chapter 2 :
the bootstrap method

- Introduction
- Glivenko-Cantelli Theorem
- The Monte Carlo method
- Bootstrap
- Parametric Bootstrap

Motivating example

Case of a random event with binary (Bernoulli) outcome $Z \in \{0, 1\}$ such that $\mathbb{P}(Z = 1) = p$

Observations $z_1, \ldots, z_n$ (iid) put to use to approximate p by

$$\hat{p} = \hat{p}(z_1, \ldots, z_n) = {}^1\!/_n \sum_{i=1}^{n} z_i$$

Illustration of a (moment/unbiased/maximum likelihood) estimator of p

intrinsic statistical randomness

$\boxed{\textbf{inference based on a random sample implies uncertainty}}$

Since it depends on a **random** sample, an estimator

$$\delta(X_1, \ldots, X_n)$$

also is a **random** variable

Hence "error" in the reply: an estimator produces a different estimation of the same quantity θ each time a new sample is used (data does produce the model)

intrinsic statistical randomness

inference based on a random sample implies uncertainty

Since it depends on a **random** sample, an estimator

$$\delta(X_1, \ldots, X_n)$$

also is a **random** variable

Hence "error" in the reply: an estimator produces a different estimation of the same quantity θ each time a new sample is used (data does produce the model)

intrinsic statistical randomness

inference based on a random sample implies uncertainty

Since it depends on a **random** sample, an estimator

$$\delta(X_1, \ldots, X_n)$$

also is a **random** variable

Hence "error" in the reply: an estimator produces a different estimation of the same quantity θ each time a new sample is used (data does produce the model)

infered variation

inference based on a random sample implies uncertainty

Question 1 :

How much does $\delta(X_1, \ldots, X_n)$ vary when the sample varies?

Question 2 :

What is the variance of $\delta(X_1, \ldots, X_n)$?

Question 3 :

What is the distribution of $\delta(X_1, \ldots, X_n)$?

infered variation

inference based on a random sample implies uncertainty

Question 1 :

How much does $\delta(X_1, \ldots, X_n)$ vary when the sample varies?

Question 2 :

What is the variance of $\delta(X_1, \ldots, X_n)$?

Question 3 :

What is the distribution of $\delta(X_1, \ldots, X_n)$?

infered variation

inference based on a random sample implies uncertainty

Question 1 :

How much does $\delta(X_1, \ldots, X_n)$ vary when the sample varies?

Question 2 :

What is the variance of $\delta(X_1, \ldots, X_n)$?

Question 3 :

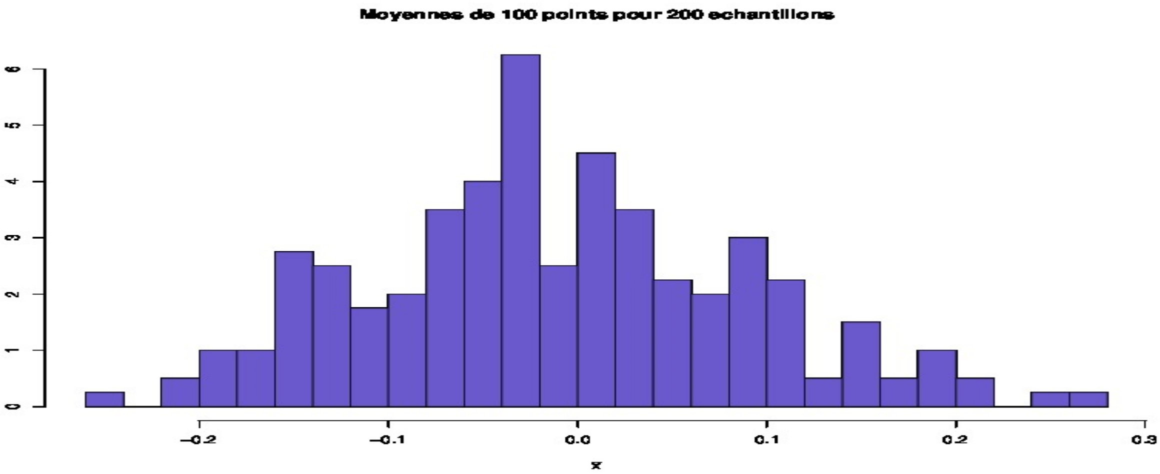
What is the distribution of $\delta(X_1, \ldots, X_n)$?

infered variation

Example (**Normal sample)**

Take $X_1, \ldots, X_{100}$ a random sample from $\mathcal{N}(\theta, 1)$. Its mean θ is estimated by

$$\hat{\theta} = \frac{1}{100} \sum_{i=1}^{100} X_i$$



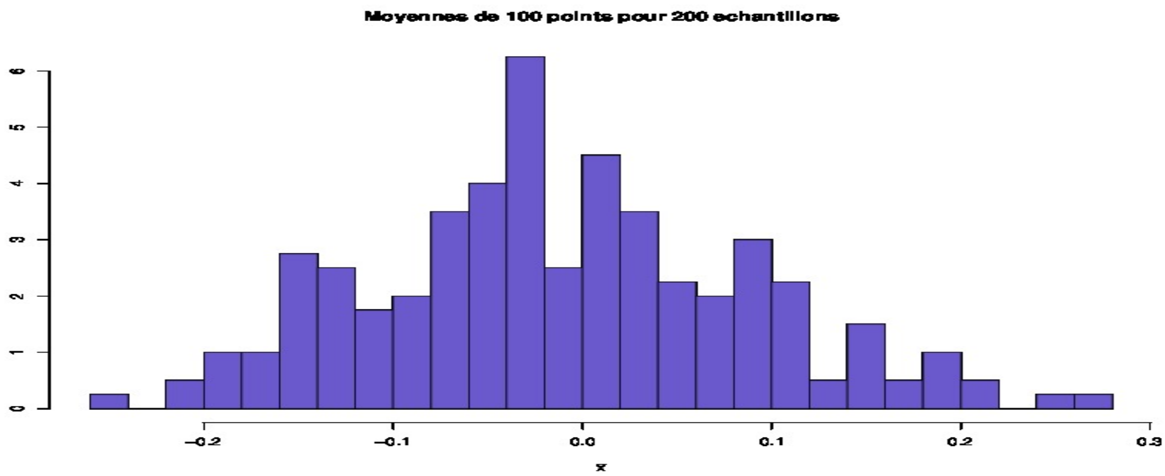
Variation compatible with the (known) theoretical distribution $\hat{\theta} \sim \mathcal{N}(\theta, 1/100)$

infered variation

Example (Normal sample)

Take $X_1, \ldots, X_{100}$ a random sample from $\mathcal{N}(\theta, 1)$. Its mean θ is estimated by

$$\hat{\theta} = \frac{1}{100} \sum_{i=1}^{100} X_i$$



Variation compatible with the (known) theoretical distribution $\hat{\theta} \sim \mathcal{N}(\theta, 1/100)$

Associated difficulties (illustrations)

- Observation of a **single** sample $x_1, \ldots, x_n$ in most cases
- The sampling distribution F is often unknown
- The evaluation of the average variation of $\delta(X_1, \ldots, X_n)$ is paramount for the construction of confidence intervals and for testing/answering questions like

$$H_0 : \theta \leqslant 0$$

- In the **normal** case, the **true** θ stands with high probability in the interval

$$[\hat{\theta} - 2\sigma, \hat{\theta} + 2\sigma] .$$

$\boxed{\textbf{Quid of } \sigma \textbf{ ?!}}$

Associated difficulties (illustrations)

- Observation of a **single** sample $x_1, \ldots, x_n$ in most cases
- The sampling distribution F is often unknown
- The evaluation of the average variation of $\delta(X_1, \ldots, X_n)$ is paramount for the construction of confidence intervals and for testing/answering questions like

$$H_0 : \theta \leqslant 0$$

- In the **normal** case, the **true** θ stands with high probability in the interval

$$[\hat{\theta} - 2\sigma, \hat{\theta} + 2\sigma] .$$

Quid of σ ?!

Associated difficulties (illustrations)

- Observation of a **single** sample $x_1, \ldots, x_n$ in most cases
- The sampling distribution F is often unknown
- The evaluation of the average variation of $\delta(X_1, \ldots, X_n)$ is paramount for the construction of confidence intervals and for testing/answering questions like

$$H_0 : \theta \leqslant 0$$

- In the **normal** case, the **true** θ stands with high probability in the interval

$$[\hat{\theta} - 2\sigma, \hat{\theta} + 2\sigma] .$$

$\boxed{\textbf{Quid of } \sigma \textbf{ ?!}}$

Associated difficulties (illustrations)

- Observation of a **single** sample $x_1, \ldots, x_n$ in most cases
- The sampling distribution F is often unknown
- The evaluation of the average variation of $\delta(X_1, \ldots, X_n)$ is paramount for the construction of confidence intervals and for testing/answering questions like

$$H_0 : \theta \leqslant 0$$

- In the **normal** case, the **true** θ stands with high probability in the interval

$$[\hat{\theta} - 2\sigma, \hat{\theta} + 2\sigma] \,.$$

$\boxed{\textbf{Quid of } \sigma \textbf{ ?!}}$

Associated difficulties (illustrations)

- Observation of a **single** sample $x_1, \ldots, x_n$ in most cases
- The sampling distribution F is often unknown
- The evaluation of the average variation of $\delta(X_1, \ldots, X_n)$ is paramount for the construction of confidence intervals and for testing/answering questions like

$$H_0 : \theta \leqslant 0$$

- In the **normal** case, the **true** θ stands with high probability in the interval

$$[\hat{\theta} - 2\sigma, \hat{\theta} + 2\sigma] .$$

Quid of σ ?!

Estimation of the repartition function

Extension/application of the LLN to the approximation of the cdf:

For an i.i.d. sample $X_1, \ldots, X_n$, empirical cdf

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{]-\infty, x]}(X_i) \\ &= \frac{\text{card}\{X_i;\, X_i \leqslant x\}}{n}, \end{aligned}$$

Step function corresponding to the empirical distribution

$$1/n \sum_{i=1}^{n} \delta_{X_i}$$

where δ Dirac mass

Estimation of the repartition function

Extension/application of the LLN to the approximation of the cdf:
For an i.i.d. sample $X_1, \ldots, X_n$, empirical cdf

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{]-\infty, x]}(X_i) \\ &= \frac{\text{card}\,\{X_i;\, X_i \leqslant x\}}{n}, \end{aligned}$$

Step function corresponding to the empirical distribution

$$1/n \sum_{i=1}^{n} \delta_{X_i}$$

where δ Dirac mass

convergence of the empirical cdf

Glivenko-Cantelli Theorem

$$\|\widehat{F}_n - F\|_\infty = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0$$

[Glivenko, 1933;Cantelli, 1933]

$\widehat{F}_n(x)$ **is a convergent estimator of the cdf** $F(x)$

convergence of the empirical cdf

Dvoretzky–Kiefer–Wolfowitz inequality

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}}\big|\widehat{F}_n(x) - F(x)\big| > \varepsilon\Big) \leqslant e^{-2n\varepsilon^2}$$

for every $\varepsilon \geqslant \varepsilon_n = \sqrt{1/2n \ln 2}$

[Massart, 1990]

$\widehat{F}_n(x)$ **is a convergent estimator of the cdf** $F(x)$

convergence of the empirical cdf

Donsker's Theorem

The sequence

$$\sqrt{n}(\hat{F}_n(x) - F(x))$$

converges in distribution to a Gaussian process G with zero mean and covariance

$$\mathrm{cov}[G(s), G(t)] = \mathbb{E}[G(s)G(t)] = \min\{F(s), F(t)\} - F(s)F(t).$$

[Donsker, 1952]

$\hat{F}_n(x)$ **is a convergent estimator of the cdf** $F(x)$

statistical consequences of Glivenko-Cantelli

Moments

$$\mathbb{E}[\widehat{F}_n(x)] = F(x)$$

$$\mathrm{var}[\widehat{F}_n(x)] = \frac{F(x)(1 - F(x))}{n}$$

statistical consequences of Glivenko-Cantelli

Confidence band

If

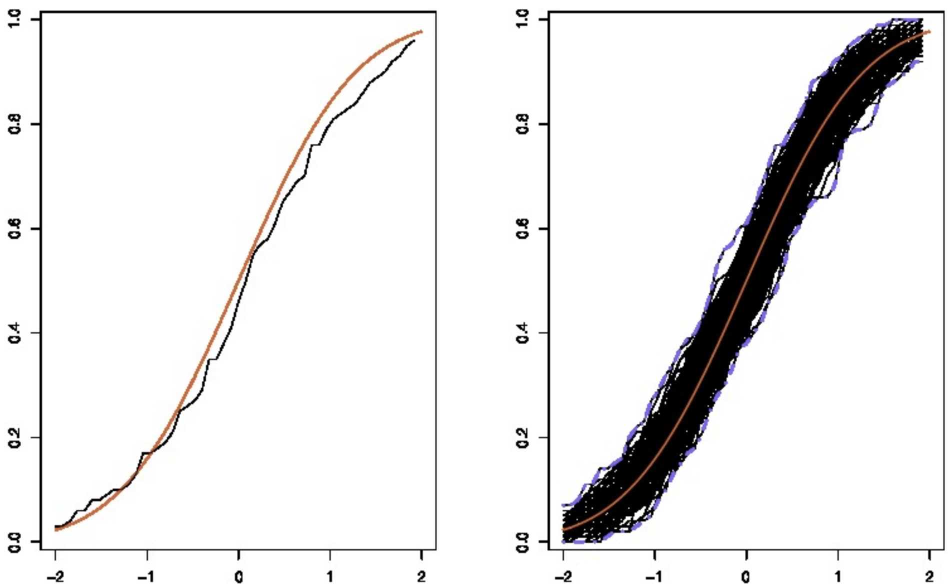
$$L_n(x) = \max\left\{\hat{F}_n(x) - \epsilon_n, 0\right\}, U_n(x) = \min\left\{\hat{F}_n(x) + \epsilon_n, 1\right\},$$

then, for $\varepsilon_n = \sqrt{1/2n \ln 2/\alpha}$,

$$\mathbb{P}\big(L_n(x) \leqslant F(x) \leqslant U_n(x) \text{ for all } x\big) \geqslant 1 - \alpha$$

Glivenko-Cantelli in action

Example (**Normal sample**)



Estimation of the cdf F from a normal sample of 100 points and variation of this estimation over 200 normal samples

Properties

- Estimator of a *non-parametric* nature : it is not necessary to know the distribution or the shape of the distribution of the sample to derive this estimator

  © it is always available

- **Robustess versus efficiency:** If the [parameterised] shape of the distribution is known, there exists a better approximation based on this shape, but if the shape is wrong, the parametric result can be completely off!

Properties

- Estimator of a *non-parametric* nature : it is not necessary to know the distribution or the shape of the distribution of the sample to derive this estimator

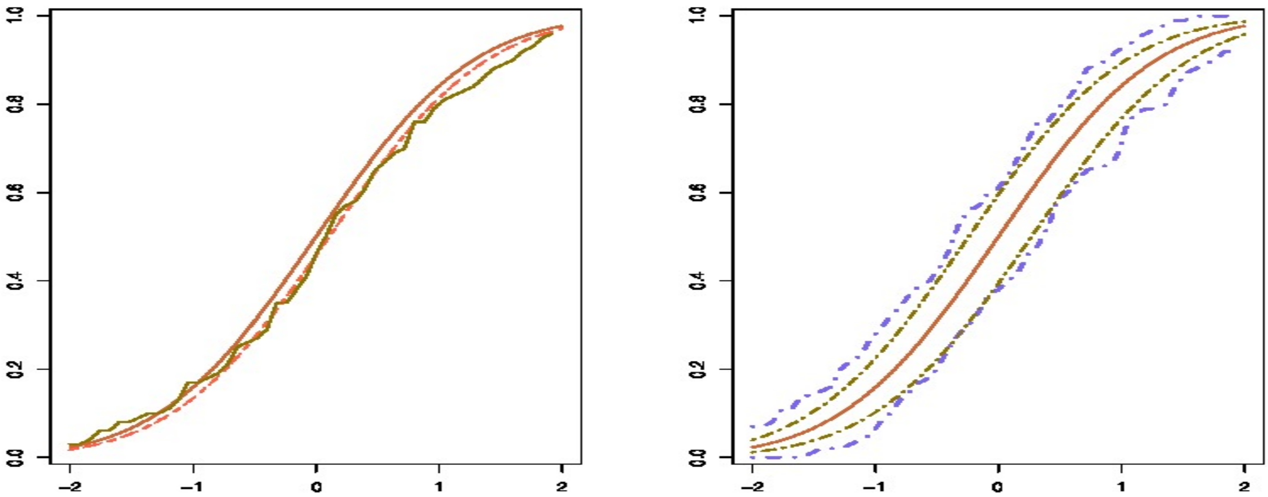
  ⓒ **it is always available**

- **Robustess versus efficiency:** If the [parameterised] shape of the distribution is known, there exists a better approximation based on this shape, but if the shape is wrong, the parametric result can be completely off!

parametric versus non-parametric inference

Example (**Normal sample**)

cdf of $\mathcal{N}(\theta, 1)$, $\Phi(x - \theta)$



Estimation of $\Phi(\cdot - \theta)$ by $\hat{F}_n$ and by $\Phi(\cdot - \hat{\theta})$ based on 100 points and maximal variation of those estimations over 200 replications

parametric versus non-parametric inference

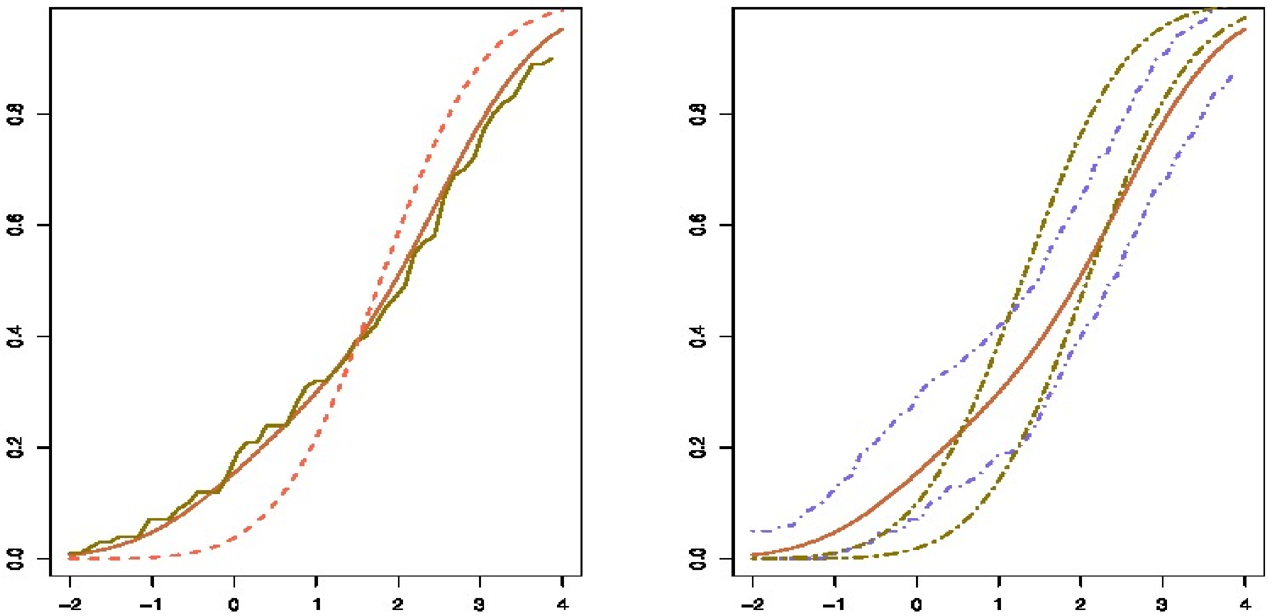
Example (**Non-normal sample)**

Sample issued from

$$0.3\mathcal{N}(0,1) + 0.7\mathcal{N}(2.5,1)$$

wrongly allocated to a normal distribution $\Phi(\cdot - \theta)$

parametric versus non-parametric inference



Estimation of F by $\widehat{F}_n$ and by $\Phi(\cdot - \widehat{\theta})$ based on 100 points and maximal variation of those estimations over 200 replications

Extension to functionals of F

For any quantity $\theta(F)$ depending on F, for instance,

$$\theta(F) = \int h(x)\, dF(x)\,,$$

[Functional of the cdf]

use of the plug-in approximation $\theta(\hat{F}_n)$, for instance,

$$\begin{aligned}\widehat{\theta(F)} &= \int h(x)\, d\hat{F}_n(x) \\ &= {}^1\!/_n \sum_{i=1}^{n} h(X_i)\end{aligned}$$

[Moment estimator]

Extension to functionals of F

For any quantity $\theta(F)$ depending on F, for instance,

$$\theta(F) = \int h(x)\,dF(x)\,,$$

[Functional of the cdf]

use of the plug-in approximation $\theta(\hat{F}_n)$, for instance,

$$\begin{aligned}\widehat{\theta(F)} &= \int h(x)\,d\hat{F}_n(x) \\ &= {}^1\!/_n \sum_{i=1}^{n} h(X_i)\end{aligned}$$

[Moment estimator]

examples

variance estimator

If

$$\theta(F) = \mathsf{var}(X) = \int \left(x - \mathbb{E}_F[X]\right)^2 dF(x)$$

then

$$\begin{aligned}\theta(\widehat{F}_n) &= \int \left(x - \mathbb{E}_{\widehat{F}_n}[X]\right)^2 d\widehat{F}_n(x) \\ &= {}^1\!/_n \sum_{i=1}^{n} \left(X_i - \mathbb{E}_{\widehat{F}_n}[X]\right)^2 = {}^1\!/_n \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2\end{aligned}$$

which differs from the (unbiased) sample variance

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

examples

median estimator

If $\theta(F)$ is the median of F, it is defined by

$$\mathbb{P}_F(X \leqslant \theta(F)) = 0.5$$

$\theta(\widehat{F}_n)$ is thus defined by

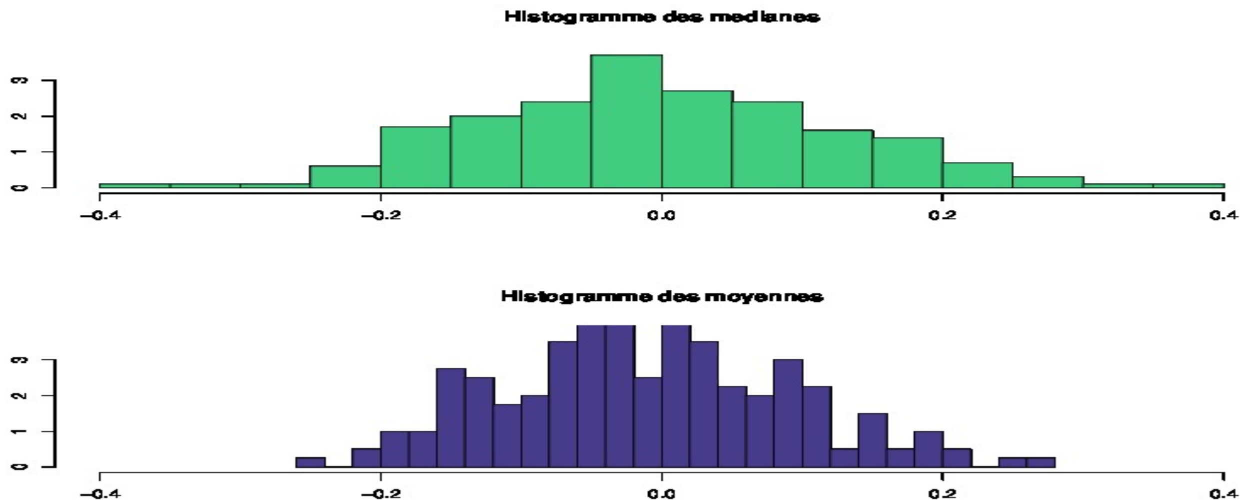
$$\mathbb{P}_{\widehat{F}_n}(X \leqslant \theta(\widehat{F}_n)) = {}^1\!/_n \sum_{i=1}^{n} \mathbb{I}(X_i \leqslant \theta(\widehat{F}_n)) = 0.5$$

which implies that $\theta(\widehat{F}_n)$ is the median of $X_1, \ldots, X_n$, namely $X_{(n/2)}$

median estimator

Example (**Normal sample)**

θ also is the median of $\mathcal{N}(\theta, 1)$, hence another estimator of θ is the median of $\hat{F}_n$, i.e. the median of $X_1, \ldots, X_n$, namely $X_{(n/2)}$



Comparison of the variations of sample means and sample medians over 200 normal samples

q-q plots

Graphical test of adequation for dataset $x_1, \ldots, x_n$ and targeted dsitribution F:
Plot sorted $x_1, \ldots, x_n$ against $F^{-1}(1/n+1), \ldots, F^{-1}(n/n+1)$

Example

Normal $\mathcal{N}(0,1)$ sample
against

- $\mathcal{N}(0,1)$
- $\mathcal{N}(0,2)$
- $\mathcal{E}(3)$

theoretical distributions

q-q plots

Graphical test of adequation for dataset $x_1, \ldots, x_n$ and targeted dsitribution F:

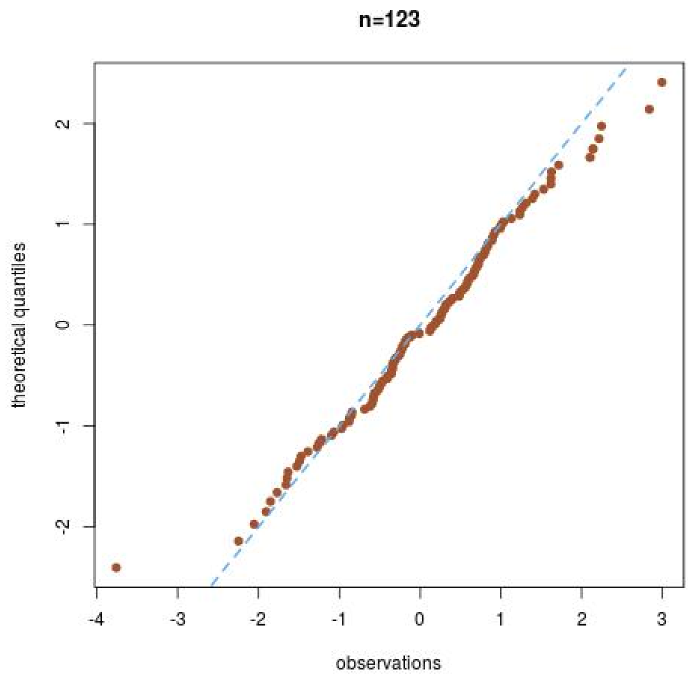
Plot sorted $x_1, \ldots, x_n$ against $F^{-1}(1/n+1), \ldots, F^{-1}(n/n+1)$

Example

Normal $\mathcal{N}(0, 1)$ sample against

- $\mathcal{N}(0, 1)$
- $\mathcal{N}(0, 2)$
- $\mathcal{E}(3)$

theoretical distributions



q-q plots

Graphical test of adequation for dataset $x_1, \ldots, x_n$ and targeted dsitribution F:

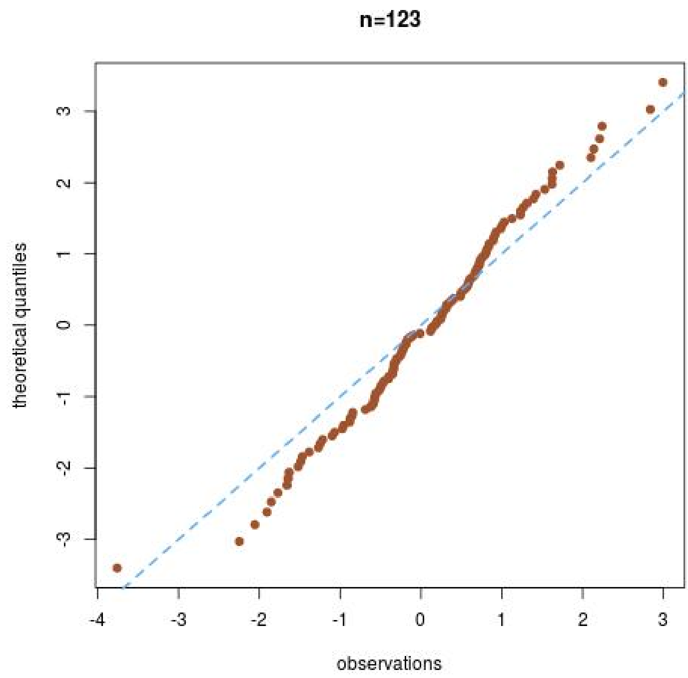
Plot sorted $x_1, \ldots, x_n$ against $F^{-1}(1/n+1), \ldots, F^{-1}(n/n+1)$

Example

Normal $\mathcal{N}(0,1)$ sample against

- $\mathcal{N}(0,1)$
- $\mathcal{N}(0,2)$
- $\mathcal{E}(3)$

theoretical distributions



q-q plots

Graphical test of adequation for dataset $x_1, \ldots, x_n$ and targeted dsitribution F:

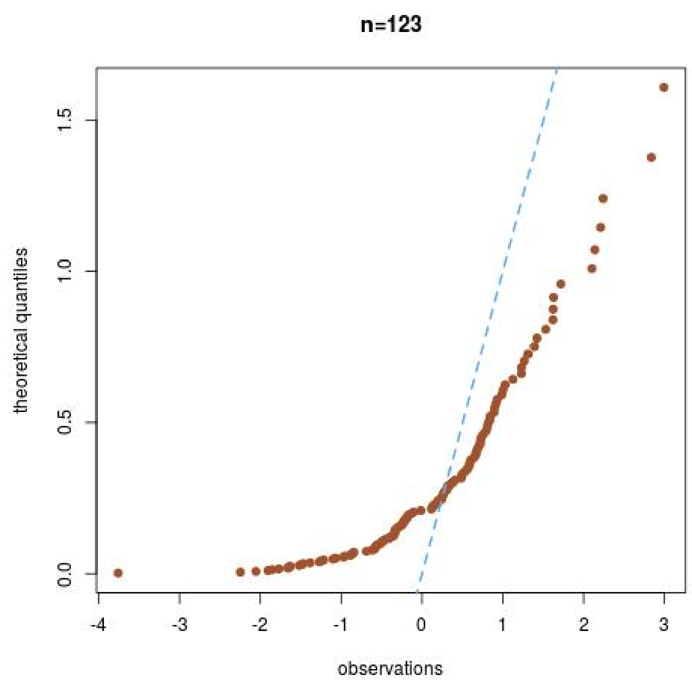
Plot sorted $x_1, \ldots, x_n$ against $F^{-1}(1/n+1), \ldots, F^{-1}(n/n+1)$

Example

Normal $\mathcal{N}(0, 1)$ sample against

- $\mathcal{N}(0, 1)$
- $\mathcal{N}(0, 2)$
- $\mathcal{E}(3)$

theoretical distributions



basis of Monte Carlo simulation

Recall the

Law of large numbers

If $X_1, \ldots, X_n$ simulated from f,

$$\widehat{\mathbb{E}[h(X)]}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[h(X)]$$

Result fundamental for the use of computer-based simulation

basis of Monte Carlo simulation

Recall the

Law of large numbers

If $X_1, \ldots, X_n$ simulated from f,

$$\widehat{\mathbb{E}[h(X)]}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[h(X)]$$

Result fundamental for the use of computer-based simulation

computer simulation

Principle

- produce by a computer program an arbitrary long sequence

$$x_1, x_2, \ldots \overset{\text{iid}}{\sim} F$$

- exploit the sequence as if it were a truly iid sample

© Mix of algorithmic, statistics, and probability theory

computer simulation

Principle

- produce by a computer program an arbitrary long sequence

$$x_1, x_2, \ldots \overset{\text{iid}}{\sim} F$$

- exploit the sequence *as if* it were a truly iid sample

ⓒ Mix of algorithmic, statistics, and probability theory

Monte Carlo simulation in practice

- For a given distribution F, call the corresponding pseudo-random generator in an arbitrary computer language

```
> x=rnorm(10)
> x
 [1] -0.02157345 -1.13473554  1.35981245 -0.88757941  0
 [7] -0.74941846  0.50629858  0.83579100  0.47214477
```

- use the sample as a statistician would do

```
> mean(x)
[1] 0.004892123
> var(x)
[1] 0.8034657
```

to approximate quantities related with F

Monte Carlo integration

Approximation of integrals related with F:

Law of large numbers

If $X_1, \dots, X_n$ simulated from f,

$$\hat{\mathfrak{I}}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{\text{a.s.}} \mathfrak{I} = \int h(x)\,dF(x)$$

Convergence a.s. as $n \to \infty$

Monte Carlo principle

1. Call a computer pseudo-random generator of F to produce $x_1, \dots, x_n$
2. Approximate $\mathfrak{I}$ with $\hat{\mathfrak{I}}_n$
3. Check the precision of $\hat{\mathfrak{I}}_n$ and if needed increase n

Monte Carlo integration

Approximation of integrals related with F:

Law of large numbers

If $X_1, \ldots, X_n$ simulated from f,

$$\hat{\mathfrak{I}}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i) \xrightarrow{\text{a.s.}} \mathfrak{I} = \int h(x)\, dF(x)$$

Convergence a.s. as $n \to \infty$

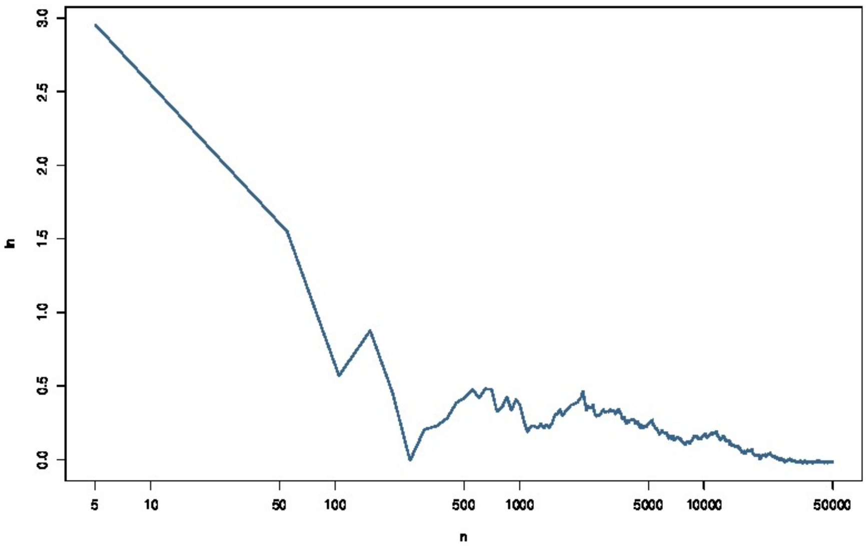
Monte Carlo principle

1. Call a computer pseudo-random generator of F to produce $x_1, \ldots, x_n$
2. Approximate $\mathfrak{I}$ with $\hat{\mathfrak{I}}_n$
3. Check the precision of $\hat{\mathfrak{I}}_n$ and if needed increase n

example: normal moment

For a Gaussian distribution, $\mathbb{E}[X^4] = 3$. Via Monte Carlo integration,

n	5	50	500	5000	50,000	500,000
$\widehat{\mathfrak{I}}_n$	1.65	5.69	3.24	3.13	3.038	3.029



How can one approximate the distribution of $\theta(\widehat{F}_n)$?

Given an estimate $\theta(\widehat{F}_n)$ of $\theta(F)$, its variability is required to evaluate precision

bootstrap principle

Since

$$\theta(\widehat{F}_n) = \theta(X_1, \ldots, X_n) \quad \textbf{with} \quad X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$$

replace F **with** $\widehat{F}_n$ **:**

$$\theta(\widehat{F}_n) \approx \theta(X_1^*, \ldots, X_n^*) \quad \textbf{with} \quad X_1^*, \ldots, X_n^* \overset{\text{iid}}{\sim} \widehat{F}_n$$

How can one approximate the distribution of $\theta(\widehat{F}_n)$?

Given an estimate $\theta(\widehat{F}_n)$ of $\theta(F)$, its variability is required to evaluate precision

bootstrap principle

Since

$$\theta(\widehat{F}_n) = \theta(X_1, \ldots, X_n) \quad \textbf{with} \quad X_1, \ldots, X_n \stackrel{\textbf{iid}}{\sim} F$$

replace F **with** $\widehat{F}_n$ **:**

$$\theta(\widehat{F}_n) \approx \theta(X_1^*, \ldots, X_n^*) \quad \textbf{with} \quad X_1^*, \ldots, X_n^* \stackrel{\textbf{iid}}{\sim} \widehat{F}_n$$

How can one approximate the distribution of $\theta(\hat{F}_n)$?

Given an estimate $\theta(\hat{F}_n)$ of $\theta(F)$, its variability is required to evaluate precision

bootstrap principle

Since

$$\theta(\hat{F}_n) = \theta(X_1, \ldots, X_n) \quad \textbf{with} \quad X_1, \ldots, X_n \overset{\textbf{iid}}{\sim} F$$

replace F **with** $\hat{F}_n$ **:**

$$\theta(\hat{F}_n) \approx \theta(X_1^*, \ldots, X_n^*) \quad \textbf{with} \quad X_1^*, \ldots, X_n^* \overset{\textbf{iid}}{\sim} \hat{F}_n$$

bootstrap principle

Since

$$\theta(\hat{F}_n) = \theta(X_1, \ldots, X_n) \quad \textbf{with} \quad X_1, \ldots, X_n \overset{\textbf{iid}}{\sim} F$$

replace F **with** $\hat{F}_n$ **:**

$$\theta(\hat{F}_n) \approx \theta(X_1^*, \ldots, X_n^*) \quad \textbf{with} \quad X_1^*, \ldots, X_n^* \overset{\textbf{iid}}{\sim} \hat{F}_n$$

illustration: bootstrap variance

For a given estimator $\theta(\widehat{F}_n)$, a random variable, its (true) variance is defined as

$$\sigma^2 = \mathbb{E}_F\left[(\theta(\widehat{F}_n) - \mathbb{E}_F[\theta(\widehat{F}_n)])^2\right]$$

bootstrap approximation

$$\mathbb{E}_{\widehat{F}_n}\left[(\theta(\widehat{\widehat{F}_n}) - \mathbb{E}_{\widehat{F}_n}[\theta(\widehat{F}_n)])^2\right] = \mathbb{E}_{\widehat{F}_n}\left[\theta(\widehat{\widehat{F}_n})^2\right] - \theta(\widehat{F}_n)^2$$

meaning that the random variable $\theta(\widehat{\widehat{F}_n})$ in the first expectation is now a transform of

$$X_1^*, \ldots, X_n^* \overset{\text{iid}}{\sim} \widehat{F}_n$$

while the second $\theta(\widehat{F}_n)$ is the original estimate

screen snapshot

bootstrap

/ˈbuːtstrap/

noun

noun: **bootstrap**; plural noun: **bootstraps**

1. a loop at the back of a boot, used to pull it on.
2. COMPUTING
   a technique of loading a program into a computer by means of a few initial instructions which enable the introduction of the rest of the program from an input device.
3. the technique of starting with existing resources to create something more complex and effective.
   "we see the creative act as a bootstrap process"

verb

verb: **bootstrap**; 3rd person present: **bootstraps**; gerund or present participle: **bootstrapping**; past tense: **bootstrapped**; past participle: **bootstrapped**

1. COMPUTING
   fuller form of boot[1] (sense 2 of the verb).
2. start up (an Internet-based business or other enterprise) with minimal financial resources.
   - get (oneself or something) into or out of a situation using existing resources.
     "the company is bootstrapping itself out of a marred financial past"

Remarks

- `bootstrap` because the sample itself is used to build an evaluation of its own distribution
- a bootstrap sample is obtained by n samplings with replacement in $(X_1, \ldots, X_n)$
- that is, X_1^* sampled from $(X_1, \ldots, X_n)$, then X_2^* independently sampled from $(X_1, \ldots, X_n)$, ...
- a bootstrap sample can thus take n^n values (or $\binom{2n-1}{n}$ values if the order does not matter)
- combinatorial complexity prevents analytic derivations

Remarks

- `bootstrap` because the sample itself is used to build an evaluation of its own distribution
- a bootstrap sample is obtained by n samplings with replacement in $(X_1, \ldots, X_n)$
- that is, X_1^* sampled from $(X_1, \ldots, X_n)$, then X_2^* independently sampled from $(X_1, \ldots, X_n)$, ...
- a bootstrap sample can thus take n^n values (or $\binom{2n-1}{n}$ values if the order does not matter)
- combinatorial complexity prevents analytic derivations

bootstrap by simulation

Implementation

Since $\widehat{F}_n$ is known, it is possible to **simulate** from $\widehat{F}_n$, therefore one can approximate the distribution of $\theta(X_1^*, \ldots, X_n^*)$ [instead of $\theta(X_1, \ldots, X_n)$]

The distribution corresponding to

$$\widehat{F}_n(x) = \text{card}\{X_i; X_i \leqslant x\}/n$$

allocates a probability of $1/n$ to each point in $\{x_1, \ldots, x_n\}$:

$$\Pr^{\widehat{F}_n}(X^* = x_i) = 1/n$$

Simulating from $\widehat{F}_n$ is equivalent to sampling **with replacement** in $(X_1, \ldots, X_n)$

[in R, sample(x,n,replace=TRUE)]

bootstrap algorithm

Monte Carlo implementation

1. For $b = 1, \ldots, B$,
   1. generate a sample $X_1^b, \ldots, X_n^b$ from $\hat{F}_n$
   2. construct the corresponding value

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

2. Use the sample

$$\hat{\theta}^1, \ldots, \hat{\theta}^B$$

   to approximate the distribution of

$$\theta(X_1, \ldots, X_n)$$

bootstrap algorithm

Monte Carlo implementation

1. For $b = 1, \ldots, B$,
   1. generate a sample $X_1^b, \ldots, X_n^b$ from $\hat{F}_n$
   2. construct the corresponding value

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

2. Use the sample

$$\hat{\theta}^1, \ldots, \hat{\theta}^B$$

   to approximate the distribution of

$$\theta(X_1, \ldots, X_n)$$

bootstrap algorithm

Monte Carlo implementation

1. For $b = 1, \dots, B$,
   1. generate a sample $X_1^b, \dots, X_n^b$ from $\hat{F}_n$
   2. construct the corresponding value

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

2. Use the sample

   $$\hat{\theta}^1, \dots, \hat{\theta}^B$$

   to approximate the distribution of

   $$\theta(X_1, \dots, X_n)$$

mixture illustration

- Observation of a sample [here simulated from $0.3\mathcal{N}(0,1) + 0.7\mathcal{N}(2.5,1)$ as illustration]

```
> x=rnorm(250)+(runif(250)<.7)*2.5 #n=250
```

- Interest in the distribution of $\bar{X} = 1/n \sum_i X_i$

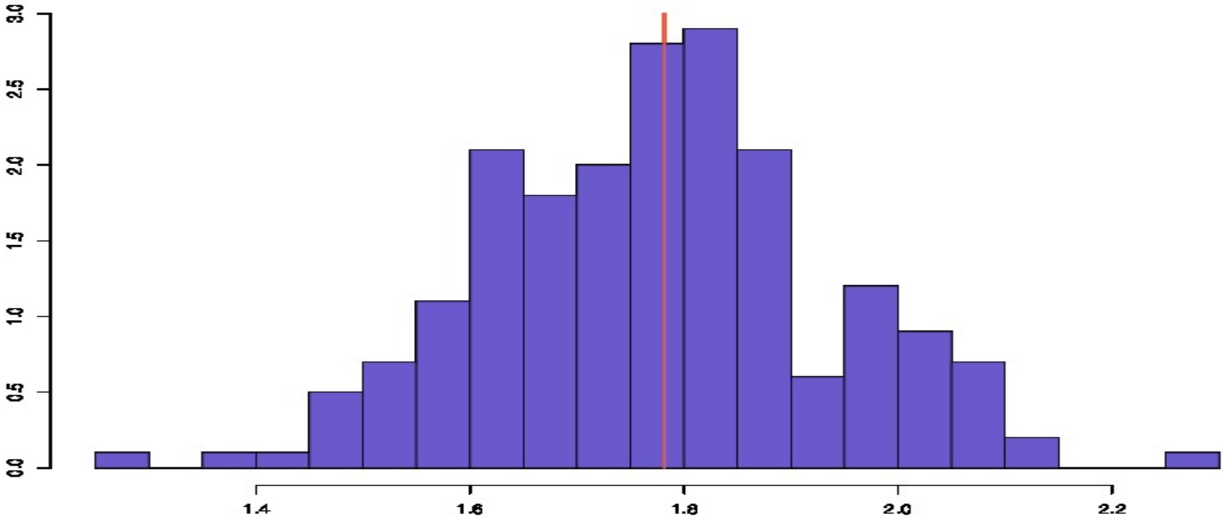
```
> xbar=mean(x)
[1] 1.73696
```

- Bootstrap sample of $\bar{X}^*$

```
> bobar=rep(0,1000) #B=1000
> for (t in 1:1000)
+ bobar[t]=mean(sample(x,250,rep=TRUE))
> hist(bobar)
```

mixture illustration

Example (Sample $0.3\mathcal{N}(0,1) + 0.7\mathcal{N}(2.5,1)$)



Variation of the empirical means over 200 bootstrap samples versus observed average

mixture illustration

Example (Derivation of the average variation)

For an estimator $\theta(X_1, \ldots, X_n)$, the standard deviation is given by

$$\eta(F) = \sqrt{\mathsf{E}^F\left[\{\theta(X_1, \ldots, X_n) - \mathsf{E}^F[\theta(X_1, \ldots, X_n)]\}^2\right]}$$

and its bootstrap approximation is

$$\eta(\hat{F}_n) = \sqrt{\mathsf{E}^{\hat{F}_n}\left[\{\theta(X_1, \ldots, X_n) - \mathsf{E}^{\hat{F}_n}[\theta(X_1, \ldots, X_n)]\}^2\right]}$$

mixture illustration

Example (Derivation of the average variation)

For an estimator $\theta(X_1, \ldots, X_n)$, the standard deviation is given by

$$\eta(F) = \sqrt{\mathbb{E}^F \left[\{\theta(X_1, \ldots, X_n) - \mathbb{E}^F[\theta(X_1, \ldots, X_n)]\}^2\right]}$$

and its bootstrap approximation is

$$\eta(\hat{F}_n) = \sqrt{\mathbb{E}^{\hat{F}_n} \left[\{\theta(X_1, \ldots, X_n) - \mathbb{E}^{\hat{F}_n}[\theta(X_1, \ldots, X_n)]\}^2\right]}$$

mixture illustration

Example (Derivation of the average variation)

Approximation itself approximated by Monte-Carlo:

$$\hat{\eta}(\hat{F}_n) = \left({}^1/_B \sum_{b=1}^{B} (\theta(X_1^b, \ldots, X_n^b) - \bar{\theta})^2 \right)^{1/2}$$

where

$$\bar{\theta} = {}^1/_B \sum_{b=1}^{B} \theta(X_1^b, \ldots, X_n^b)$$

bootstrap confidence intervals

Several ways to implement the bootstrap principle to get confidence intervals, that is intervals $C(X_1, \dots, X_n)$ on $\theta(F)$ such that

$$\mathbb{P}\big(C(X_1, \dots, X_n) \ni \theta(F)\big) = 1 - \alpha$$

[$1 - \alpha$-level confidence intervals]

1 rely on the normal approximation

$$\theta(\widehat{F}_n) \approx N(\theta(F), \eta(F)^2)$$

and use the interval

$$\big[\theta(\widehat{F}_n) + z_{\alpha/2}\eta(\widehat{F}_n), \theta(\widehat{F}_n) - z_{\alpha/2}\eta(\widehat{F}_n)\big]$$

bootstrap confidence intervals

Several ways to implement the bootstrap principle to get confidence intervals, that is intervals $C(X_1, \dots, X_n)$ on $\theta(F)$ such that

$$\mathbb{P}\big(C(X_1, \dots, X_n) \ni \theta(F)\big) = 1 - \alpha$$

[$1 - \alpha$-level confidence intervals]

2 generate a bootstrap approximation to the cdf of $\theta(\widehat{F}_n)$

$$\widehat{H}(r) = {}^1\!/_B \sum_{b=1}^{B} \mathbb{I}(\theta(X_1^b, \dots, X_n^b) \leqslant r)$$

and use the interval

$$\big[\widehat{H}^{-1}(\alpha/2), \widehat{H}^{-1}(1 - \alpha/2)\big]$$

which is also

$$\big[\theta^*_{(n\{\alpha/2\})}, \theta^*_{(n\{1-\alpha/2\})}\big]$$

bootstrap confidence intervals

Several ways to implement the bootstrap principle to get confidence intervals, that is intervals $C(X_1, \ldots, X_n)$ on $\theta(F)$ such that

$$\mathbb{P}\big(C(X_1, \ldots, X_n) \ni \theta(F)\big) = 1 - \alpha$$

[$1 - \alpha$-level confidence intervals]

3 generate a bootstrap approximation to the cdf of $\theta(\widehat{F}_n) - \theta(F)$,

$$\widehat{H}(r) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}((\theta(X_1^b, \ldots, X_n^b) - \theta(\widehat{F}_n) \leqslant r)$$

and use the interval

$$\big[\theta(\widehat{F}_n) - \widehat{H}^{-1}(1 - \alpha/2), \theta(\widehat{F}_n) - \widehat{H}^{-1}(\alpha/2)\big]$$

which is also

$$\big[2\theta(\widehat{F}_n) - \theta^*_{(n\{1-\alpha/2\})}, 2\theta(\widehat{F}_n) - \theta^*_{(n\{\alpha/2\})}\big]$$

exemple: median confidence intervals

Take $X_1, \ldots, X_n$ an iid random sample and $\theta(F)$ as the median of F, then

$$\theta(F_n) = X_{(n/2)}$$

```
> x=rnorm(123)
> median(x)
[1] 0.03542237
> T=10^3
> bootmed=rep(0,T)
> for (t in 1:T) bootmed[t]=median(sample(x,123,rep=TRUE))
> sd(bootmed)
[1] 0.1222386
> median(x)-2*sd(bootmed)
[1] -0.2090547
> median(x)+2*sd(bootmed)
[1] 0.2798995
```

exemple: median confidence intervals

Take $X_1, \ldots, X_n$ an iid random sample and $\theta(F)$ as the median of F, then

$$\theta(F_n) = X_{(n/2)}$$

```
> x=rnorm(123)
> median(x)
[1] 0.03542237
> T=10^3
> bootmed=rep(0,T)
> for (t in 1:T) bootmed[t]=median(sample(x,123,rep=TRUE))
> quantile(bootmed,prob=c(.025,.975))
      2.5%       97.5%
-0.2430018  0.2375104
```

exemple: median confidence intervals

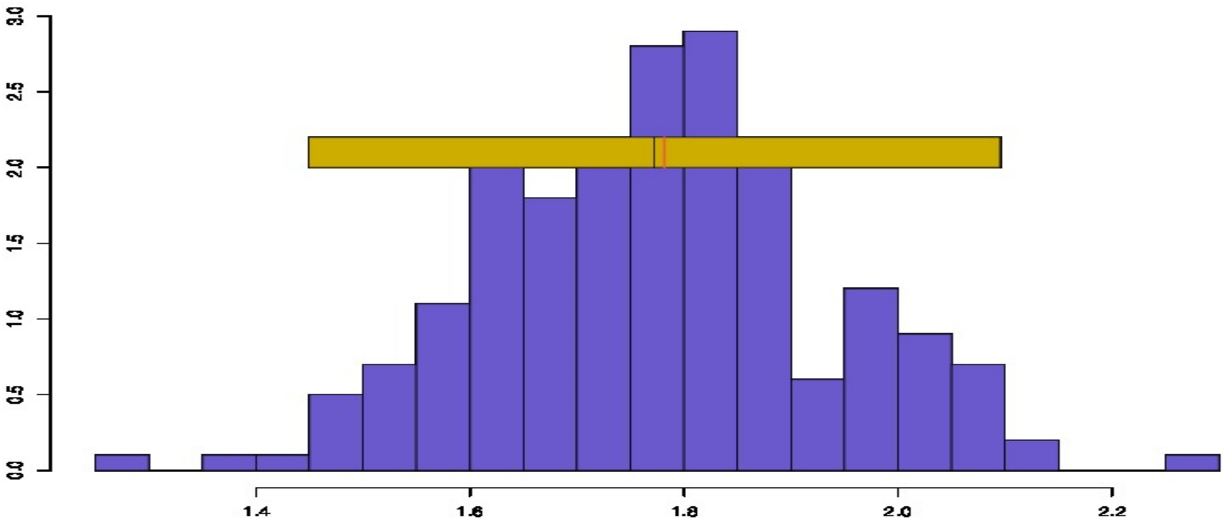
Take $X_1, \ldots, X_n$ an iid random sample and $\theta(F)$ as the median of F, then

$$\theta(F_n) = X_{(n/2)}$$

```
> x=rnorm(123)
> median(x)
[1] 0.03542237
> T=10^3
> bootmed=rep(0,T)
> for (t in 1:T) bootmed[t]=median(sample(x,123,rep=TRUE))
> 2*median(x)-quantile(bootmed,prob=c(.975,.025))
     97.5%       2.5%
-0.1666657  0.3138465
```

example: mean bootstrap variation

Example (**Sample** $0.3\mathcal{N}(0,1) + 0.7\mathcal{N}(2.5,1)$)



Interval of bootstrap variation at $\pm 2\hat{\eta}(\hat{F}_n)$ **and average of the observed sample**

example: mean bootstrap variation

Example (Normal sample)

Sample

$$(X_1, \ldots, X_{100}) \overset{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$$

Comparison of the confidence intervals

$$[\bar{x} - 2 * \hat{\sigma}_x/10, \bar{x} + 2 * \hat{\sigma}_x/10] = [-0.113, 0.327]$$

[normal approximation]

$$[\bar{x}^* - 2 * \hat{\sigma}^*, \bar{x}^* + 2 * \hat{\sigma}^*] = [-0.116, 0.336]$$

[normal bootstrap approximation]

$$[q^*(0.025), q^*(0.975)] = [-0.112, 0.336]$$

[generic bootstrap approximation]

example: mean bootstrap variation

Example (**Normal sample)**

Sample

$$(X_1, \ldots, X_{100}) \overset{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$$

Comparison of the confidence intervals

$$[\bar{x} - 2 * \hat{\sigma}_x/10, \bar{x} + 2 * \hat{\sigma}_x/10] = [-0.113, 0.327]$$

[normal approximation]

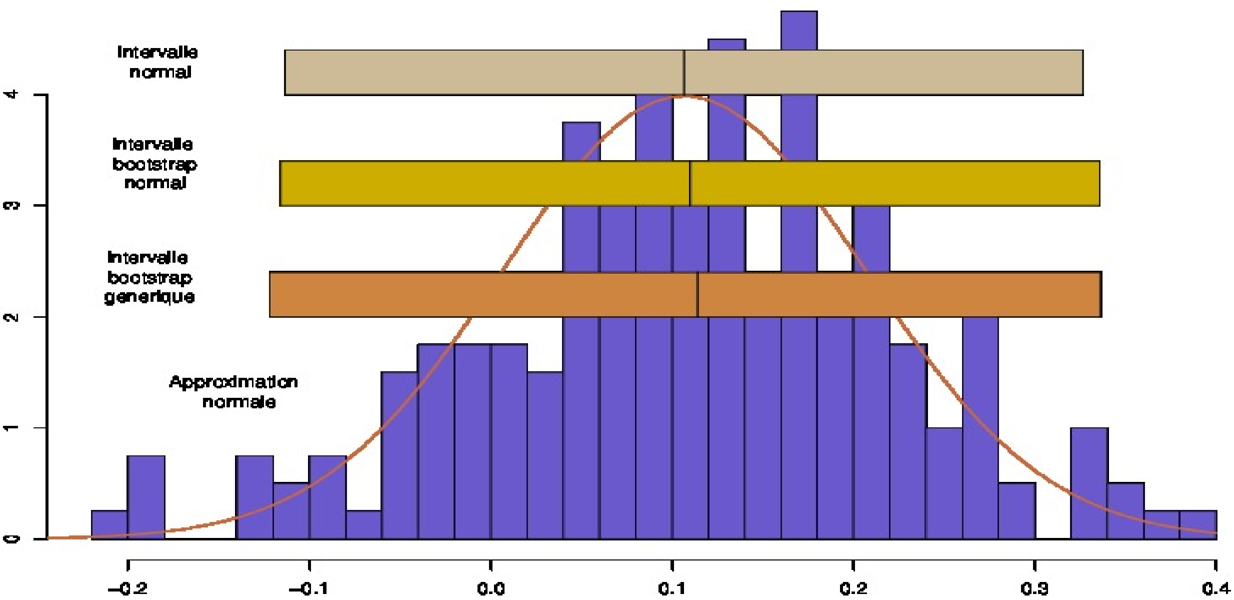
$$[\bar{x}^* - 2 * \hat{\sigma}^*, \bar{x}^* + 2 * \hat{\sigma}^*] = [-0.116, 0.336]$$

[normal bootstrap approximation]

$$[q^*(0.025), q^*(0.975)] = [-0.112, 0.336]$$

[generic bootstrap approximation]

example: mean bootstrap variation



Variation ranges at 95% for a sample of 100 points and 200 bootstrap replications

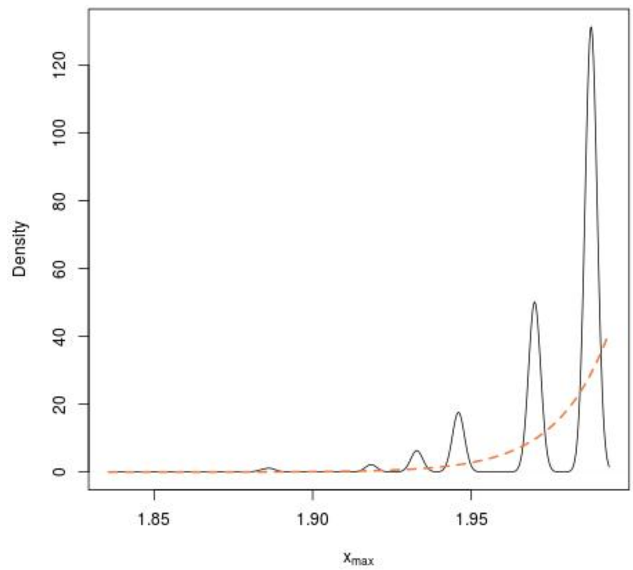
a counter-example

Consider $X_1, \dots, X_n \sim \mathcal{U}(0, \theta)$ then

$$\theta = \theta(F) = \mathbb{E}_\theta \left[\frac{n}{n-1} X_{(n)} \right]$$

Using bootstrap, distribution of $\,^{n-1}/_{n}\theta(\widehat{F}_n)$ far from truth

$$f_{\max}(x) = \,^{n\,x^{n-1}}/_{\theta^n}\, \mathbb{I}_{(0,\theta)}(x)$$



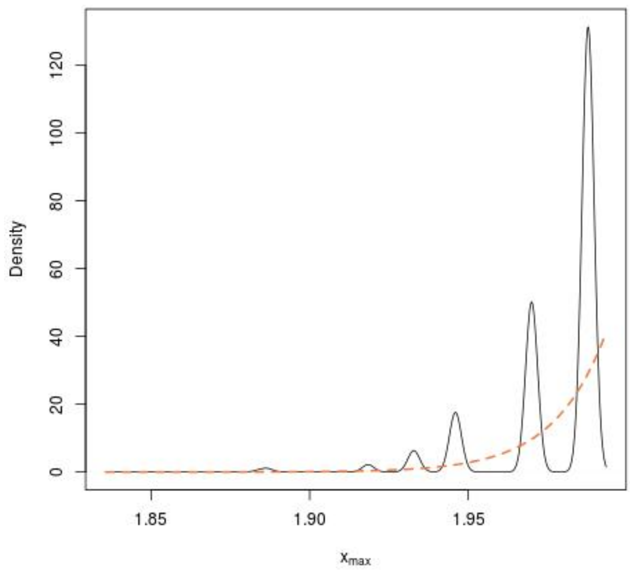
a counter-example

Consider $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$ then

$$\theta = \theta(F) = \mathbb{E}_\theta \left[\frac{n}{n-1} X_{(n)} \right]$$

Using bootstrap, distribution of ${}^{n-1}/_{n}\theta(\widehat{F}_n)$ far from truth

$$f_{\max}(x) = {}^{n\,x^{n-1}}/_{\theta^n}\, \mathbb{I}_{(0,\theta)}(x)$$



Parametric Bootstrap

If the parametric shape of F is known,

$$F(\cdot) = \Phi_\lambda(\cdot) \qquad \lambda \in \Lambda\,,$$

an evaluation of F more efficient than $\hat{F}_n$ is provided by

$$\Phi_{\hat{\lambda}_n}$$

where $\hat{\lambda}_n$ is a convergent estimator of λ

[Cf Example 3]

Parametric Bootstrap

If the parametric shape of F is known,

$$F(\cdot) = \Phi_\lambda(\cdot) \qquad \lambda \in \Lambda\,,$$

an evaluation of F more efficient than $\hat{F}_n$ is provided by

$$\Phi_{\hat{\lambda}_n}$$

where $\hat{\lambda}_n$ is a convergent estimator of λ

[Cf Example 3]

Parametric Bootstrap

Approximation of the distribution of

$$\theta(X_1, \ldots, X_n)$$

by the distribution of

$$\theta(X_1^*, \ldots, X_n^*) \qquad X_1^*, \ldots, X_n^* \overset{\text{iid}}{\sim} \Phi_{\hat{\lambda}_n}$$

May avoid Monte Carlo simulation approximations in some cases

Parametric Bootstrap

Approximation of the distribution of

$$\theta(X_1, \dots, X_n)$$

by the distribution of

$$\theta(X_1^*, \dots, X_n^*) \qquad X_1^*, \dots, X_n^* \overset{\text{iid}}{\sim} \Phi_{\hat{\lambda}_n}$$

May avoid Monte Carlo simulation approximations in some cases

example of parametric Bootstrap

Example (**Exponential Sample**)

Take

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}\text{xp}(\lambda)$$

and $\lambda = 1/\mathbb{E}_\lambda[X]$ to be estimated

A possible estimator is

$$\hat{\lambda}(x_1, \ldots, x_n) = \frac{n}{\sum_{i=1}^{n} x_i}$$

but this estimator is biased

$$\mathbb{E}_\lambda[\hat{\lambda}(X_1, \ldots, X_n)] \neq \lambda$$

example of parametric Bootstrap

Example (**Exponential Sample**)

Take

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}\text{xp}(\lambda)$$

and $\lambda = 1/E_\lambda[X]$ to be estimated

A possible estimator is

$$\hat{\lambda}(x_1, \ldots, x_n) = \frac{n}{\sum_{i=1}^n x_i}$$

but this estimator is biased

$$E_\lambda[\hat{\lambda}(X_1, \ldots, X_n)] \neq \lambda$$

example of parametric Bootstrap

Example (Exponential Sample (2))

Questions :

- What is the bias

$$\lambda - E_\lambda[\hat{\lambda}(X_1, \ldots, X_n)]$$

  of this estimator ?

- What is the distribution of this estimator ?

example of parametric Bootstrap

Example (**Exponential Sample (2)**)

Questions :

- What is the bias

$$\lambda - E_\lambda[\hat{\lambda}(X_1, \dots, X_n)]$$

  of this estimator ?
- What is the distribution of this estimator ?

Bootstrap evaluation of the bias

Example (**Exponential Sample (3)**)

$$\hat{\lambda}(x_1, \ldots, x_n) - \mathrm{E}_{\hat{\lambda}(x_1,\ldots,x_n)}[\hat{\lambda}(X_1, \ldots, X_n)]$$

[parametric version]

$$\hat{\lambda}(x_1, \ldots, x_n) - \mathrm{E}_{\hat{F}_n}[\hat{\lambda}(X_1, \ldots, X_n)]$$

[non-parametric version]

example: bootstrap bias evaluation

Example (Exponential Sample (4))

In the first (parametric) version,

$$1/\hat{\lambda}(X_1, \ldots, X_n) \sim \mathcal{G}a(n, n\lambda)$$

and

$$\mathbb{E}_\lambda[\hat{\lambda}(X_1, \ldots, X_n)] = \frac{n}{n-1}\lambda$$

therefore the bias is **analytically** evaluated as

$$-\lambda/n-1$$

and estimated by

$$-\frac{\hat{\lambda}(X_1, \ldots, X_n)}{n-1} = -0.00787$$

example: bootstrap bias evaluation

Example (**Exponential Sample (4)**)

In the first (parametric) version,

$$1/\hat{\lambda}(X_1, \ldots, X_n) \sim \mathcal{G}a(n, n\lambda)$$

and

$$\mathbb{E}_\lambda[\hat{\lambda}(X_1, \ldots, X_n)] = \frac{n}{n-1}\lambda$$

therefore the bias is **analytically** evaluated as

$$-\lambda/n-1$$

and estimated by

$$-\frac{\hat{\lambda}(X_1, \ldots, X_n)}{n-1} = -0.00787$$

example: bootstrap bias evaluation

Example (Exponential Sample (5))

In the second (nonparametric) version, evaluation by Monte Carlo,

$$\hat{\lambda}(x_1, \ldots, x_n) - \mathbb{E}_{\hat{F}_n}[\hat{\lambda}(X_1, \ldots, X_n)] = 0.00142$$

which achieves the **"wrong"** sign

example: bootstrap bias evaluation

Example (**Exponential Sample (6)**)

Construction of a confidence interval on λ

By parametric bootstrap,

$$\Pr_\lambda\left(\hat\lambda_1 \leqslant \lambda \leqslant \hat\lambda_2\right) = \Pr\left(\omega_1 \leqslant \lambda/\hat\lambda \leqslant \omega_2\right) = 0.95$$

can be deduced from

$$\lambda/\hat\lambda \sim \mathcal{G}a(n, n)$$

[In R, qgamma(0.975,n,1/n)]

$$[\hat\lambda_1, \hat\lambda_2] = [0.452, 0.580]$$

example: bootstrap bias evaluation

Example (Exponential Sample (7))

In nonaramettric bootstrap, one replaces

$$\Pr_F\left(q(.025) \leqslant \lambda(F) \leqslant q(.975)\right) = 0.95$$

with

$$\Pr_{\hat{F}_n}\big(q^*(.025) \leqslant \lambda(\hat{F}_n) \leqslant q^*(.975)\big) = 0.95$$

Approximation of quantiles $q^*(.025)$ and $q^*(.975)$ of $\lambda(\hat{F}_n)$ by bootstrap (Monte Carlo) sampling

$$[q^*(.025), q^*(.975)] = [0.454, 0.576]$$

example: bootstrap bias evaluation

Example (**Exponential Sample (7)**)

In nonarametric bootstrap, one replaces

$$\Pr_F\left(q(.025) \leqslant \lambda(F) \leqslant q(.975)\right) = 0.95$$

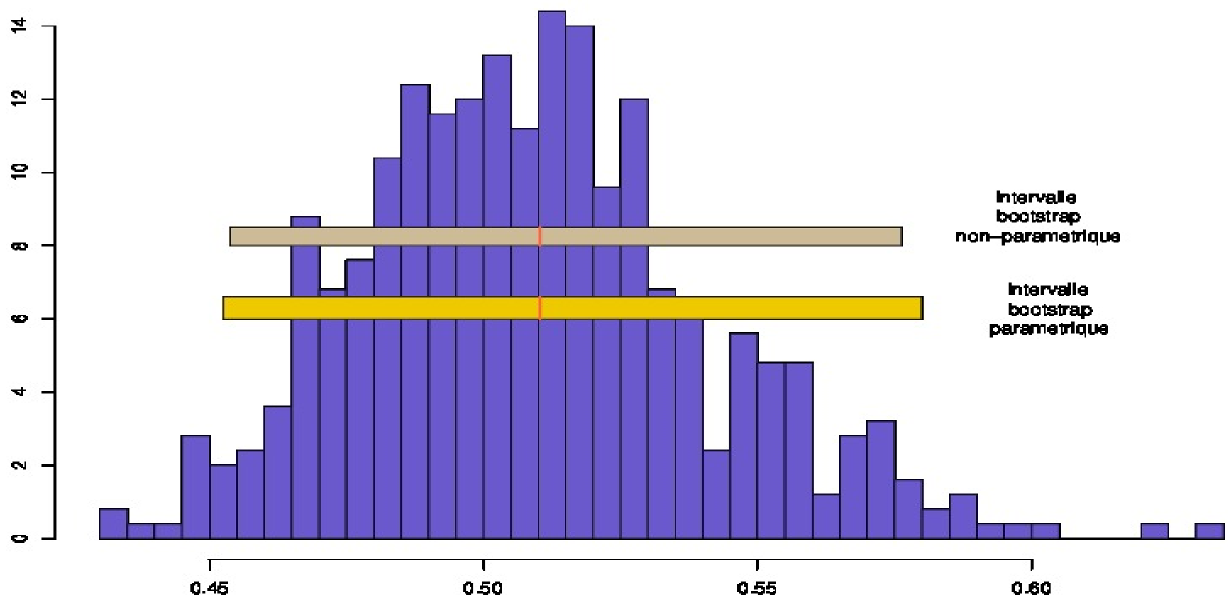
with

$$\Pr_{\hat{F}_n}\left(q^*(.025) \leqslant \lambda(\hat{F}_n) \leqslant q^*(.975)\right) = 0.95$$

Approximation of quantiles $q^*(.025)$ and $q^*(.975)$ of $\lambda(\hat{F}_n)$ by bootstrap (Monte Carlo) sampling

$$[q^*(.025), q^*(.975)] = [0.454, 0.576]$$

example: bootstrap bias evaluation



example: bootstrap distribution evaluation

Example (Student Sample)

Take

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathfrak{T}(5, \mu, \tau^2) \overset{\text{def}}{=} \mu + \tau \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_5/5}}$$

μ and τ could be estimated by

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\tau}_n = \sqrt{\frac{5-2}{5}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2} = \sqrt{\frac{5-2}{5}}\,\hat{\sigma}_n$$

example: bootstrap distribution evaluation

Example (Student Sample (2))

Problem $\hat{\mu}_n$ is not distributed from a Student $\mathfrak{T}(5, \mu, \tau^2/n)$ distribution

The distribution of $\hat{\mu}_n$ can be reproduced by bootstrap sampling

example: bootstrap distribution evaluation

Example (Student Sample (3))

Comparison of confidence intervals

$$[\hat{\mu}_n - 2 * \hat{\sigma}_n/10, \hat{\mu}_n + 2 * \hat{\sigma}_n/10] = [-0.068, 0.319]$$

[normal approximation]

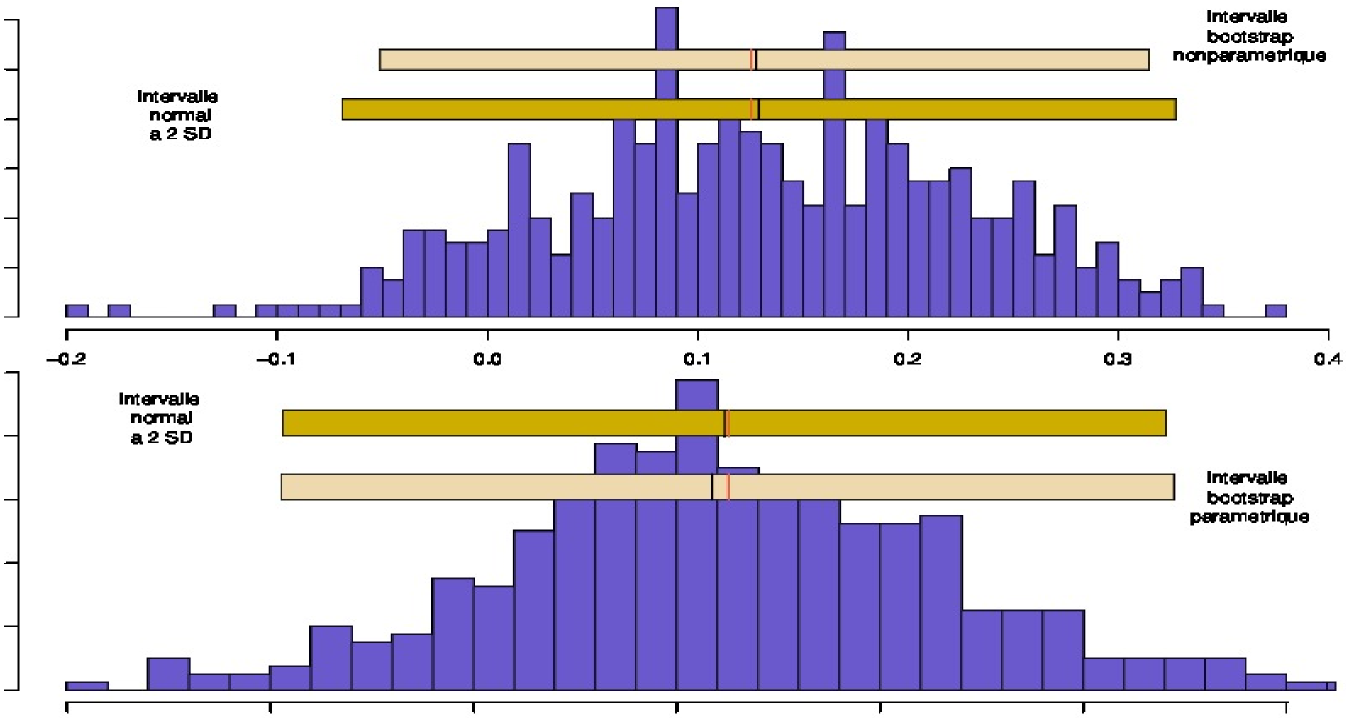
$$[q^*(0.05), q^*(0.95)] = [-0.056, 0.305]$$

[parametric boostrap approximation]

$$[q^*(0.05), q^*(0.95)] = [-0.094, 0.344]$$

[non parametric boostrap approximation]

example: bootstrap distribution evaluation



95% **variation interval for a** 150 **points sample with** 400 **bootstrap replicas** *(top)* **nonparametric and** *(bottom)* **parametric**