots, n$, calculer
$$Y_i^* = \hat{g}(X_i) + \hat{\epsilon}_i^*$$
3. A partir de $(Y_1^*, X_1, \dots, Y_n^*, X_n)$, calculer $\hat{g}_b^*$
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

- 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.