Technische Universität Braunschweig



Bootstrap

Jens–Peter Kreiß, TU Braunschweig

Based on joint work with

Efstathios Paparoditis, Cyprus Academy of Sciences, Letters, and Arts

September 8, 2026

Overview

Introduction

Bootstrap Procedure

Bootstrap Asymptotics

Bootstrap Consistency

Bootstrap Failure

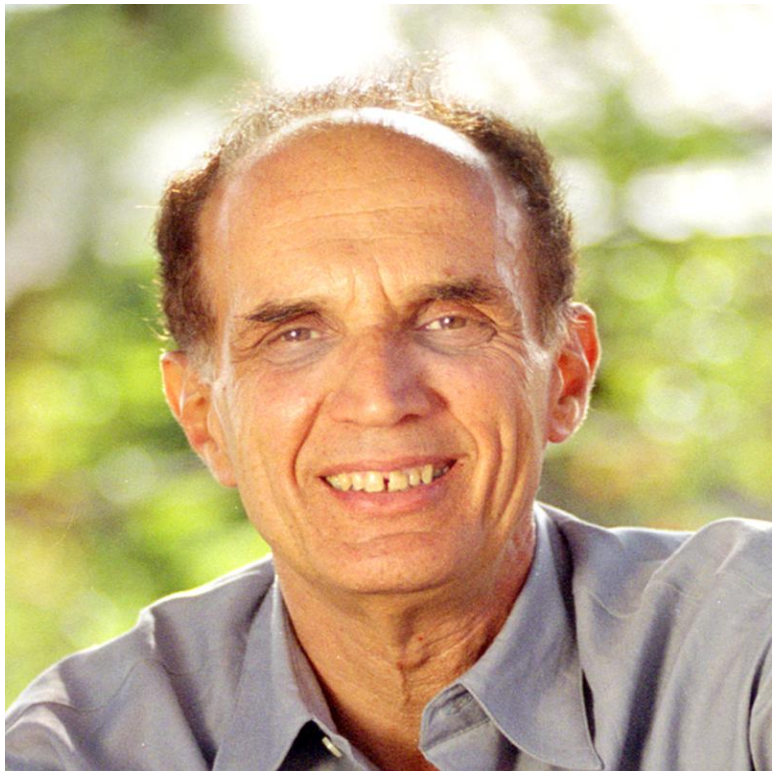
Bootstrap for Time Series

References

Subsampling

Second-Order Properties of Bootstrap

The Inventor of Bootstrap: Professor Bradley Efron (Stanford University)



Efron, B. (1979): Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7, 1–16.

30.148 Google scholar citations!

Together with his monograph "An Introduction to the Bootstrap" (joint with Robert J. Tibshirani) close to 100.000 citations.

Introduction

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample of real-valued random variables with unknown mean $\mathrm{E}(X_1) = \mu$ and variance $0 < \sigma_X^2 = \mathrm{Var}(X_1) < \infty$. Consider the problem of estimating the unknown mean μ by means of the sample mean $\overline{X}_n = n^{-1} \sum_{j=1}^{n} X_j$.

We are interested in estimating the distribution of $\overline{X}_n$. Why?
The CLT leads for $n \to \infty$ to

$$S_n = \sqrt{n}(\overline{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X_1)).$$

Assume for a moment that the cdf $F(x) = P\{X_1 \leq x\}$ is known.

In principle, $\mathcal{L}(S_n)$ can then be computed. But that could be difficult!
A computationally intensive workaround:

Simulate any number of samples of length n from F and use these samples to estimate the distribution of $\sqrt{n}(\overline{X}_n - \mu)$ empirically.

More specific, denote by $X_1^+, X_2^+, \ldots, X_n^+$ an i.i.d. random sample stemming from F and let $S_n^+ = \sqrt{n}(\overline{X}_n^+ - \mu)$.

Notice that $\sqrt{n}(\overline{X}_n^+ - \mu)$ is a copy of $\sqrt{n}(\overline{X}_n - \mu)$ since both $X_1, X_2, \ldots, X_n$ and $X_1^+, X_2^+, \ldots, X_n^+$ are samples of i.i.d. random variables having the same distribution F.
Now, since we can generate as many copies of S_n as we wish, say $S_{n,1}^+, \ldots, , S_{n,B}^+$, where $S_{n,b}^+ = \sqrt{n}(\overline{X}_{n,b}^+ - \mu)$ and $\overline{X}_{n,b}^+ = n^{-1}\sum_{j=1}^n X_{j,b}^+$, we can estimate the distribution function of interest, which is

$$G_n(x) = P(\sqrt{n}(\overline{X}_n - \mu) \leq x), \ x \in \mathbb{R},$$

by means of the empirical distribution function

$$\widehat{G}_{n,B}(x) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{(-\infty,x]}(S_{n,b}^+).$$

as accurately as our computer allows!

$\widehat{G}_{n,B}$ is calculated from the B independent copies $S^+_{n,1}, \ldots, S^+_{n,B}$ of $\sqrt{n}(\overline{X}_n - \mu)$. Notice that since the $S^+_{n,b}$ are independent and identically distributed, it follows by Glivenko-Cantelli's Theorem, that $\sup_{x\in\mathbb{R}} |\widehat{G}_{n,B}(x) - G_n(x)| \to 0$, almost surely, as $B \to \infty$.

Practical implementation of this idea faces of course the problem that the distribution function F is unknown!

The solution that the bootstrap proposes, is to replace F by an estimator of F based on the sample $X_1, X_2, \ldots, X_n$. One strong candidate is of course the empirical distribution function $\widehat{F}_n$,

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(-\infty,x]}(X_j), \ x \in \mathbb{R}, \tag{1}$$

because of its nice properties.
Parametric models $\{F_\theta : \theta \in \Theta\} \ni F$ would also be an option.

A random sample from $\widehat{F}_n$ is called a bootstrap sample and it will be denoted by $X_1^*, X_2^*, \ldots, X_n^*$.

The random variables X_j^* are defined on some underlying probability space $(\mathcal{X}^*, \mathcal{F}^*, P^*)$ and take values in $\{X_1, X_2, \ldots, X_n\}$, where each observation is listed in the set, regardless of whether observed values are identical or not. This ensures that the set $\{X_1, X_2, \ldots, X_n\}$ always contains exactly n elements.

Observe that $X_j^* \sim \widehat{F}_n$ means each X_j^* takes one of the values in the set $\{X_1, X_2, \ldots, X_n\}$ with probability $1/n$, in other words, each X_j^* is obtained by **drawing with replacement** from $\{X_1, X_2, \ldots, X_n\}$.

The expectation of any X_j^* for example is equal to

$$\sum_{j=1}^{n} X_j \cdot P^*(X_1^* = X_j) = \overline{X}_n.$$

Therefore, replacing F by its estimator $\widehat{F}_n$, we obtain the analogue

$$S_n^* = \sqrt{n}(\overline{X}_n^* - \overline{X}_n),$$

of $S_n^+ = \sqrt{n}(\overline{X}_n^+ - \mu)$, where $\overline{X}_n^* = n^{-1}\sum_{j=1}^n X_j^*$ is the sample mean based on the bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$.

Note that $\overline{X}_n$, replacing μ, is the expectation of X_1^*.

What the bootstrap proposes, is nothing else but to estimate the distribution of S_n by the distribution of S_n^*.

It might again be difficult to calculate the distribution of S_n^* analytically. But because $\widehat{F}_n$ is **known**, we can approximate the distribution of S_n^* by Monte Carlo.

That is, we generate many copies, say $S_{n,1}^*, S_{n,2}^*, \ldots, , S_{n,B}^*$ of S_n^*, where $S_{n,b}^* = \sqrt{n}(\overline{X}_{n,b}^* - \overline{X}_n)$ and $\overline{X}_{n,b}^* = n^{-1}\sum_{j=1}^n X_{j,b}^*$.

Recall that each bootstrap sample consists of n independent random variables X_j^* stemming from $\widehat{F}_n$.

Then we can estimate the distribution function G_n of S_n (in which we are interested in) by means of the empirical distribution function

$$\widehat{G}_n^*(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{(-\infty,x]}(S_{n,b}^*).$$

At first glance, the double stochastic structure of the bootstrap, namely on the one hand the original random variables $X_1, \ldots, X_n$ and on the other hand the bootstrap random variables $X_1^*, \ldots, X_n^*$, may look a bit strange or confusing.
Considering bootstrap as an additional random experiment, that heavily depends on and varies with the underlying sample $X_1, \ldots, X_n$, hopefully makes things more clear.

The distribution of $S_n^* = \sqrt{n}\,(\overline{X}_n^* - \overline{X}_n)$ depends on the observed data $X_1, \ldots, X_n$ and therefore is random itself (a random measure).

Note that in the expression $S_n^* = \sqrt{n}\,(\overline{X}_n^* - \overline{X}_n)$ evaluated conditional on $X_1, X_2, \ldots, X_n$, the quantity $\overline{X}_n^*$ remains a random variable (on the bootstrap probability space $(\mathcal{X}^*, \mathcal{F}^*, P^*)$) but $\overline{X}_n$ is not!

The distribution of S_n^* on the statistical bootstrap experiment is therefore nothing else but a (complicated and measure-valued) random quantity depending on the underlying sample $(X_1, \ldots, X_n) =: \underline{X}_n$.

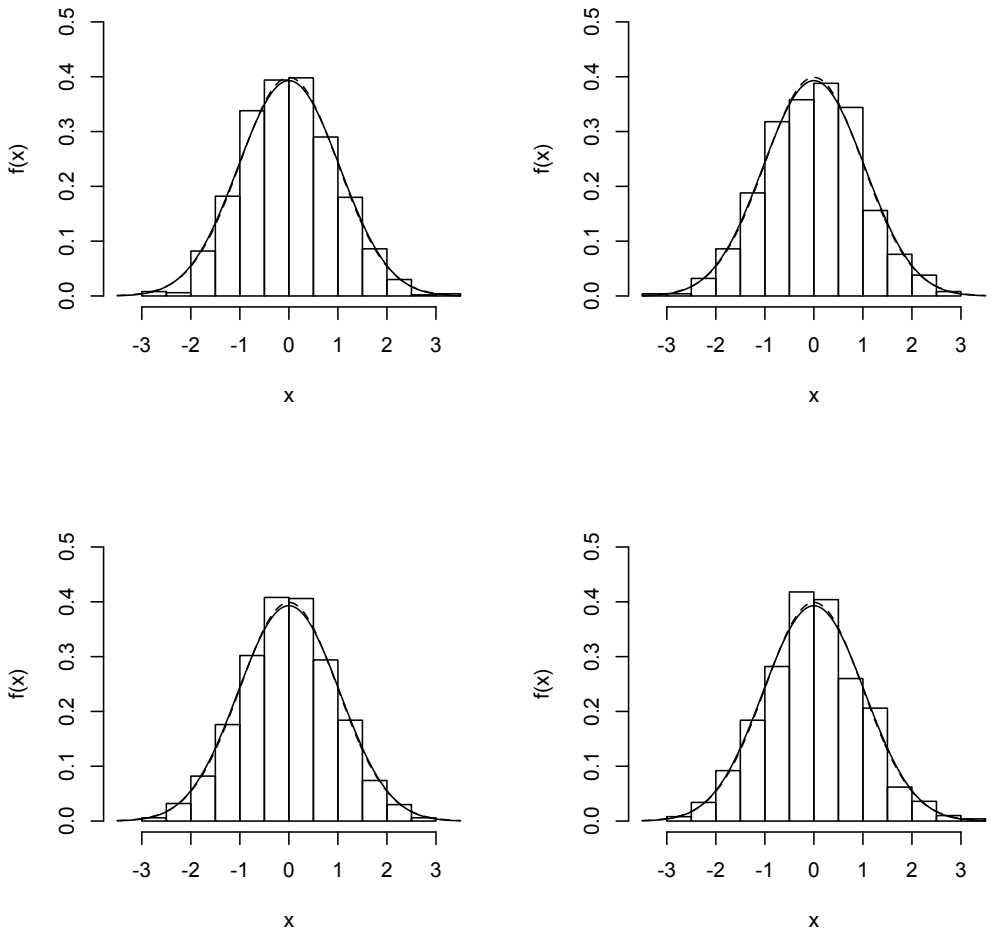
It is intended to serve as an estimator of the distribution of $S_n = \sqrt{n}(\overline{X}_n - \mu)$.

The notation $\mathcal{L}(S_n^* | \underline{X}_n)$, or in this specific case $\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n) | \underline{X}_n)$, is intended to emphasize this and will be used throughout the presentation.

The estimation of the distribution of of $\mathcal{L}(S_n^* | \underline{X}_n)$ by Monte Carlo is only a step which is due to the fact that analytical calculation of the distribution of S_n^* is typically difficult if not impossible.

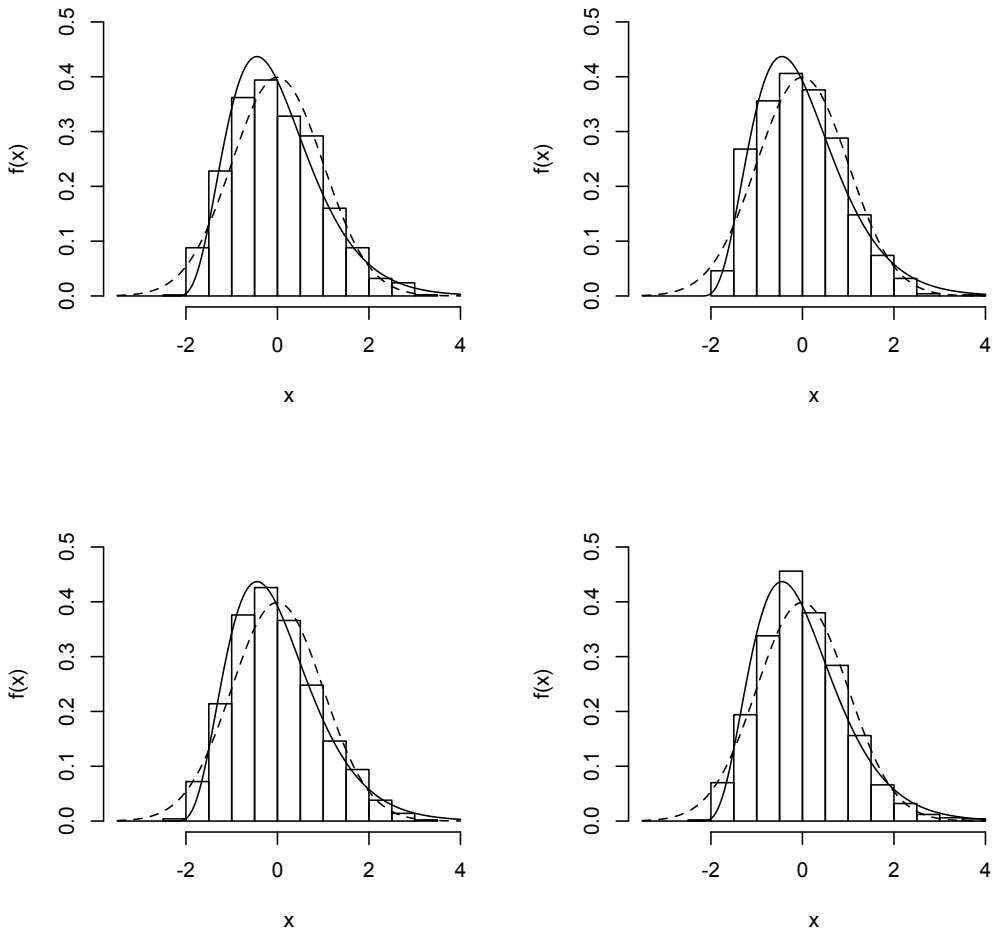
Let us illustrate the preceding discussion by bootstrap approximations for a standardized version $L_n := \sqrt{n}(\overline{X}_n - \mu)/\sigma_X$ of S_n in two examples.

An Example



Uniform case: Distribution of L_n and its different approximations based on four randomly selected samples of length $n = 10$. Solid line is the density of the exact distribution of L_{10}, dashed line is the density of the standard Gaussian distribution and the histograms refer to the bootstrap approximations.

An Example



Chi-square case: Distribution of L_n and its different approximations based on four randomly selected samples of length $n = 10$. Solid line is the density of the exact distribution of L_{10}, dashed line is the density of the standard Gaussian distribution and the histograms refer to the bootstrap approximations.

Bootstrap Procedure

A common way to describe an inference problem in mathematical statistics is by means of a statistical experiment, that is a probability space $(\Omega_n, \mathcal{A}_n, P_n)$ and a sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$ taking values in Ω_n. Typically,

1. $\Omega_n = \mathbb{R}^n$, the sample space, i.e., the set of all possible realizations of the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$,
2. $\mathcal{A}_n$ a σ-algebra of subsets of Ω_n, if $\Omega_n = \mathbb{R}^n$ then $\mathcal{A}_n = \mathbb{B}_n$, the Borel σ-algebra on $\mathbb{R}^n$, and,
3. P_n an *unknown* probability measure on $\mathcal{A}_n$ which describes the (joint) distribution of $\underline{X}_n = (X_1, X_2, \ldots, X_n)$.

$(\Omega_n, \mathcal{A}_n, P_n)$: Underlying statistical experiment.

Consider a statistic, that is a measurable function

$$T_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}^d, \mathbb{B}_d).$$

We often write T_n instead of $T_n(X_1, X_2, \ldots, X_n)$.

We are typically interested in the distribution of a properly centered and scaled version of the statistic T_n or at least in moments of this distribution, like for instance the variance or the standard deviation.

To elaborate, assume that a sequence $(r_n, n \in \mathbb{N})$ of non-negative real numbers increasing to infinity and a corresponding centering sequence $(\theta_n, n \in \mathbb{N})$ exists such that the sequence of random variables $(S_n := r_n(T_n - \theta_n), n \in \mathbb{N})$, converge weakly to a non-degenerated limit distribution G, that is

$$\mathcal{L}(S_n) \xrightarrow{\mathcal{D}} G,$$

as $n \to \infty$.

In order to infer properties of T_n or S_n one can of course try to estimate the distribution $P_n = \mathcal{L}(X_1, X_2, \ldots, X_n)$ of the sample $(X_1, X_2, \ldots, X_n)$ and then, based on this estimator to estimate the distribution of T_n, respectively, S_n.

However, (asymptotically) consistent estimation of P_n is a rather hard task and it is typically only feasible in limited situations (i.i.d. set up).

Moreover, the actual distributions of interest, that is, the distribution of T_n, respectively, of S_n, often do not depend on all the information about P_n, but only certain relevant features or characteristics of it. Now, the bootstrap is a methodology to approximate the finite sample distributions $\mathcal{L}(T_n)$ or $\mathcal{L}(S_n)$.

To do so, the basic idea is the following:

Instead of estimating the entire distribution P_n itself, a "bootstrap" or "pseudo" statistical experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$, which depends on the observed sample $X_1, X_2, \ldots, X_n$, is proposed and corresponding "bootstrap" or "pseudo" random variables $X_i^* : (\Omega_n^*, \mathcal{A}_n^*) \to (\Omega_n, \mathcal{A}_n)$, $i = 1, 2, \ldots, n$, are considered, such that $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^*)$, respectively $S_n^* = r_n(T_n^* - \theta_n^*)$ are able to mimic relevant features of T_n, respectively, S_n. Here θ_n^* is a suitable centering sequence.

θ_n^* may also depend on the underlying sample $X_1, X_2, \ldots, X_n$.

The symbol * in the notation θ_n^* is solely used in order to indicate that this is the sequence used to properly center the bootstrap statistic T_n^*. In the context of the pseudo statistical experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$,

1. Ω_n^* denotes the set containing at least all possible realizations of the bootstrap sample $\underline{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$,
2. $\mathcal{A}_n^*$ is a σ-algebra on Ω_n^*, and,
3. P_n^* is a *known* probability measure on $\mathcal{A}_n^*$ which describes the (joint) distribution of $(X_1^*, X_2^*, \ldots, X_n^*)$.

Essentially, what changes by switching from the underlying statistical experiment $(\Omega_n, \mathcal{A}_n, P_n)$ to the new, pseudo-statistical experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$ is that the unknown probability measure P_n generating the sample $X_1, \ldots, X_n$, has been replaced by a known probability measure P_n^*, which mimics relevant features of P_n and in particular those affecting the distribution $\mathcal{L}(T_n)$, respectively, $\mathcal{L}(S_n)$.

The rationale behind the bootstrap idea is that if the pseudo-statistical experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$ succeeds in sufficiently well mimicking relevant features of the underlying statistical experiment $(\Omega_n, \mathcal{A}_n, P_n)$, then the distribution of S_n^* under the probability measure P_n^* will be close to that of S_n under P_n.

P_n^* is therefore, not (necessarily) an estimator of P_n but a known probability measure based on the sample $(X_1, X_2, \ldots, X_n)$, which mimics sufficiently well relevant features of P_n.

Note that this replacement of P_n by the known probability measure P_n^* in the bootstrap proposal, simplifies the statistical inference problem considerably.

To elaborate, recall that $\mathcal{L}(X_1^*, X_2^*, \ldots, X_n^*) = P_n^*$ and that $S_n^* = r_n(T_n^* - \theta_n^*)$ is the bootstrap analogue of S_n, with $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^*)$.

Furthermore, the unknown distribution P_{n,S_n} (or just P_{S_n} for short) of $S_n = r_n(T_n - \theta_n)$ is approximated by the corresponding distribution $P_{n,S_n^*}^* = \mathcal{L}(S_n^* | \underline{X}_n)$ (for short $P_{S_n^*}^*$).

Now since P_n^* is known, one can generate many bootstrap samples $X_1^*, X_2^*, \ldots, X_n^*$, and consequently many copies of the bootstrap statistic S_n^*.
This allows to estimate $P_{S_n^*}^*$ by using the empirical distribution of the many copies of S_n^* generated.

In some relevant cases, it could be that the bootstrap statistic T_n^* or S_n^* is not explicitly defined via a bootstrap sample, i.e. according to $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^*)$, but it is defined directly as a statistic on a skillfully chosen bootstrap experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$.

Having such situations also in mind, we avoided mentioning explicitly a bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ in the following Definition and in Diagram 1.1 below.

For a comprehensive understanding of the bootstrap, however, it is good to imagine, at least initially, the explicit situation in which the bootstrap statistic $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^*)$ is defined via an explicit bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ mimicking sufficiently well relevant features of the underlying sample $X_1, X_2, \ldots, X_n$.

The previous discussion lead us to the following formal definition of a bootstrap procedure.

Definition 1

Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a statistical experiment and let $T_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}^d, \mathbb{B}_d)$, $T_n = T_n(X_1, X_2, \ldots, X_n)$, be a statistic, where $\underline{X}_n = (X_1, X_2, \ldots, X_n)$ is a sample stemming from $(\Omega_n, \mathcal{A}_n, P_n)$.

Consider a statistical experiment $(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$, where P_n^* is a probability measure on $\mathcal{A}_n^*$ depending on $(X_1, X_2, \ldots, X_n)$, and a statistic $T_n^* : (\Omega_n^*, \mathcal{A}_n^*) \to (\mathbb{R}^d, \mathbb{B}_d)$.

$(\Omega_n^*, \mathcal{A}_n^*, P_n^*)$ together with T_n^* is called a *bootstrap or pseudo experiment* or a *bootstrap procedure*.

T_n^* is called a *bootstrap statistic*.

The following Diagram 1.1 illustrates the relation between the underlying statistical experiment and its bootstrap counterpart.
It demonstrates how the bootstrap experiment builds upon the original statistical experiment, in that the bootstrap probability measure P_n^* is based on the sample $X_1, X_2, \ldots, X_n$.

Every subsequent statistical analysis within the bootstrap statistical experiment depends, therefore, on the observed sample $X_1, X_2, \ldots, X_n$. This is also true for bootstrap statistics T_n^* and their distributions.

After a proper centering and scaling, the distribution $P^*_{S_n^*}$ of $S_n^* = r_n(T_n^* - \theta_n^*)$ under P_n^*, is used to estimate the distribution P_{S_n} of $S_n = r_n(T_n - \theta_n)$, of interest.

Notice that, typically, it is the distribution of S_n under P_n in which we are interested in, and not necessarily the probability measure P_n itself.

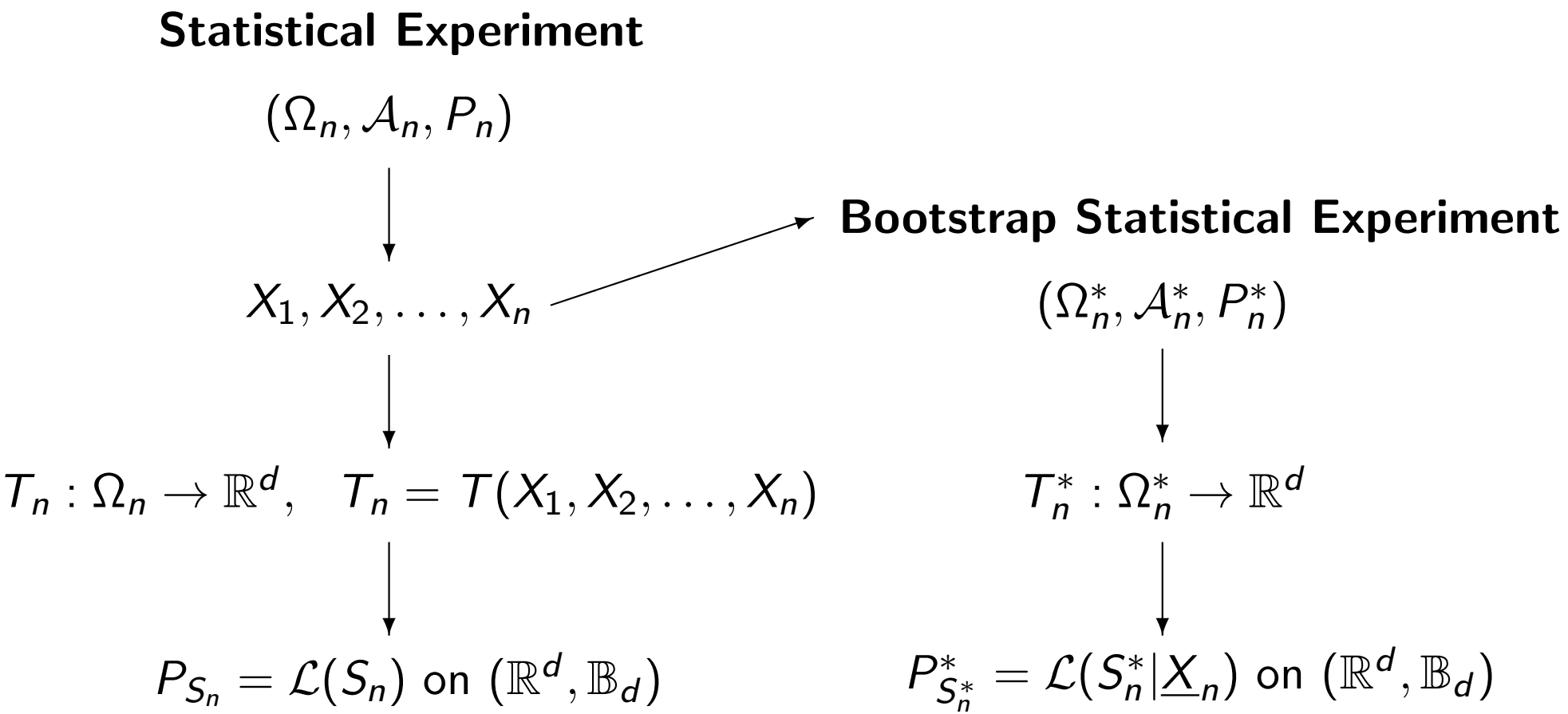


Diagram 1.1: Schematic representation of underlying and bootstrap statistical experiment.

That is, the success of a bootstrap proposal does not necessarily depend on how well the bootstrap experiment $(\Omega_n^*, \mathcal{F}_n^*, P_n^*)$ or the probability measure P_n^* replicates the entire structure of the statistical experiment $(\Omega_n, \mathcal{F}_n, P_n)$ or of P_n.

As a matter of fact, in complex situations it could be very hard or even impossible to imitate the entire structure of P_n by that of P_n^*.

Bootstrap Asymptotics

We want to be able to prove properties of functions of S_n^* or better of $\mathcal{L}(S_n^*|\underline{X}_n)$.

Relevant functions of S_n^* are for example $\mathrm{E}^*(S_n^*)$, $\mathrm{Var}^*(S_n^*)$ or even the complete distribution $\mathcal{L}(S_n^*|\underline{X}_n)$, where we denote by E^*, Var^* and Cov^* the expectation, variance and covariance under P_n^*.

All quantities are considered as random quantities on $(\Omega_n, \mathcal{A}_n, P_n)$ taking values in $\mathbb{R}$ or $\mathbb{R}^d$ or (in the latter case) in the space of probability measures on $\mathbb{R}$ or $\mathbb{R}^d$.

Notions like $\mathrm{E}\big(\mathrm{E}^* S_n^*\big)$ or $P\{|\mathrm{Var}^*(S_n^*) - a| > \epsilon\}$, $P_n^*\{|S_n^*| > \epsilon\} \to 0$ a.s. and $P_n^*\{|S_n^*| > \epsilon\} \to 0$ in P_n-probability would then make sense.

We would like to prove results like:

The following applies

$$d_K(\mathcal{L}(S_n^*|\underline{X}_n), \mathcal{L}(S_n)) = o_{P_n}(1). \tag{2}$$

Here $d_K(P_n, P) := sup_{x\in\mathbb{R}}|P_n((-\infty, x)) - P((-\infty, x))|$ denotes the Kolmogorov-distance.

We typically know

$$\mathcal{L}(S_n) \xrightarrow{\mathcal{D}} Q,$$

where Q often is a normal distribution.

It then suffices to show

$$\mathcal{L}(S_n^*|\underline{X}_n) \xrightarrow{\mathcal{D}} Q \text{ in } P_n\text{-probability}.$$

But what does this mean?

Definition 2

A sequence $(S_n^*, n \in \mathbb{N})$ with S_n^* defined on Ω_n^*, $n \in \mathbb{N}$, has a feature M in P_n-probability (abbreviated: S_n^* satisfies M in P_n-probability)

if for all $\epsilon > 0$ and all $n \in \mathbb{N}$, sets $B_{n,\epsilon} \in \mathcal{A}_n$ with $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$ exist, such that for all sequences $\{\underline{x}_n, n \in \mathbb{N}\}$ with $\underline{x}_n \in B_{n,\epsilon}$,

$$\{S^*_{n,\underline{x}_n}, n \in \mathbb{N}\} \text{ has feature } M.$$

See Lohse (1984).

Does this make sense?

Lemma 3

For $\mathbb{R}^d$-valued random elements Z_n on $(\Omega_n, \mathcal{A}_n, P_n)$ the following two statements are equivalent.

(i) $Z_n \xrightarrow{P_n} a$ *as* $n \to \infty$, *that is, for all* $\delta > 0$, $P_n(\|Z_n - a\| \geq \delta) \to 0$, *as* $n \to \infty$.

(ii) *For all* $\epsilon > 0$ *and all* $n \in \mathbb{N}$, *there exist sets* $B_{n,\epsilon} \in \mathcal{A}_n$ *with* $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$, *such that for all sequences* $(\underline{x}_n, n \in \mathbb{N})$, *satisfying* $\underline{x}_n \in B_{n,\epsilon}$,

$$Z_n(\underline{x}_n) \to a \text{ as } n \to \infty$$

holds true. Here "$\to$" stands for usual convergence in $\mathbb{R}^d$.

Remark 1

*(i) For random quantities Z_n derived from S_n^* (**for example like** $\mathbf{Var^*(S_n^*)}$), which are usual random variables on $(\Omega_n, \mathcal{A}_n, P_n)$, we immediately have from the above Lemma and Definition that the following two statements are equivalent*

$$(Z_n) \text{ has the feature "convergence to } a \in \mathbb{R}^d \text{" in } P_n\text{-probability}$$
$$\text{(abbreviated: } Z_n - a = o(1) \text{ in } P_n\text{-probability)}$$

and

$$Z_n \xrightarrow{P_n} a \quad \text{as } n \to \infty \text{ (abbreviated: } Z_n - a = o_{P_n}(1)).$$

The advantage of the above Definition is, that it easily extends to other types of convergence and even more general "features" as well.

For example, it clarifies what weak convergence of $\mathcal{L}(S_n^*|\underline{X}_n)$ (including asymptotic normality) in P_n-probability, and $S^*_{n,\underline{X}_n} = o_{P_n^*}(1)$, or $\mathcal{O}_{P_n^*}(1)$, in P_n-probability, means.

(ii) For easy reference we specialize the definition to the following cases relevant for the bootstrap. Let (a_n) and (b_n) be non-negative sequences of real numbers.

(I) $S_n^* = o_{P_n^*}(a_n)$ in P_n-probability means, that
$\forall \epsilon > 0$ and $\forall n \in \mathbb{N}\ \exists B_{n,\epsilon} \in \mathcal{A}_n$ with $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$,
such that $S^*_{n,\underline{x}_n} = o_{P_n^*}(a_n)$ for all sequences
$\{\underline{x}_n, n \in \mathbb{N}\}$ with $\underline{x}_n \in B_{n,\epsilon}$.

(II) $S_n^* = \mathcal{O}_{P_n^*}(b_n)$ in P_n-probability means, that
$\forall \epsilon > 0$ and $\forall n \in \mathbb{N}\ \exists B_{n,\epsilon} \in \mathcal{A}_n$ with $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$,
such that $S^*_{n,\underline{x}_n} = \mathcal{O}_{P_n^*}(b_n)$ for all sequences
$\{\underline{x}_n, n \in \mathbb{N}\}$ with $\underline{x}_n \in B_{n,\epsilon}$.

and

(III) $\mathcal{L}(S_n^*|\underline{X}_n) \xrightarrow{\mathcal{D}} Q$ in P_n-probability means, that
$\forall \epsilon > 0$ and $\forall n \in \mathbb{N}\ \exists B_{n,\epsilon} \in \mathcal{A}_n$ with $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$,
such that $\mathcal{L}(S_n^*|\underline{X}_n = \underline{x}_n) \xrightarrow{\mathcal{D}} Q$ for all sequences
$\{\underline{x}_n, n \in \mathbb{N}\}$ with $\underline{x}_n \in B_{n,\epsilon}$.

(iii) Clearly, if a sequence (S_n^*) has a feature M almost surely, i.e. the feature is fulfilled on a set of probability one, then obviously (S_n^*) has the feature M in P_n-probability as well.

We can prove results like

Theorem 4

In the situation of Diagram 1.1 we have

(i) *For any non-negative sequence* (a_n)*:*

$$\begin{aligned} S_n^* = o_{P_n^*}(a_n) \text{ in } P_n\text{-probability} \\ \iff \forall \delta > 0 : \; P_n^* \left\{ \frac{\|S_n^*\|}{a_n} > \delta \right\} = o_{P_n}(1). \end{aligned} \tag{3}$$

(ii) *Assume that the probability measure* Q *has a continuous cumulative distribution function. Then,*

$$\mathcal{L}(S_n^* | \underline{X}_n) \xrightarrow{\mathcal{D}} Q \text{ in } P_n\text{-probability} \Leftrightarrow d_K(\mathcal{L}(S_n^* | \underline{X}_n), Q) = o_{P_n}(1). \tag{4}$$

Example 5

We return to the sample mean of i.i.d. random variables $X_1, X_2, \ldots$ taking values in $(\mathbb{R}, \mathbb{B})$ with $\mathrm{E}X_1 = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$.

We have $\mathrm{E}(\sqrt{n}(\overline{X}_n - \mu)) = 0$, $\mathrm{Var}(\sqrt{n}(\overline{X}_n - \mu)) = \sigma^2$ and from the central limit theorem of Lindeberg-Lévy

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Now suppose that we have observed a sample $X_1, \ldots, X_n$ and have defined a bootstrap sample $X_1^*, \ldots, X_n^*$ by drawing independently and with replacement from the original sample.

It can easily be seen that $\mathrm{E}^*(X_i^*) = \overline{X}_n$, which immediately leads to $\mathrm{E}^*(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)) = 0$.

Similarly we get, left as excecise,

$$\mathrm{Var}^*\left(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)\right) = \mathrm{Var}^*(X_1^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Since under our assumptions $(1/n)\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \to \sigma^2$ in probability (even almost surely), which is nothing else but $(1/n)\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \to \sigma^2$ in P_n-probability in the spirit of the definition, we conclude that $\mathrm{Var}^*(\sqrt{n}(\overline{X}_n^* - \overline{X}_n))$ converges towards σ^2 in P_n-probability .

We can even show that

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{L}(\sqrt{n}(\overline{X}_n - \mu))\right) \to 0 \text{ in } P_n\text{-probability.}$$

As mentioned earlier, the notation $\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n)$ is to make clear that this distribution is a random quantity that depends on the observed sample $\underline{X}_n = (X_1, \ldots, X_n)$. It is a distribution in the statistical bootstrap experiment, where everything depends on the observed underlying sample. In particular, $\overline{X}_n$ is not random in this context. This is different for $\mathcal{L}(\sqrt{n}(\overline{X}_n - \mu))$, which denotes the distribution of the sample mean in the underlying statistical experiment.
The assertion

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{L}(\sqrt{n}(\overline{X}_n - \mu))\right) \to 0 \text{ in } P_n\text{-probability},$$

typically is called **asymptotic consistency of the bootstrap**; in this case for the sample mean.

To check

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{L}(\sqrt{n}(\overline{X}_n - \mu))\right) \to 0 \text{ in } P_n\text{-probability},$$

one can apply the triangle inequality and note that

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n - \mu)), \mathcal{N}(0, \sigma^2)\right) \to 0,$$

due to the central limit theorem of Lindeberg-Lévy.

It then remains to show

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{N}(0, \sigma^2)\right) \to 0 \text{ in } P_n\text{-probability}.$$

In the end the last statement (which we have to show) is a consequence of the central limit theorem of Lindeberg-Lévy for row-wise i.i.d. triangular arrays.

In particular we show that for every $\epsilon > 0$ and every $n \in \mathbb{N}$ we can find $B_{n,\epsilon} \in \mathbb{B}_n$, satisfying $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$, such that for all sequences $(\underline{x}_n, n \in \mathbb{N})$, with $\underline{x}_n \in B_{n,\epsilon}$,

$$d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)\middle|\underline{X}_n = \underline{x}_n), \mathcal{N}(0, \sigma^2)\right) \to 0. \tag{5}$$

Recall the notation $\mathcal{L}(\cdot|\underline{X}_n = \underline{x}_n)$, which denotes the distribution under the probability measure P_n^* in the specific case in which the underlying original sample is $\underline{x}_n$. In the same way, expectations $\mathrm{E}^*(\cdot)$ are also random variables depending on $\underline{X}_n$.

The convergence (5) follows if the assumptions of a CLT can be fulfilled in P_n-probability! That is

$$\mathrm{E}^*\left((X_1^* - \overline{X}_n)^2\right) \to \sigma^2, \tag{6}$$

in P_n-probability,

and the Lindeberg-Condition (also in P_n-probability)

$$\mathrm{E}^*\left[(X_1^* - \overline{X}_n)^2 \mathbb{1}\{|X_1^* - \overline{X}_n| > \delta\sqrt{n}\}\right] \to 0. \tag{7}$$

In order to establish (6), note that
$\mathrm{E}^*(X_1^* - \overline{X}_n)^2 = (1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 \xrightarrow{P_n} \sigma^2$ by the WLLN.
So we find for every given $\epsilon > 0$ and every $n \in \mathbb{N}$ sets $B_{n,\epsilon}^{(1)} \in \mathbb{B}_n$ with $P_n(B_{n,\epsilon}^{(1)}) \geq 1 - \epsilon/2$ such that for all sequences $(\underline{x}_n)$, with $\underline{x}_n \in B_{n,\epsilon}^{(1)}$, we have $\mathrm{E}^*\left((X_1^* - \overline{X}_n)^2\right) \to \sigma^2$ (here $\to$ denotes usual convergence in $\mathbb{R}$).
Along the same lines we obtain from

$$\begin{aligned}
&\mathrm{E}^*\left[(X_1^* - \overline{X}_n)^2 \mathbb{1}\{|X_1^* - \overline{X}_n| > \delta\sqrt{n}\}\right] \\
&\qquad = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \mathbb{1}\{|X_i - \overline{X}_n| > \delta\sqrt{n}\} \xrightarrow{P_n} 0, \text{ as } n \to \infty,
\end{aligned}$$

which is not complicated but a bit tedious to verify, the existence of sets $B_{n,\epsilon}^{(2)}$ with $P_n(B_{n,\epsilon}^{(2)}) \geq 1 - \epsilon/2$ such that for $(\underline{x}_n)$, with $\underline{x}_n \in B_{n,\epsilon}^{(2)}$, (7) is fulfilled.
Let $B_{n,\epsilon} := B_{n,\epsilon}^{(1)} \cap B_{n,\epsilon}^{(2)}$ and observe $P_n(B_{n,\epsilon}) \geq 1 - \epsilon$. For every sequence $(\underline{x}_n)$, with $\underline{x}_n \in B_{n,\epsilon}$, we have (6) and (7) by construction. From this (5) follows.

By the definition of a "feature in P_n-probability" this is equivalent to $d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{N}(0, \sigma^2)\right) \to 0$ in P_n-probability.

Finally, we obtain altogether, which once again is nothing else but the usual convergence of $d_K\left(\mathcal{L}(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n), \mathcal{L}(\sqrt{n}(\overline{X}_n - \mu))\right)$ to zero in probability.
Also Lyapunov's Condition can be stated in P_n-probability, which for the specific situation considered here is much easier to verify.

Theorem 6

Assume that $Z^*_{j,n}, j = 1, 2, \ldots, k(n), n \in \mathbb{N}$, *is a triangular array of random variables on* $(\Omega^*_n, \mathcal{A}^*_n, P^*_n)$ *taking values in* $\mathbb{R}^d$, *which are for each fixed* $n \in \mathbb{N}$ *mutually independent and have a distribution* $\mathcal{L}(Z^*_{j,n}|\underline{X}_n)$, *which depends on* $\underline{X}_n$. *Further, assume that, as* $n \to \infty$ $\sum_{j=1}^{k(n)} \mathrm{Cov}^*\left(Z^*_{j,n}\right) \xrightarrow{P_n} \Sigma$, *for a positive-definite matrix* Σ, *and that the Lindeberg condition,*

$$(LC) \quad \sum_{j=1}^{k(n)} \mathrm{E}^*\left[\|Z^*_{j,n} - \mathrm{E}^*(Z^*_{j,n})\|^2 \mathbb{1}\{\|Z^*_{j,n} - \mathrm{E}^*(Z^*_{j,n})\| > \varepsilon\}\right] \xrightarrow{P_n} 0, \forall \epsilon > 0, \tag{8}$$

is satisfied. Then,

$$\mathcal{L}\left(\sum_{j=1}^{k(n)} (Z^*_{j,n} - \mathrm{E}^*(Z^*_{j,n}))\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma) \quad \textit{in } P_n\textit{-probability.} \tag{9}$$

Bootstrap Consistency

We say, that, the bootstrap proposal S_n^* is (asymptotically) **consistent** for S_n if for some distance measure d

$$d\big(\mathcal{L}(S_n), \mathcal{L}(S_n^*|\underline{X}_n)\big) \to_{n\to\infty} 0 \text{ in } P_n\text{-probability}$$

applies.

Instead of "(asymptotically) consistent" we also say "valid" or simply that "bootstrap works".

Another very useful distance measure (in addition to the Kolmogorov distance d_K) for investigating properties of bootstrap methods is the so-called **Mallows or Wasserstein distance** which is described in the following definition.

Definition 7

For $r \geq 1$ the Mallows distance $d_r : \mathcal{F}_r \times \mathcal{F}_r \to [0, +\infty)$ is defined by

$$d_r(P_1, P_2) := \inf_{\mathcal{L}(X_1, X_2)} \left(\mathrm{E} \left\| X_1 - X_2 \right\|^r \right)^{1/r},$$

where the infimum is taken over all distributions $\mathcal{L}(X_1, X_2)$ of the pair (X_1, X_2) such that $X_1 \sim P_1$ and $X_2 \sim P_2$. This means that the marginal distributions of (X_1, X_2) are fixed but the dependence between X_1 and X_2 is completely unrestricted.

Lemma 8

d_r is a metric on $\mathcal{F}_r$.

and

Lemma 9

Let F_1 and F_2 be distribution functions on $\mathbb{R}$ with corresponding probability measures P_1 and P_2, respectively. Assume that $P_1, P_2 \in \mathcal{F}_r$ for an arbitrary $r \geq 1$. Then,

$$d_r(P_1, P_2) = \left(\int_0^1 \left| F_1^{-1}(x) - F_2^{-1}(x) \right|^r dx \right)^{1/r}.$$

For $r = 1$ the following additional equality applies

$$d_1(P_1, P_2) = \int_0^1 \left| F_1^{-1}(x) - F_2^{-1}(x) \right| dx = \int_{\mathbb{R}} \left| F_1(x) - F_2(x) \right| dx\,.$$

Lemma 10

Let $\{P_n, n \in \mathbb{N}\}$ be a sequence of probability measures and P a probability measure on $(\mathbb{R}^d, \mathbb{B}_d)$ all of which are in $\mathcal{F}_r$ for some $r \geq 1$. Then, $d_r(P_n, P) \to 0$ as $n \to \infty$ is equivalent to

$$P_n \xrightarrow{\mathcal{D}} P \quad \text{and} \quad \int \|x\|^r dP_n \to \int \|x\|^r dP \text{ as } n \to \infty. \tag{10}$$

$P_n \xrightarrow{\mathcal{D}} P$ is defined as $F_n \xrightarrow{\mathcal{D}} F$, where F_n, F denote the cumulative distribution functions of P_n, P, respectively.

The following lemma shows that under bootstrap consistency, quantiles of the bootstrap distribution function can be used, if the distribution function of the target distribution is continuous; see Müller (2017).

Lemma 11

Under bootstrap consistency and assuming that $\mathcal{L}(S_n) \xrightarrow{\mathcal{D}} Z$ *and* $\mathcal{L}(S_n^*|\underline{X}_n) \xrightarrow{\mathcal{D}} Z$ *in* P_n*-probability, for some real-valued random variable* Z*. Further assume that the distribution function* F_Z *of* Z *is continuous. Let* $F_{S_n^*}$ *be the distribution function of* S_n^* *and for* $\beta \in (0,1)$*, denote by* $q_{n,\beta}^*$ *the* β*-quantile of the distribution of* S_n^**, that is,*

$$q_{n,\beta}^* = \inf\{y \in \mathbb{R} \,|\, F_{S_n^*}(y) \geq \beta\}.$$

It then holds true that

$$\lim_{n\to\infty} P_n\big(S_n \leq q_{n,\beta}^*\big) = \beta.$$

Theorem 12

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a distribution with distribution function $F \in \mathcal{F}_2$. Moreover let $X_1^+, X_2^+, \ldots, X_n^+$ be a bootstrap sample, where the X_j^+'s also are i.i.d. but with distribution function $\widetilde{\mathbf{F}}_\mathbf{n}$, an estimator of F depending on the (observed) sample $\underline{X}_n$.

If $\widetilde{F}_n$ satisfies $d_2(\widetilde{F}_n, F) \to 0$, in P_n-probability, as $n \to \infty$, then

$$d_2\Big(\mathcal{L}\big(\sqrt{n}(\overline{X}_n - \mathrm{E}(X_1))\big), \mathcal{L}\big(\sqrt{n}(\overline{X}_n^+ - \mathrm{E}^*(X_1^+))\big)\Big) \to 0 \text{ in } P_n\text{-probability},$$

where $\overline{X}_n^+ = n^{-1}\sum_{j=1}^n X_j^+$.

That means, the bootstrap proposal (asymptotically) works.

Consider a linear regression model with stochastic design given by

$$V_j = \theta U_j + \varepsilon_j, \quad j = 1, \ldots, n, \tag{11}$$

where U_j and ε_j, $j = 1, 2, \ldots, n$, are i.i.d. random variables.
We also assume that all U_j's are independent of the ε_j's, that $\mathrm{E}\varepsilon_1 = 0$, $0 < \mathrm{E}(U_1^2) = \tau_U^2 < \infty$, $\mathrm{E}(U_j^4) < \infty$ and that $0 < \mathrm{E}(\varepsilon_1^2) = \mathrm{Var}(\varepsilon_1) = \sigma_\varepsilon^2 < \infty$.

Consider the least squares estimator of the parameter θ obtained by minimizing $\sum_{j=1}^n (V_j - c\,U_j)^2$ with respect to $c \in \mathbb{R}$. Simple algebra leads to the estimator

$$\widehat{\theta}_n = T_n = \sum_{j=1}^n U_j V_j \Big/ \sum_{r=1}^n U_r^2. \tag{12}$$

Consider next the sequence

$$S_n = \sqrt{n}(T_n - \theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j \varepsilon_j \Big/ \Big(\frac{1}{n} \sum_{r=1}^n U_r^2\Big). \tag{13}$$

By the weak law of large numbers and by the central limit theorem for i.i.d. random variables, we get that

$$\frac{1}{n}\sum_{r=1}^{n} U_r^2 \xrightarrow{P_n} \tau_U^2 \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{j=1}^{n} U_j\varepsilon_j \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \tau_U^2\sigma_\varepsilon^2\right),$$

respectively. Combining these results using Slutsky's theorem we get

$$S_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\varepsilon^2/\tau_U^2\right),$$

as $n \to \infty$, which also shows that $r_n = \sqrt{n}$ is an appropriate choice for the rate of standardization.

We are now interested in designing a bootstrap procedure for estimating the distribution of S_n. Three such procedures are discussed in the following and are denoted by Bootstrap Proposal I, II and III.

The independence of the pairs (V_j, U_j), $j = 1, 2, \ldots, n$, motivates the following bootstrap procedure.
Generate a bootstrap sample of n pairs by drawing with replacement from the set $\{(V_j, U_j), j = 1, 2, \ldots, n\}$ of available pairs. Denote by (V_j^*, U_j^*), $j = 1, 2, \ldots, n$, the pairs selected.
We may think of (V_j^*, U_j^*) as generated by $(V_j^*, U_j^*) = (V_{I_j^*}, U_{I_j^*})$, $j = 1, 2, \ldots, n$, where the random variables $I_1^*, I_2^*, \ldots, I_n^*$ are i.i.d. having the uniform distribution over the set $\{1, 2, \ldots, n\}$.

After generating the bootstrap pairs, one can define the bootstrap analogue

$$\widehat{\theta}_n^* = T_n^* = \sum_{j=1}^{n} U_{I_j^*} V_{I_j^*} \Big/ \sum_{j=1}^{n} U_{I_j^*}^2,$$

of the least squares estimator $\widehat{\theta}_n$.

As a normalized bootstrap estimator one can then consider

$$S^*_{1,n} = \sqrt{n}\big(T^*_n - \widehat{\theta}_n\big) = \sqrt{n}\Big(\sum_{j=1}^{n} U_{I^*_j} V_{I^*_j} \Big/ \sum_{j=1}^{n} U^2_{I^*_j} - \sum_{j=1}^{n} U_j V_j \Big/ \sum_{r=1}^{n} U^2_r\Big),$$

The distribution of $S^*_{1,n}$ can be used to estimate the distribution of S_n.

Let $\underline{(V,U)}_n := \big((V_j, U_j), j = 1, 2, \ldots, n\big)$. It can then be shown that

$$\mathcal{L}\big(S^*_{1,n} | \underline{(V,U)}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2_\varepsilon/\tau^2_U\big) \text{ in } P_n\text{-probability.} \qquad (14)$$

The proof of this statement is left as an exercise.

Observe that in Bootstrap Proposal I, only the independence of the pairs (V_j, U_j) has been used and not the structural relation between the random variable V_j and U_j as this is described by the linear regression equation (11).
One can of course take advantage of this knowledge in proposing a procedure to generate the bootstrap sample (V_j^*, U_j^*), $j = 1, 2, \ldots, n$.

Toward this goal, we first calculate the residuals

$$\widetilde{\varepsilon}_j = V_j - \widehat{\theta}_n U_j, \quad j = 1, 2, \ldots, n,$$

and define $\widehat{\varepsilon}_j = \widetilde{\varepsilon}_j - \overline{\varepsilon}_n$, where $\overline{\varepsilon}_n = \sum_{j=1}^n \widetilde{\varepsilon}_j / n$.

The centered residuals $\widehat{\varepsilon}_j$ will be used as proxies for the non observed random variables ε_j.

A random sample of n i.i.d. bootstrap variables $\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*$ is then generated by drawing with replacement from the set $\{\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n\}$.

That is, $\varepsilon_j^* = \varepsilon_{I_j^*}$, where $I_1^*, I_2^*, \ldots, I_n^*$ are i.i.d. random variables having a uniform distribution on the set $\{1, 2, \ldots, n\}$.

Note that, like the ε_j, the bootstrap variables ε_j^* are i.i.d. and that the centering by $\overline{\varepsilon}_n$ used in the definition of $\widehat{\varepsilon}_j$ ensures that

$$\mathrm{E}^*(\varepsilon_1^*) = \sum_{j=1}^{n} \widehat{\varepsilon}_j \cdot P(\varepsilon_1^* = \widehat{\varepsilon}_j) = \frac{1}{n} \sum_{j=1}^{n} \widehat{\varepsilon}_j = \frac{1}{n} \sum_{j=1}^{n} \left(\widetilde{\varepsilon}_j - \overline{\varepsilon}_n\right) = 0.$$

We also generate an i.i.d. bootstrap sample $U_1^*, U_2^*, \ldots, U_n^*$ of the regressors U_j by drawing with replacement from the set $\{U_1, U_2, \ldots, U_n\}$.

Given the bootstrap variables U_j^* and ε_j^*, $j = 1, 2 \ldots, n$, we obtain bootstrap replicates of V_j by using the linear relation, that is,

$$V_j^* = \widehat{\theta}_n U_j^* + \varepsilon_j^*, \quad j = 1, 2, \ldots, n.$$

A bootstrap analogue of the least squares estimator $\widehat{\theta}_n$ is then given by

$$\widehat{\theta}_n^* = T_n^* = \sum_{j=1}^{n} U_j^* V_j^* / \sum_{r=1}^{n} (U_r^*)^2$$

and the distribution of S_n can be estimated by the distribution of the bootstrap analogue

$$S_{2,n}^* = \sqrt{n}(T_n^* - \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j^* \varepsilon_j^* \Big/ \frac{1}{n} \sum_{r=1}^{n} (U_r^*)^2. \qquad (15)$$

Let us investigate the limiting distribution of the sequence $(S^*_{2,n})$.

First note that the following applies to the denominator of (15)

$$\frac{1}{n}\sum_{r=1}^{n}(U_r^*)^2 \xrightarrow{P_n^*} \tau_U^2 \text{ in } P_n\text{-probability.} \tag{16}$$

This follows since by the weak law of large numbers,

$$\mathrm{E}^*\big(n^{-1}\sum_{r=1}^{n}(U_r^*)^2\big) = \mathrm{E}^*(U_1^*)^2 = n^{-1}\sum_{j=1}^{n}U_j^2 \xrightarrow{P_n} \tau_U^2$$

Furthermore, by the weak law of large numbers and $\mathrm{E}(U_1^4) < \infty$

$$\mathrm{Var}^*\big(n^{-1}\sum_{r=1}^{n}(U_r^*)^2\big) = n^{-1}\mathrm{Var}^*((U_1^*)^2)$$

$$= n^{-1}\Big(\frac{1}{n}\sum_{j=1}^{n}U_j^4 - \big(\frac{1}{n}\sum_{j=1}^{n}U_j^2\big)^2\Big) = \mathcal{O}_{P_n}(n^{-1}).$$

Next, we show that

$$\mathcal{L}\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} U_j^*\varepsilon_j^* \Big| \underline{(V,U)}_n\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\sigma_\varepsilon^2\tau_U^2\big) \text{ in } P_n\text{-probability.} \qquad (17)$$

First, due to $\widetilde{\varepsilon}_j - \varepsilon_j = (\widehat{\theta}_n - \theta)U_j$, $\overline{\varepsilon}_n = \mathcal{O}_{P_n}(n^{-1/2})$ and

$$\begin{aligned}
\widehat{\sigma}_\varepsilon^2 := \mathrm{E}^*(\varepsilon_j^*)^2 &= \frac{1}{n}\sum_{j=1}^{n}(\widetilde{\varepsilon}_j - \overline{\varepsilon}_n)^2 = \frac{1}{n}\sum_{j=1}^{n}\widetilde{\varepsilon}_j^2 - (\overline{\varepsilon}_n)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j^2 + (\theta - \widehat{\theta}_n)\frac{2}{n}\sum_{j=1}^{n}U_j\varepsilon_j + (\theta-\widehat{\theta}_n)^2\frac{1}{n}\sum_{j=1}^{n}U_j^2 - (\overline{\varepsilon}_n)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j^2 + \mathcal{O}_{P_n}(|\widehat{\theta}_n - \theta|) + \mathcal{O}_{P_n}(|\widehat{\theta}_n-\theta|^2) + \mathcal{O}_{P_n}(n^{-1}) \\
&= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j^2 + o_{P_n}(1) \xrightarrow{P_n} \sigma_\varepsilon^2, \text{ because } \widehat{\theta}_n - \theta = \mathcal{O}_{P_n}(n^{-1/2}).
\end{aligned}$$

Furthermore, let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ be the random variables corresponding to the $\widetilde{\varepsilon}_j = V_j - \widehat{\theta}_n U_j$, $j = 1, 2, \ldots, n$, when $\widehat{\theta}_n$ is replaced by θ.

Define further pseudo random variables ε_j^+, $j = 1, 2, \ldots, n$, as $\varepsilon_j^+ = \varepsilon_{I_j^*}$ where $I_1^*, I_2^*, \ldots, I_n^*$ are the same i.i.d. random variables as those used to generate the ε_j^*'s.

Let $Y_j^* = U_j^* \varepsilon_j^*$ and $Y_j^+ = U_j^* \varepsilon_j^+, j = 1, \ldots, n$, and observe that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j^* = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j^+ + o_{P_n^*}(1).$$

This holds true because

$$n^{-1} \mathrm{E}^* \Big(\sum_{j=1}^{n} (Y_j^* - Y_j^+) \Big)^2 = \mathrm{E}^*(U_1^*)^2 \mathrm{E}^*(\varepsilon_1^* - \varepsilon_1^+)^2 = o_{P_n}(1).$$

The last convergence to zero in probability can be seen as follows

$$
\begin{aligned}
n^{-1}\mathrm{E}^*\Big(\sum_{j=1}^{n}(Y_j^* - Y_j^+)\Big)^2 &= \mathrm{E}^*(U_1^*)^2\mathrm{E}^*(\varepsilon_1^* - \varepsilon_1^+)^2 \\
&= \mathcal{O}_{P_n}(1)\frac{1}{n}\sum_{j=1}^{n}(\widehat{\varepsilon}_j - \varepsilon_j)^2 \\
&\leq \mathcal{O}_{P_n}(1)\Big(\frac{1}{n}\sum_{j=1}^{n}(\widetilde{\varepsilon}_j - \varepsilon_j)^2 + \overline{\varepsilon}_n^2\Big) \\
&= \mathcal{O}_{P_n}(1)\mathcal{O}_{P_n}(n^{-1}) \text{ , see above} \\
&= o_{P_n}(1).
\end{aligned}
$$

To verify (17), we show that $(n^{-1/2}\sum_{j=1}^{n} Y_{j,n}^+, n \in \mathbb{N})$ satisfies the assumptions of the CLT in P_n-probability.

For $\sum_{j=1}^{k(n)} \mathrm{Cov}^*\left(Z_{j,n}^*\right) \xrightarrow{P_n} \Sigma$, we have,

$$n^{-1}\sum_{j=1}^{n}\mathrm{Var}^*(Y_{j,n}^+) = \mathrm{E}^*(U_1^*)^2\mathrm{E}^*(\varepsilon_1^+)^2$$

$$= \Big(\frac{1}{n}\sum_{j=1}^{n} U_j^2\Big)\Big(\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j^2\Big) \xrightarrow{P_n} \tau_U^2\sigma_\varepsilon^2.$$

Condition (8) also is satisfied, since for every $\epsilon > 0$, we have

$$n^{-1}\sum_{j=1}^{n}\mathrm{E}^*\left[|Y_{j,n}^+|^2\mathbb{1}(|Y_{j,n}^+| > \epsilon\sqrt{n})\right] = \mathrm{E}^*\left[|U_1^*\varepsilon_1^+|^2\mathbb{1}(|U_1^*\varepsilon_1^+| > \epsilon\sqrt{n})\right]$$

$$= \frac{1}{n}\sum_{j=1}^{n}(U_j\varepsilon_j)^2\mathbb{1}(|U_j\varepsilon_j| > \epsilon\sqrt{n})\Big] =: L_n.$$

Now, $\mathrm{E}(L_n) = \mathrm{E}\big[(U_1\varepsilon_1)^2 \mathbb{1}(|U_1\varepsilon_1| > \epsilon\sqrt{n})\big] \to 0$ and $\mathrm{Var}(L_n) = \mathcal{O}(n^{-1}) \to 0$, as $n \to \infty$.

By (16) and (17) we finally conclude that

$$\mathcal{L}\big(S^*_{2,n} | \underline{(V,U)}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2_\varepsilon/\tau^2_U\big) \text{ in } P_n\text{-probability,}$$

which shows that Bootstrap Proposal II works.

Bootstrap Proposals I and II considered so far in this example, generate bootstrap samples (V_j^*, U_j^*), $j = 1, 2, \ldots, n$, from which bootstrap replicates $T_n^* = \widehat{\theta}_n^*$ of the statistic $T_n = \widehat{\theta}_n$ are obtained.

However, bootstrap replicates of the statistic T_n can be generated without generating bootstrap replicates (V_j^*, U_j^*) of the pairs (V_j, U_j) themselves. To elaborate, observe that expression (12) for the least squares estimator, also can be written as $T_n = \sum_{j=1}^n Y_j$, where (recall $V_j = \theta U_j + \varepsilon_j$)

$$\begin{aligned} Y_j &:= U_j V_j \Big/ \sum_{r=1}^n U_r^2 \\ &= \theta\, U_j^2 \Big/ \sum_{r=1}^n U_r^2 + U_j\, \varepsilon_j \Big/ \sum_{r=1}^n U_r^2, \quad j = 1, \ldots, n. \end{aligned} \tag{18}$$

That is, T_n is n times the sample mean of $Y_1, Y_2, \ldots, Y_n$.

This motivates the generation of a bootstrap sample $Y_1^*, Y_2^*, \ldots, Y_n^*$ of the Y_j's which allows the generation of a bootstrap analogue of T_n by simple averaging the bootstrap variables Y_j^* generated.

Due to the normalization by $\sum_{r=1}^{n} U_r^2$, the random variables Y_j are not independent which implies that drawing with replacement from the set $\{Y_1, Y_2, \ldots, Y_n\}$ is not a proper procedure to generate the bootstrap sample of interest.

One can, however, proceed as follows:

Observe that in expression (18) the random variables Y_j are affected by two independent sources of randomness.

The one is due to the U_j's and the other is due to the ε_j's.

Now, ignoring the randomness due to the U_j's we can generate bootstrap random variables Y_j^* by using bootstrap replicates of the ε_j's only.

To be more specific, let $\widehat{\sigma}_\varepsilon^2$ be a consistent estimator of σ_ε^2.

Generate i.i.d. random variables $W_1^*, W_2^*, \ldots, W_n^*$ which are independent of the U_j, have mean zero, variance $\widehat{\sigma}_\varepsilon^2$ and satisfy $\mathrm{E}^*|W_1^*|^{2+\delta} < \infty$ for some $\delta > 0$.

Replacing then θ in (18) by its estimator $\widehat{\theta}_n$, the following proposal leads to bootstrap replicates of the Y_j's,

$$Y_j^* = \widehat{\theta}_n \, U_j^2 \Big/ \sum_{r=1}^{n} U_r^2 + U_j \, W_j^* \Big/ \sum_{r=1}^{n} U_r^2, \quad j = 1, \ldots, n. \tag{19}$$

Notice that we do not further specify the distribution of the i.i.d. random variables $W_j^*, j = 1, \ldots, n$.
In fact W_j^* may have an arbitrary distribution provided it has mean zero, variance $\widehat{\sigma}_\varepsilon^2$ and $\mathrm{E}^*|W_1^*|^{2+\delta} < \infty$ is satisfied.

That is, for the W_j^*'s it suffices if they properly imitate the first and second order moments of the ε_j's. The bootstrap analogue of T_n is then given by $T_n^* = \sum_{i=1}^n Y_i^*$ and the distribution of

$$S_{3,n}^* = \sqrt{n}(T_n^* - \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j W_j^* / (\frac{1}{n} \sum_{r=1}^n U_r^2)$$

can be used to estimate the distribution of S_n. It is easily seen that $\mathrm{E}^*(S_{3,n}^*) = 0$ while

$$\mathrm{Var}^*(S_{3,n}^*) = \widehat{\sigma}_\varepsilon^2 / (\frac{1}{n} \sum_{r=1}^n U_r^2) \xrightarrow{P_n} \sigma_\varepsilon^2 / \tau_U^2.$$

Now, letting $u_{j,n} := U_j/(n^{-1}\sum_{r=1}^{n} U_r^2)$, one gets

$$\mathcal{L}\big(S_{3,n}^*\big|\underline{(V,U)}_n\big) = \mathcal{L}\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} u_{j,n}W_j^*\Big|\underline{(V,U)}_n\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\sigma_\varepsilon^2/\tau_U^2\big),$$

in P_n-probability.

The later convergence holds true since by the CLT in P_n-probability with Lyapunov's condition in P_n-probability and for any $\delta > 0$ such that $\mathrm{E}^*|W_1^*|^{2+\delta} < \infty$ and $\mathrm{E}|U_1|^{2+\delta} < \infty$, we have

$$\frac{1}{n^{1+\delta/2}}\sum_{j=1}^{n}\mathrm{E}^*|u_{j,n}W_j^*|^{2+\delta} = \mathrm{E}^*|W_1^*|^{2+\delta}\frac{1}{n^{1+\delta/2}}\sum_{j=1}^{n}|U_j|^{2+\delta}\Big/(n^{-1}\sum_{r=1}^{n}U_r^2)$$

$$= \mathcal{O}_{P_n}(n^{-\delta/2}) = o_{P_n}(1),$$

as $n \to \infty$.

The above derivations make it clear why despite the fact that the randomness of the U_j's in generating the Y_j^* in (19) has been ignored, bootstrap consistency for $S_{3,n}^*$ still holds true.

Furthermore, it suffices for the i.i.d. pseudo random variables W_j^*'s to only imitate the first and second order moments of the ε_j's to achieve bootstrap consistency.

As the discussion in the above example shows, all three bootstrap proposals $S^*_{1,n}$, $S^*_{2,n}$ and $S^*_{3,n}$, lead to a consistent estimation of the distribution of S_n, that is the corresponding distributions $\mathcal{L}(S^*_{j,n}|\underline{(V,U)}_n)$, $j = 1, 2, 3$, converges in P_n-probability and as $n \to \infty$, to the same Gaussian limit as the distribution $\mathcal{L}(S_n)$ does.

By the triangular inequality and the continuity of the Gaussian distribution function, this implies that, as $n \to \infty$,

$$d_K\Big(\mathcal{L}\big(S^*_{j,n}|\underline{(V,U)}_n\big), \mathcal{L}\big(S_n\big)\Big) \to 0 \quad \text{in } P_n\text{-probability},$$

for every $j \in \{1, 2, 3\}$.

However, while one may think of $S^*_{1,n}$ as generated using a "model free" bootstrapping scheme, $S^*_{2,n}$ is obtained by a bootstrap procedure which fully takes into account the model structure generating the observed sample (V_j, U_j), $j = 1, 2, \ldots, n,$.

At the same time, in the generation of $S_{3,n}^*$ the underlying model structure only partially has been used.

Furthermore, although the three bootstrap proposals (implicitly) imitate different aspects of the underlying statistical experiment and in particular of the probability measure P_n governing the generation of the sample (V_j, U_j), $j = 1, 2, \ldots, n$, they all succeed in properly mimicking those features of P_n that affect the (limiting) distribution of the statistic of interest T_n, more precisely, of its centered and normalized version S_n.

This statement holds true, since, as we have seen, all three bootstrap proposals are consistent. An interesting question of course will be whether any one of the three bootstrap proposals considered outperforms the others. Although we will not discuss this issue here, we provide in the next section some useful tools which allow to investigate properties of the bootstrap beyond consistency.

Bootstrap Failure

So far we have developed a number of tools for investigating properties of bootstrap proposals and for proving their consistency. However, there are situations in which a developed bootstrap proposal does not "work", i.e., it is not consistent, and in this section we demonstrate this by means of some examples.

Example 13

Consider an i.i.d. random sample $X_1, \ldots, X_n$ with $\mathrm{E}(X_j) = \mu$ and $0 < \mathrm{Var}(X_j) = \sigma^2 < \infty$. The distribution of interest is the distribution of a properly standardized version of the difference $D_n = g(\overline{X}_n) - g(\mu)$. The real valued function g is assumed to be three times continuously differentiable.

Suppose that we want to approximate the distribution of D_n by the distribution of $D_n^* = g(\overline{X}_n^*) - g(\overline{X}_n)$ with $\overline{X}_n^* = n^{-1} \sum_{j=1}^n X_j^*$ for an i.i.d. sample $X_1^*, X_2^*, \ldots, X_n^*$ generated from the empirical distribution function $\widehat{F}_n$, that is, $X_i^* \sim \widehat{F}_n$.

One might be tempted to assume that the aforementioned bootstrap procedure works for this situation, since we know consistency for the sample mean of i.i.d. sequences (cf. Theorem 12) and because of the δ-method. But, if $g'(\mu) = 0$ and $g''(\mu) \neq 0$, the proposed bootstrap will fail.

To elaborate, first we look at the limiting distribution of D_n. From a Taylor series expansion of $g(\overline{X}_n)$ around $g(\mu)$ we have because of $g'(\mu) = 0$, that

$$g(\overline{X}_n) - g(\mu) = \frac{1}{2}(\overline{X}_n - \mu)^2 g''(\mu) + \frac{1}{6}(\overline{X}_n - \mu)^3 g'''(U_n),$$

where U_n is a random value between $\overline{X}_n$ and μ. Continuity of g''' together with the fact that $\sqrt{n}(\overline{X}_n - \mu) = \mathcal{O}_{P_n}(1)$ implies

$$\frac{1}{6}(\overline{X}_n - \mu)^3 g'''(U_n) = \mathcal{O}_{P_n}(n^{-3/2}),$$

that is,

$$\begin{aligned} S_n &= n\big(g(\overline{X}_n) - g(\mu)\big) \\ &= \frac{1}{2} n\big(\overline{X}_n - \mu\big)^2 g''(\mu) + o_{P_n}(1) \\ &\xrightarrow{\mathcal{D}} \frac{1}{2} g''(\mu)\sigma^2 \chi_1^2. \end{aligned} \tag{20}$$

Recall that $\sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

This means that in this case it turns out that n instead of $\sqrt{n}$ is the appropriate standardization for $g(\overline{X}_n) - g(\mu)$.

Next, we consider the bootstrap approximation S_n^* of S_n. Notice that $g'(\overline{X}_n)$ is not necessarily equal to zero, and, therefore, the corresponding Taylor series expansion multiplied by n, yields

$$
\begin{aligned}
S_n^* &= n\big(g(\overline{X}_n^*) - g(\overline{X}_n)\big) \\
&= n(\overline{X}_n^* - \overline{X}_n)g'(\overline{X}_n) + \frac{1}{2}n\big(\overline{X}_n^* - \overline{X}_n\big)^2 g''(\overline{X}_n) \\
&\quad + \frac{1}{6}n\big(\overline{X}_n^* - \overline{X}_n\big)^3 g'''(\widetilde{U}_n),
\end{aligned}
$$

with $\widetilde{U}_n$ a random value between $\overline{X}_n^*$ and $\overline{X}_n$.
Now, as before, we have that

$$
\frac{1}{6}\big(\overline{X}_n^* - \overline{X}_n\big)^3 g'''(\widetilde{U}_n) = \mathcal{O}_{P_n^*}(n^{-3/2}) \text{ in } P_n\text{-probability},
$$

by continuity of g''' and because $\mathcal{L}\big(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)|\underline{X}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ in P_n-probability.

Thus $n(\overline{X}_n^* - \overline{X}_n)^3 g'''(\widetilde{U}_n) = \mathcal{O}_{P_n^*}(n^{-1/2})$ as $n \to \infty$.

Furthermore,

$$\mathcal{L}\Big(\frac{1}{2} n\, g''(\overline{X}_n)(\overline{X}_n^* - \overline{X}_n)^2 | \underline{X}_n\Big) \xrightarrow{\mathcal{D}} \frac{1}{2} g''(\mu)\sigma^2 \chi_1^2, \tag{21}$$

since $g''(\overline{X}_n) \xrightarrow{P_n} g''(\mu)$, $\widehat{\sigma}^2 \xrightarrow{P_n} \sigma^2$ and

$\mathcal{L}\big(\sqrt{n}(\overline{X}_n^* - \overline{X}_n)/\widehat{\sigma} | \underline{X}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ in P_n-probability.

However, $n(\overline{X}_n^* - \overline{X}_n)g'(\overline{X}_n) \nrightarrow 0$ in P_n-probability, since $\sqrt{n}(\overline{X}_n^* - \overline{X}_n)$ converges weakly towards a normal distribution in P_n-probability (see above) and because of

$$\begin{aligned}
\sqrt{n}\, g'(\overline{X}_n) &= \sqrt{n}\big(g'(\overline{X}_n) - g'(\mu)\big) \\
&= \sqrt{n}\big(\overline{X}_n - \mu\big) g''(\mu) + \frac{1}{2}\sqrt{n}\big(\overline{X}_n - \mu\big)^2 g'''(\overline{U}_n) \qquad (22) \\
&\xrightarrow{\mathcal{D}} \; g''(\mu) Z,
\end{aligned}$$

for a further random value $\overline{U}_n$ between $\overline{X}_n$ and μ and with $Z \sim \mathcal{N}(0, \sigma^2)$.

Thus, $n\big(g(\overline{X}_n) - g(\mu)\big)$ and $n\big(g(\overline{X}_n^*) - g(\overline{X}_n)\big)$ do not have the same limiting distribution and therefore the proposed bootstrap procedure fails.

The reason for this failure lies in the fact that $g'(\mu) = 0$ but $g'(\overline{X}_n) \neq 0$ for $\overline{X}_n \neq \mu$.

Even though $g'(\overline{X}_n) \xrightarrow{P_n} g'(\mu) = 0$, the convergence is not fast enough to ensure that $\sqrt{n}g'(\overline{X}_n)$ is asymptotically negligible.

Instead $\sqrt{n}g'(\overline{X}_n)$ converges towards a non-degenerated limit distribution (cf. (22)).

A simple way to get around the bootstrap failure in this approach would be to look directly at the statistics

$$\frac{1}{2} n\, g''(\overline{X}_n)(\overline{X}_n^* - \overline{X}_n)^2$$

as a bootstrap approximation of the distribution of $n\big(g(\overline{X}_n) - g(\mu)\big)$.

From (21) it immediately follows that this bootstrap proposal works.

Of course this proposal is not really applicable, since in practice we don't know μ and therefore we generally cannot decide whether or not $g'(\mu) = 0$ (unless $g'(x) \neq 0$ for all x).

Nevertheless, this example shows that the phrase "the devil is in the details" also applies for the bootstrap.

A more promising alternative bootstrap procedure to resolve the above bootstrap failure is to resample fewer than n observations, namely to generate a i.i.d. sample $X_1^*, \ldots, X_m^*$ with $m \ll n$ and $X_i^* \sim \widehat{F}_n$.

If $m = m(n)$ fulfills both $m(n) \to \infty$ and $m(n)/n \to 0$ as $n \to \infty$, then one can show that the distribution of $m\big(g(\overline{X}_m^*) - g(\overline{X}_n)\big)$ converges weakly towards the same scaled χ^2 limiting distribution as $n\big(g(\overline{X}_n) - g(\mu)\big)$ does.

Moreover, the described bootstrap procedure, which is called the m-out-of-n bootstrap, also works in case $g(\mu) \neq 0$. In this case the distributions of $\sqrt{n}\big(g(\overline{X}_n) - g(\mu)\big)$ and of $\sqrt{m}\big(g(\overline{X}_m^*) - g(\overline{X}_n)\big)$ both weakly converge to the same limiting normal distribution $\mathcal{N}\big(0, \sigma^2(g'(\mu))^2\big)$.

Details are left as an exercise.

The m-out-of-n bootstrap appears to be less sensitive to sudden changes (discontinuities) of the asymptotic distribution of a statistic depending on the underlying parameter.

A close inspection of the reasoning in this example shows that it is sufficient to require that the differentiability assumptions for g only hold true in a neighborhood of μ.

Let us consider two further examples.

Example 14

In this example we suppose that $X_1, X_2, \ldots, X_n$ is an i.i.d. random sample from the uniform distribution $U[0, \vartheta]$, $\vartheta > 0$, over the interval $[0, \vartheta]$. Consider the following random sequences of interest:

$$\text{Case I:} \quad S_n^{(1)} = \sqrt{n}\big(2\overline{X}_n - \vartheta\big),$$

$$\text{Case II:} \quad S_n^{(2)} = n\big(\max\{X_1, X_2, \ldots, X_n\} - \vartheta\big).$$

Note that $2\,\overline{X}_n$ is a consistent moment estimator of ϑ, while the maximum likelihood estimator for ϑ is given by $\max\{X_1, X_2, \ldots, X_n\}$. In Case I validity of the bootstrap procedure based on resampling from the empirical distribution function $\widehat{F}_n$ of $X_1, X_2, \ldots, X_n$ immediately follows from Theorem 12. In fact, the distributions of $S_n^{(1)}$ and $\sqrt{n}\big(2\overline{X}_n^* - 2\overline{X}_n\big)$ both converge in distribution towards a centered normal distribution with variance $4\,\mathrm{Var}(X_1) = \vartheta^2/3$.

However, for Case II the same bootstrap proposal fails. To see this first note that for $x \in \mathbb{R}$

$$P\big\{n\,(\max\{X_1,X_2,\ldots,X_n\}-\vartheta) \le x\big\} \\ = \big(P\big\{X_1 \le \vartheta + x/n\big\}\big)^n \to \mathrm{e}^{x/\vartheta}\mathbb{1}_{\{x\le 0\}} + \mathbb{1}_{\{x>0\}},$$

as $n \to \infty$, where the limiting cumulative distribution function is everywhere continuous.

In contrast, we have for a bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$, stemming from the empirical distribution function $\widehat{F}_n$, and if the underlying observations $X_1, \ldots, X_n$ are pairwise different (which happens with probability one), on the one hand that,

$$P_n^*\big(n(\max\{X_1^*, X_2^*, \ldots, X_n^*\} - \max\{X_1, X_2, \ldots, X_n\}) \le 0\big) = 1$$

and on the other hand that

$$
\begin{aligned}
&P_n^*\big(n(\max\{X_1^*, X_2^*, \ldots, X_n^*\} - \max\{X_1, X_2, \ldots, X_n\}) \geq 0\big) \\
&\quad = P_n^*\big(\max\{X_1^*, X_2^*, \ldots, X_n^*\} \geq \max\{X_1, X_2, \ldots, X_n\}\big) \\
&\quad = 1 - P_n^*\big(X_j^* < \max\{X_1, X_2, \ldots, X_n\} \text{ for all } j = 1, \ldots, n\big) \\
&\quad = 1 - \Big(P_n^*\big(X_1^* < \max\{X_1, X_2, \ldots, X_n\}\big)\Big)^n \\
&\quad = 1 - \left(1 - \frac{1}{n}\right)^n, \text{ since the } X_j \text{ are all different} \\
&\quad \to 1 - \mathrm{e}^{-1} \; > \; 0, \text{ as } n \to \infty.
\end{aligned}
$$

This means that in the limit, the distribution of

$$n(\max\{X_1^*, X_2^*, \ldots, X_n^*\} - \max\{X_1, X_2, \ldots, X_n\})$$

has a discontinuity at $x = 0$.

The fact that we know that $X_j \sim U[0, \vartheta]$ rather motivates a parametric bootstrap proposal, where resampling is not done from the empirical distribution, as above, but from a uniform distribution $U[0, \widehat{\vartheta}_n]$, where $\widehat{\vartheta}_n$ denotes an arbitrary but consistent estimator for ϑ.

The details are left as an exercise.

Example 15

In this last example we assume that $X_1, X_2, \ldots, X_n$ is an i.i.d. random sample from a distribution with $\mathrm{E}(X_j) = \mu$ and $0 < \mathrm{Var}(X_j) = \sigma^2 < \infty$. Suppose that we are interested in the distribution of the statistic $S_n = \sqrt{n}\left(|\overline{X}_n| - |\mu|\right)$. This is exactly the same situation as in Example 13 with $g(x) = |x|$.

We again consider the bootstrap proposal $S_n^* = \sqrt{n}\big(|\overline{X}_n^*| - |\overline{X}_n|\big)$, where $\overline{X}_n^* = n^{-1}\sum_{j=1}^n X_j^*$ and $X_1^*, \ldots, X_n^*$ are generated according to the empirical distribution function $\widehat{F}_n$ of the underlying observations $X_1, \ldots, X_n$.

From Theorem 12 we know that $\mathcal{L}\big(\sqrt{n}\big(\overline{X}_n^* - \overline{X}_n\big)|\underline{X}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ in P_n-probability applies.

Along the lines of Example 13 we immediately obtain that if $\mu \neq 0$ then this bootstrap works.

For the specific case $\mu = 0$ we observe an anomaly in the asymptotic distribution of S_n, namely that for $x \geq 0$ it holds

$$\begin{aligned} P(S_n \leq x) &= P\big(\sqrt{n}\,|\overline{X}_n|/\sigma \leq x/\sigma\big) \\ &= P\big(-x/\sigma \leq \sqrt{n}\,\overline{X}_n/\sigma \leq x/\sigma\big) \\ &\to \big(\Phi(x/\sigma) - \Phi(-x/\sigma)\big)\mathbb{1}_{[0,+\infty)}(x) \text{ as } n \to \infty. \end{aligned}$$

That is, it holds for $\mu = 0$ that $S_n = \sqrt{n}\,|\overline{X}_n| \xrightarrow{\mathcal{D}} |Z|$ for $Z \sim \mathcal{N}(0, \sigma^2)$.
Recall that if $\mu \neq 0$ then S_n is asymptotically normal.
Now we investigate the distribution of S_n^* in case of $\mu = 0$. Skorokhod's representation theorem (cf. Billingsley (1995), Theorem 25.6) implies that because of $\sqrt{n}\,\overline{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ we can find a further common probability space and i.i.d. random variables $X_1', \ldots, X_n'$ with $\sqrt{n}\,\overline{X}_n' \sim \sqrt{n}\,\overline{X}_n$, for all $n \in \mathbb{N}$, and $Z' \sim \mathcal{N}(0, \sigma^2)$, such that $\sqrt{n}\,\overline{X}_n' \to Z'$ almost surely.

Assume that a further bootstrap sample $X_1'^*, \ldots, X_n'^*$ is generated from the empirical distribution function of $X_1', \ldots, X_n'$.

Since $\mathcal{L}\left(\sqrt{n}(\overline{X}_n'^* - \overline{X}_n') \middle| \underline{X}_n'\right) \xrightarrow{\mathcal{D}} Z''$ in P_n-probability for a further random variable $Z'' \sim \mathcal{N}(0, \sigma^2)$, where Z'' by construction is independent of Z', and since $\sqrt{n}\,\overline{X}_n' \to Z'$ almost surely, we obtain, as $n \to \infty$,

$$\begin{aligned}
&\mathcal{L}\left(\sqrt{n}\,(|\overline{X}_n'^*| - |\overline{X}_n'|) \middle| \underline{X}_n'\right) \\
&= \mathcal{L}\left(\left|\sqrt{n}\,(\overline{X}_n'^* - \overline{X}_n') + \sqrt{n}\,\overline{X}_n'\right| - \left|\sqrt{n}\,\overline{X}_n'\right| \middle| \underline{X}_n'\right) \\
&\xrightarrow{\mathcal{D}} |Z'' + Z'| - |Z'| \text{ in } P_n\text{-probability},
\end{aligned}$$

for Z' and Z'' independent and identically distributed Gaussian random variables with variance σ^2.

Therefore the limiting distribution of $\sqrt{n}\big(|\overline{X}_n'^*| - |\overline{X}_n'|\big)$ is different from that of $\sqrt{n}|\overline{X}_n|$.

This means that the proposed bootstrap does not work for the case $\mu = 0$.

If, unrealistically, $\mu = 0$ can be assumed to be known, then for the modified bootstrap proposal $S_n'^* = \sqrt{n}\,|\overline{X}_n^* - \overline{X}_n|$ using Theorem 12 and the continuous mapping theorem, it can be shown that this leads to a valid bootstrap approximation of the distribution of $S_n' = \sqrt{n}\,|\overline{X}_n|$.

Bootstrap for Time Series

Consider the situation where the sample $X_1, X_2, \ldots, X_n$ follows the simple first order moving average model,

$$X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \tag{23}$$

$\theta \in \mathbb{R}$ and ε_j, $j = 0, 1, \ldots, n$ are i.i.d. with $\mathrm{E}(\varepsilon_j) = 0$ and $0 < \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_1) < \infty$.

We assume that the special model structure (23) is not known and we consider again the problem of estimating the distribution of the sample mean $\overline{X}_n = n^{-1}\sum_{j=1}^{n} X_j$.

Notice that if $\theta = 0$ we are in the i.i.d. setup.

The asymptotic distribution of $\sqrt{n}(\overline{X}_n - \mu)$ is easily derived since

$$\sqrt{n}(\overline{X}_n - \mu) = (1+\theta)\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\varepsilon_j + \frac{\theta}{\sqrt{n}}\varepsilon_0 + \frac{1}{\sqrt{n}}\varepsilon_n, \tag{24}$$

The central limit theorem for sums of i.i.d. random variables gives

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1+\theta)^2\,\sigma_\varepsilon^2\right),$$

Notice that for $h \in \mathbb{Z}$,

$$\mathrm{Cov}(X_t, X_{t+h}) = \sigma_\varepsilon^2 \mathbb{1}_{\{-1,0,+1\}}(h)\big\{(1-|h|)(1+\theta^2) \; + \; |h|\theta\big\},$$

and therefore,

$$\begin{aligned} n\mathrm{Var}(\overline{X}_n) &= \sum_{h=-1}^{1} (1-|h|/n)\mathrm{Cov}(X_1, X_{1+h}) \\ &= \big(2\big(1-1/n\big)\theta + \big(1+\theta^2\big)\big)\sigma_\varepsilon^2. \end{aligned}$$

Hence the limit of $n\mathrm{Var}(\overline{X}_n)$, which also is the variance of the limiting Gaussian distribution of $\sqrt{n}(\overline{X}_n - \mu)$, equals

$$\mathrm{Var}(X_1) + 2\,\mathrm{Cov}(X_t, X_{t+1}) = (1+\theta)^2\,\sigma_\varepsilon^2. \tag{25}$$

A bootstrap procedure based on i.i.d. resampling of the X_j's (drawing with replacement) fails in this context since it does not take into account the dependence structure of the sample $X_1, X_2, \ldots, X_n$. As (25) shows, this dependence indeed affects the distribution of $\overline{X}_n$.

To elaborate, let $\overline{X}_n^* = n^{-1} \sum_{j=1}^n X_j^*$, where $X_1^*, X_2^*, \ldots, X_n^*$ is an i.i.d. sample with $X_j^* \sim \widehat{F}_n$ and $\widehat{F}_n(x) = n^{-1} \sum_{j=1}^n \mathbb{1}_{(-\infty,x]}(X_j)$, the empirical distribution function. Then, as in Example 5, we get, as $n \to \infty$,

$$\begin{aligned} \mathrm{Var}^*(\sqrt{n}\,\overline{X}_n^*) = \mathrm{Var}^*(X_1^*) &= \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2 \\ &\xrightarrow{P_n} \mathrm{Var}(X_1) = (1+\theta^2)\sigma_\varepsilon^2 \neq (1+\theta)^2\sigma_\varepsilon^2 \quad \text{for } \theta \neq 0 \end{aligned}$$

Clearly, the limit of $\mathrm{Var}(\sqrt{n}\,\overline{X}_n^*)$ differs from that of $\mathrm{Var}(\sqrt{n}\,\overline{X}_n)$ by the term $2\,\mathrm{Cov}(X_t, X_{t-1})$, which is the autocovariance of (X_t) not mimicked by the i.i.d. pseudo random sample $X_1^*, \ldots, X_n^*$.

One might be tempted to think that, since X_t and X_s are independent for $|t-s|>1$, i.e, $X_1, X_2, \ldots, X_n$ is a m-dependent time series with $m=1$, a bootstrap procedure based on i.i.d. resampling from the pairs (X_t, X_{t+1}), $t=1,2,\ldots,n-1$, will suffice to appropriately mimic the dependence structure of $X_1, X_2, \ldots, X_n$.

To be more specific, assume for simplicity that $n=2k$, $k \in \mathbb{N}$, and denote by $X_1^+, X_2^+, \ldots, X_{n-1}^+, X_n^+$ a bootstrap sample obtained by choosing with replacement k pairs (X_t^+, X_{t+1}^+) from the set of all possible pairs $\{(X_t, X_{t+1}),\ t=1,2,\ldots,n-1\}$ and joining them together in the order selected.

Let $\overline{X}_n^+ = n^{-1}\sum_{t=1}^n X_t^+$ and use the distribution of $S_n^+ = \sqrt{n}(\overline{X}_n^+ - \mathrm{E}^+(\overline{X}_n^+))$ to approximate the distribution of $\sqrt{n}(\overline{X}_n - \mu)$. Now, due to the independence of the k pairs, we get

$$\mathrm{Var}^+(S_n^+) = \mathrm{Var}^+\Big(\frac{1}{2k}\sum_{s=1}^{k}(X_{2s-1}^+ + X_{2s}^+)\Big)$$

$$= \frac{1}{2k}\sum_{s=1}^{k}\mathrm{Var}^+(X_{2s-1}^+ + X_{2s}^+) = \frac{1}{2}\mathrm{Var}^+(X_1^+ + X_2^+) \quad (26)$$

and since $\mathrm{Var}^+(X_1^+ + X_2^+) \xrightarrow{P_n} 2\mathrm{Var}(X_0) + 2\mathrm{Cov}(X_0, X_1)$, as $n \to \infty$, left as exercise, we get that

$$\mathrm{Var}^+(S_n^+) \xrightarrow{P_n} (1+\theta+\theta^2)\sigma_\varepsilon^2 \neq (1+\theta)^2\sigma_\varepsilon^2.$$

That is, this bootstrap procedure also fails, and as equation (26) shows, the reason for this failure is the fact that this procedure uses the variance of the sum of two random variables to approximate the variance of the sum of n random variables.

As this simple example demonstrates, allowing for dependence in the sample $X_1, X_2, \ldots, X_n$ puts the bootstrap in front of new challenges and the question arises how an appropriate bootstrap procedure can be designed in such a situation.

As we will see, one option will be to resample not pairs but blocks of consecutive observations, where, in order to overcome the limitations of bootstrapping pairs, the length of the blocks resampled should be allowed to increase to infinity with the sample size.

This is the main idea behind the different block bootstrap methods proposed in the literature.

But if one assumes knowledge of the parametric structure generating the observations, one can take advantage of this knowledge and design a bootstrap procedure which generates bootstrap samples by using an estimated version of the underlying parametric structure.

Such parametric approaches will be discussed next.

Assume that the time series $X_1, X_2, \ldots, X_n$ at hand stems from a stochastic process $\{X_t, t \in \mathbb{Z}\}$, where

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{27}$$

$\phi_j \in \mathbb{R}$, $j = 1, 2, \ldots, p$, and the innovation process $\{\varepsilon_t, t \in \mathbb{Z}\}$ consists of i.i.d. random variables with cumulative distribution function F_ε having mean zero and variance $0 < \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_0) < \infty$.

It is known (cf. Brockwell and Davis (1991), Chapter 3), that the process (27) is stationary if the characteristic polynomial

$$\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j, \ z \in \mathbb{C},$$

does not vanish on the unit circle, that is if $\phi(z) \neq 0$ for all $|z| = 1$.

Furthermore, if $\phi(z) \neq 0$ for $|z| \leq 1$, then the process (X_t) is also causal, that is, X_t can be expressed as a function of the past innovations only, more precisely,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{28}$$

where $\psi_0 = 1$ and the coefficients ψ_j, $j = 1, 2, \ldots$, are those appearing in the power series $\phi^{-1}(z) = 1 + \sum_{j=1}^{\infty} \psi_j z^j$, $|z| \leq 1$.

In the following we restrict our discussion to the case where the autoregressive process $\{X_t, t \in \mathbb{Z}\}$ is stationary and causal.

We further assume that the model structure (27) is known up to the parameter vector $\phi = (\phi_1, \phi_2, \ldots, \phi_p)^\top$ and to the distribution F_ε of the innovations (including the variance σ_ε^2).

Our aim is to develop a bootstrap proposal that generates pseudo-random variables $X_1^*, X_2^*, \ldots, X_n^*$, which will be plugged in some statistic $T_n = T(X_1, X_2, \ldots, X_n)$ of interest in order to successfully approximate its distribution.

Typically, we are again interested in a centered and rescaled version of T_n denoted by $S_n = r_n(T_n - \theta)$, where r_n is a sequence of real numbers that increases to infinity as n increases to infinity.

Under the assumption that the model structure (27) is known, it is natural to design a bootstrap proposal that takes advantage of this knowledge when generating the bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$.

This can be achieved by using the following model-based bootstrap algorithm.

Step 1: Let $\widetilde{\phi}_n = (\widetilde{\phi}_1, \widetilde{\phi}_2, \ldots, \widetilde{\phi}_p)^\top$ be an estimator of $\phi = (\phi_1, \ldots, \phi_p)^\top$ and denote by

$$\widetilde{\varepsilon}_t = X_t - \sum_{j=1}^{p} \widetilde{\phi}_j X_{t-j}, \qquad t = p+1, p+2, \ldots, n,$$

the corresponding set of residuals. Let $\widehat{\varepsilon}_t = \widetilde{\varepsilon}_t - \overline{\varepsilon}_n$, $t = p+1$, $p+2, \ldots, n$, the centered residuals, where $\overline{\varepsilon}_n = (n-p)^{-1} \sum_{t=p+1}^{n} \widetilde{\varepsilon}_t$.

Step 2: Generate a sample $\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*$ of i.i.d. random variables from the empirical distribution function $\widehat{F}_{n,\varepsilon}$ of the centered residuals $\widehat{\varepsilon}_t$, that is, the ε_t^*'s are chosen with replacement from the set $\{\widehat{\varepsilon}_t,\ t = p+1, p+2, \ldots, n\}$.

Step 3: For $t = 1, 2, \ldots, n$, define $X_t^* = \sum_{j=1}^{p} \widetilde{\phi}_j X_{t-j}^* + \varepsilon_t^*$.

Step 4: Let $T_n^* = T(X_1^*, \ldots, X_n^*)$ and approximate the distribution of S_n by that of $S_n^* = r_n(T_n^* - \widetilde{\theta})$.

To prove consistency we make use of the following auxiliary results.

Lemma 16

Suppose that the sequence $\{\widetilde{\phi}_n, n \in \mathbb{N}\}$ *of autoregressive parameter estimators used in Step 1 of the Model-Based AR Bootstrap satisfies* $\sqrt{n}\|\widetilde{\phi}_n - \phi\| = \mathcal{O}_{P_n}(1)$ *and that* $\mathrm{E}(|\varepsilon_1|^{2r}) < \infty$ *for some* $r \in \mathbb{N}$. *Then, as* $n \to \infty$,

$$d_{2r}(\widehat{F}_{n,\varepsilon}, F_\varepsilon) \xrightarrow{P_n} 0.$$

Lemma 17

Suppose that the sequence $\{\widetilde{\phi}_n, n \in \mathbb{N}\}$ *of autoregressive parameter estimators used in Step 1 of the Model-Based AR Bootstrap satisfies* $\sqrt{n}\|\widetilde{\phi}_n - \phi\| = \mathcal{O}_{P_n}(1)$ *and that* $\widetilde{\phi}(z) = 1 - \sum_{j=1}^{p} \widetilde{\phi}_j z^j \neq 0$ *for* $|z| \leq 1$. *Then, there exists* $\eta > 0$ *such that*

$$\sqrt{n} \sup_{j \in \mathbb{N}} \left((1+\eta)^j |\widetilde{\psi}_j - \psi_j|\right) = \mathcal{O}_{P_n}(1).$$

We then can prove the following result for the sample mean.

Theorem 18

Assume that $\{X_t, t \in \mathbb{Z}\}$ *is a stationary and causal AR(p) process with i.i.d. innovations, cf.* (27), *which allows for an MA*(∞) *representation* (28). *Let us also assume that* $X_1^*, X_2^*, \ldots, X_n^*$ *are generated as in Model-Based AR Bootstrap and that the sequence of estimators* $\widetilde{\phi}_n$ *satisfies the conditions of Lemma 17. Then*

$$d_2(\mathcal{L}(\sqrt{n}\,\overline{X}_n), \mathcal{L}(\sqrt{n}\,\overline{X}_n^* | \underline{X}_n)) \to 0 \quad \text{in } P_n\text{-probability},$$

where $\sqrt{n}\,\overline{X}_n^* = n^{-1/2} \sum_{t=1}^n X_t^*$.

Notice that no centering is needed here since by assumption $\mathrm{E}(X_t) = \mathrm{E}^*(X_t^*) = 0$.

Idea of proof: Let $\widetilde{\psi}_j$, $j = 1, 2, \ldots$ be the coefficients in the power series expansion $\widetilde{\phi}^{-1}(z) = \sum_{j=0}^{\infty} \widetilde{\psi}_j z^j$, $|z| \leq 1$, $\widetilde{\psi}_0 = 1$. Making use of the causal representation of X_t and of X_t^*, respectively, we obtain that

$$S_n = \sqrt{n}\overline{X}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \text{ and } S_n^* = \sqrt{n}\overline{X}_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j=0}^{\infty} \widetilde{\psi}_j \varepsilon_{t-j}^* .$$

Let (u_t, v_t), $t \in \mathbb{Z}$, be i.i.d. pairs of random variables on a new probability space (depending on $\underline{X}_n$), where u_t has marginal distribution function $\widehat{F}_{n,\varepsilon}$ and , v_t has marginal distribution F_ε, such that $d_2(u_1, v_1) = \sqrt{\mathrm{E}(u_1 - v_1)^2}$; $d_2(u_1, v_1) = d_2(\widehat{F}_{n,\varepsilon}, F_\varepsilon)$.

Then,

$$d_2(S_n, S_n^*) \leq \left(n^{-1} \mathrm{E}\left(\left| \sum_{t=1}^{n} \sum_{j=0}^{\infty} (\widetilde{\psi}_j u_{t-j} - \psi_j v_{t-j}) \right|^2 \right) \right)^{1/2}$$

$$\leq \left(n^{-1} \mathrm{E}\left(\left| \sum_{t=1}^{n} \sum_{j=0}^{\infty} \widetilde{\psi}_j (u_{t-j} - v_{t-j}) \right|^2 \right) \right)^{1/2} + \left(n^{-1} \mathrm{E}\left(\left| \sum_{t=1}^{n} \sum_{j=0}^{\infty} (\widetilde{\psi}_j - \psi_j) v_{t-j} \right|^2 \right) \right)^{1/2} .$$

The Model-Based AR Bootstrap also works for so-called *generalized means*, cf. Künsch (1989):

$$T_n = f\Big(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g\big(X_t, X_{t+1}, \ldots, X_{t+m-1}\big)\Big). \tag{29}$$

Theorem 19

Suppose that $X_1^*, X_2^*, \ldots, X_n^*$ is generated as in Model-Based AR Bootstrap that the underlying process $\{X_t, t \in \mathbb{Z}\}$ fulfills (27) and that the sequence of estimators $\widetilde{\phi}_n$ satisfies the conditions of Lemma 17. Suppose some assumptions are fulfilled for some $h \geq 1$ and that $\mathrm{E}(\varepsilon_1^{2(h+2)}) < \infty$, that Lemma 16 with $r = h+2$ holds true.

Then, as $n \to \infty$,

$$d_K\big(\mathcal{L}(\sqrt{n}(T_n - f(\theta))), \mathcal{L}(\sqrt{n}(T_n^* - f(\theta^*))|\underline{X}_n)\big) \to 0 \quad \text{in } P_n\text{-probability.}$$

We generate time series of length n from the first order autoregressive process $\{X_t, t \in \mathbb{Z}\}$, where

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \phi = 0.8, \tag{30}$$

and the ε_t's are i.i.d. Gaussian random variables with mean zero and variance $\sigma_\varepsilon^2 = 1$.
We are interested in comparing the finite sample performance of the asymptotic and the bootstrap estimators of the distribution of the Yule-Walker estimator $\widetilde{\phi}_n^{YW}$, i.e., of the distribution of $S_n = \sqrt{n}(\widetilde{\phi}_n^{YW} - \phi)$.

In the following we write for simplicity of notation $\widehat{\phi}$ for $\widetilde{\phi}_n^{YW}$.
Notice that for the particular example considered $\phi = \rho(1)$ and $\widehat{\phi} = \widehat{\rho}(1)$.

It holds true that, as $n \to \infty$, $S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$, where $\tau^2 = 1 - \phi^2$; see Brockwell and Davis (1991), Theorem 8.1.1.

Thus, the asymptotic estimator of the distribution of S_n is Gaussian with mean zero and estimated variance $\widehat{\tau}^2 = 1 - \widehat{\rho}(1)^2$.

The model-based AR Bootstrap suggests to use $S_n^* = \sqrt{n}(\widehat{\phi}^* - \widehat{\phi})$ as an estimator of S_n.

To judge the finite sample performance of the bootstrap as well as of the asymptotic normal approximation, we generate $R = 100$ replications of model (30) where for each replication $r = 1, 2, \ldots, 100$, we estimated the asymptotic Gaussian distribution, denoted by $L_{n,Asy}^{(r)}$ and the corresponding model-based bootstrap distribution denoted by $L_{n,Boot}^{(r)}$.

The latter distribution has been estimated using $B = 1,000$ bootstrap replicates.

Obviously, both approximations depend on the estimated parameter $\widehat{\phi}^{(r)}$ of ϕ in the r-th replication.

Finally, $1,000$ replications of model (30) have been generated in order to estimate the exact distribution of S_n.

To compare the quality of the asymptotic and of the bootstrap approximation, we calculated the empirical d_1-distance between the distributions of $L_{n,Asy}^{(r)}$ and of $L_{n,Boot}^{(r)}$, respectively, and the (estimated) exact distribution L_n of S_n.

Recall that d_1-Mallows distance between two distributions with distribution function F and G, respectively, is defined as

$$d_1(F,G) = \int_0^1 \left|F^{-1}(t) - G^{-1}(t)\right| dt = \int_{-\infty}^{\infty} \left|F(x) - G(x)\right| dx.$$

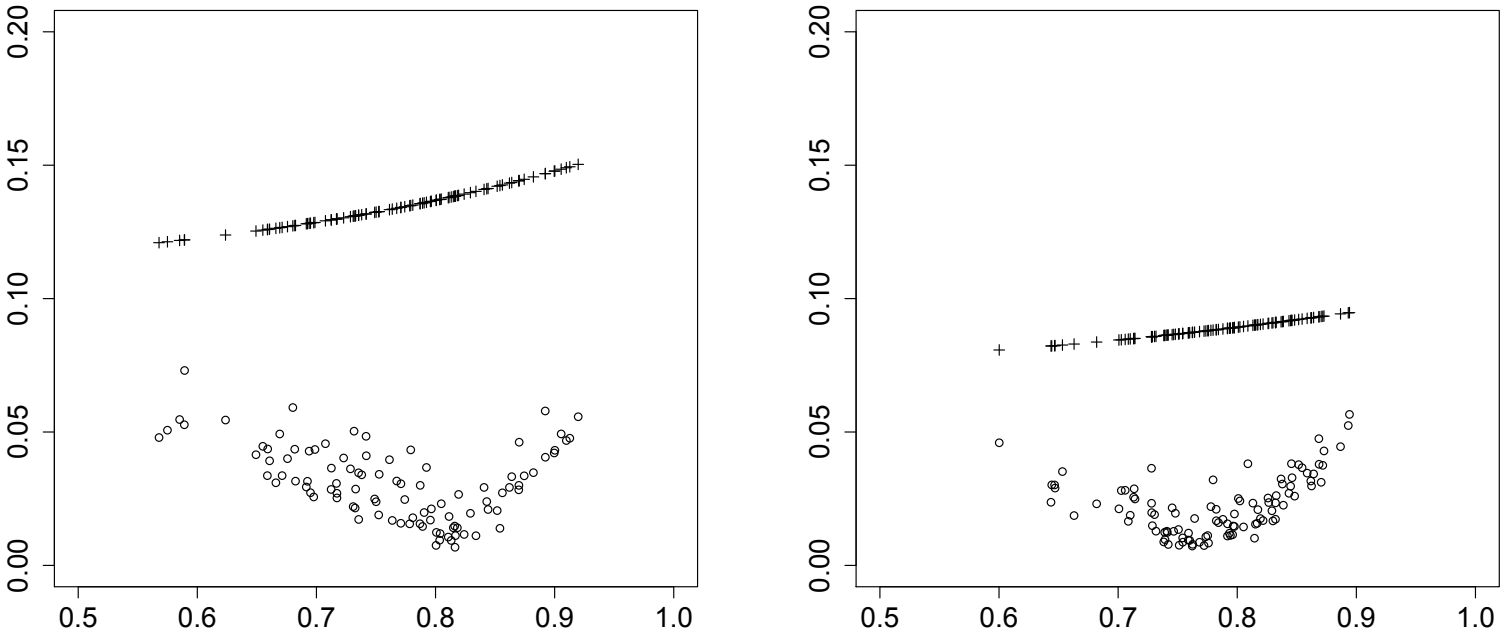
The empirical d_1-distance between two distributions $L_{1,n}$ and $L_{2,n}$ with empirical distribution functions $\widehat{F}_n$ and $\widehat{G}_n$, respectively, is defined as

$$\widehat{d}_1(L_{1,n}, L_{2,n}) = \frac{1}{K} \sum_{j=1}^{K} \left| \widehat{F}_n(x_j) - \widehat{G}_n(x_j) \right|, \tag{31}$$

where $x_1 < x_2 < \ldots < x_K$, is a set of K values at which both distribution functions are evaluated.

The results obtained for $n = 50$ and $n = 100$ are presented in Figure 3, where the empirical d_1-distances of the asymptotic and of the bootstrap approximations are plotted against the corresponding estimate $\widehat{\phi}^{(r)}$ obtained in each one of the $r = 1, 2, \ldots, R = 100$ replications generated.

A Simulation



Empirical d_1-distances (vertical axis) against the estimated parameter $\widehat{\phi}^{(r)}$ (horizontal axis) for $r = 1, 2, \ldots, R = 100$ replications of the process (30). The crosses (+) refer to the d_1-distances of the asymptotic Gaussian and the circles (o) to the d_1-distances of the bootstrap distribution to the estimated exact distribution of $\sqrt{n}(\widehat{\phi} - \phi)$, for $n = 50$ (left panel) and for $n = 100$ (right panel).

Step 1: Consider the column vectors $X_{t,p} = (X_t, X_{t-1}, \ldots, X_{t-p+1})^\top$ and the row vectors $Y_{t+1} := (X_{t+1}, X_{t,p}^\top)$ for $t = p, p+1, \ldots, n-1$. Generate with the help of independent random variables $T_j^*, j = p+1, \ldots, n-1$, vectors $Y_{j+1}^* := (X_{T_j^*+1}, X_{T_j^*,p}^\top)$, $j = p, p+1, \ldots, n-1$, of random variables with $P_n^*(Y_{j+1}^* = Y_{s+1}) = P_n^*(T_j^* = s) = 1/(n-p)$ for $s = p, p+1, \ldots, n-1$.

Step 2: Define

$$\widetilde{\phi}_n^* = \Big(\frac{1}{n-p}\sum_{j=p}^{n-1} X_{T_j^*,p} X_{T_j^*,p}^\top\Big)^{-1}\Big(\frac{1}{n-p}\sum_{j=p}^{n-1} X_{T_j^*,p} X_{T_j^*+1}\Big).$$

Step 3: Let $\widetilde{\phi}_n = (\sum_{j=p}^{n-1} X_{j,p}X_{j,p}^\top)^{-1}(\sum_{j=p}^{n-1} X_{j,p}X_{j+1})$ be the least-squares estimator of ϕ. Approximate the distribution of $\sqrt{n}(\widetilde{\phi}_n - \phi)$ through the distribution of the bootstrap analogue $\sqrt{n}(\widetilde{\phi}_n^* - \widetilde{\phi}_n)$.

Step 1: Let $\widetilde{\varepsilon}_{t+1} = X_{t+1} - \widetilde{\phi}_n^\top X_{t,p}$, $t = p, p+1, \ldots, n-1$, and define the centered residuals $\widehat{\varepsilon}_t = \widetilde{\varepsilon}_t - (n-p)^{-1} \sum_{s=p+1}^{n} \widetilde{\varepsilon}_s$. Denote by $\widehat{F}_{n,\varepsilon}$ the empirical distribution function of the $\widehat{\varepsilon}_t$, $t = p+1, p+2, \ldots, n$.

Step 2: Generate pseudo random variables X_{t+1}^*, $t = p, p+1, \ldots, n-1$, as

$$X_{t+1}^* = \widetilde{\phi}_n^\top X_{t,p} + \varepsilon_{t+1}^*,$$

where ε_t^*, $t = p+1, p+2, \ldots, n$, is a sample of independent, identically distributed random variables stemming from $\widehat{F}_{n,\varepsilon}$.

Step 3: Let $\widetilde{\phi}_n^*$ be the least squares estimator obtained by minimizing $\sum_{j=p}^{n-1} (X_{j+1}^* - c^\top X_{j,p})^2$ with respect to $c \in \mathbb{R}^p$.

Step 4: Approximate the distribution of $\sqrt{n}(\widetilde{\phi}_n - \phi)$ by the distribution of the bootstrap analogue $\sqrt{n}(\widetilde{\phi}_n^* - \widetilde{\phi}_n)$.

Step 1: Let $\widetilde{\varepsilon}_{t+1} = X_{t+1} - \widetilde{\phi}_n^\top X_{t,p}$, $t = p, p+1, \ldots, n-1$.

Step 2: Generate pseudo random variables X_{t+1}^+, $j = p, p+1, \ldots, n-1$, as

$$X_{j+1}^+ = \widetilde{\phi}_n^\top X_{j,p} + \varepsilon_{j+1}^+,$$

where $\varepsilon_j^+ = \widetilde{\varepsilon}_j \cdot \eta_j^*$ for $j = p+1, p+2, \ldots, n$ and the η_j^*'s are i.i.d. random variables satisfying $\mathrm{E}^*(\eta_j^*) = 0$, $\mathrm{E}^*((\eta_j^*)^2) = 1$ and $\mathrm{E}^*((\eta_j^*)^3) = 1$.

Step 3: Let $\widetilde{\phi}_n^+$ be the least squares estimator obtained by minimizing $\sum_{j=p}^{n-1}(X_{j+1}^+ - c^\top X_{j,p})^2$ with respect to $c \in \mathbb{R}^p$.

Step 4: Approximate the distribution of $\sqrt{n}(\widetilde{\phi}_n - \phi)$ by the distribution of the bootstrap analogue $\sqrt{n}(\widetilde{\phi}_n^+ - \widetilde{\phi}_n)$.

We investigate the finite sample performance of the four different bootstrap methods proposed so far and which can be used to approximate the distribution of the least squares estimator $\sqrt{n}(\widetilde{\phi}_n - \phi)$, where $\widetilde{\phi}_n = \widetilde{\phi}_n^{LS}$ is the least squares estimator.

We refer to the model-based bootstrap, the bootstrap of segments, the regression-type bootstrap and the regression-type bootstrap with wild errors. Recall that all four bootstrap methods are asymptotically valid. Time series of length $n = 100$ are generated stemming from the first-order autoregressive process

$$X_t = \phi X_{t-1} + \varepsilon_t, \tag{32}$$

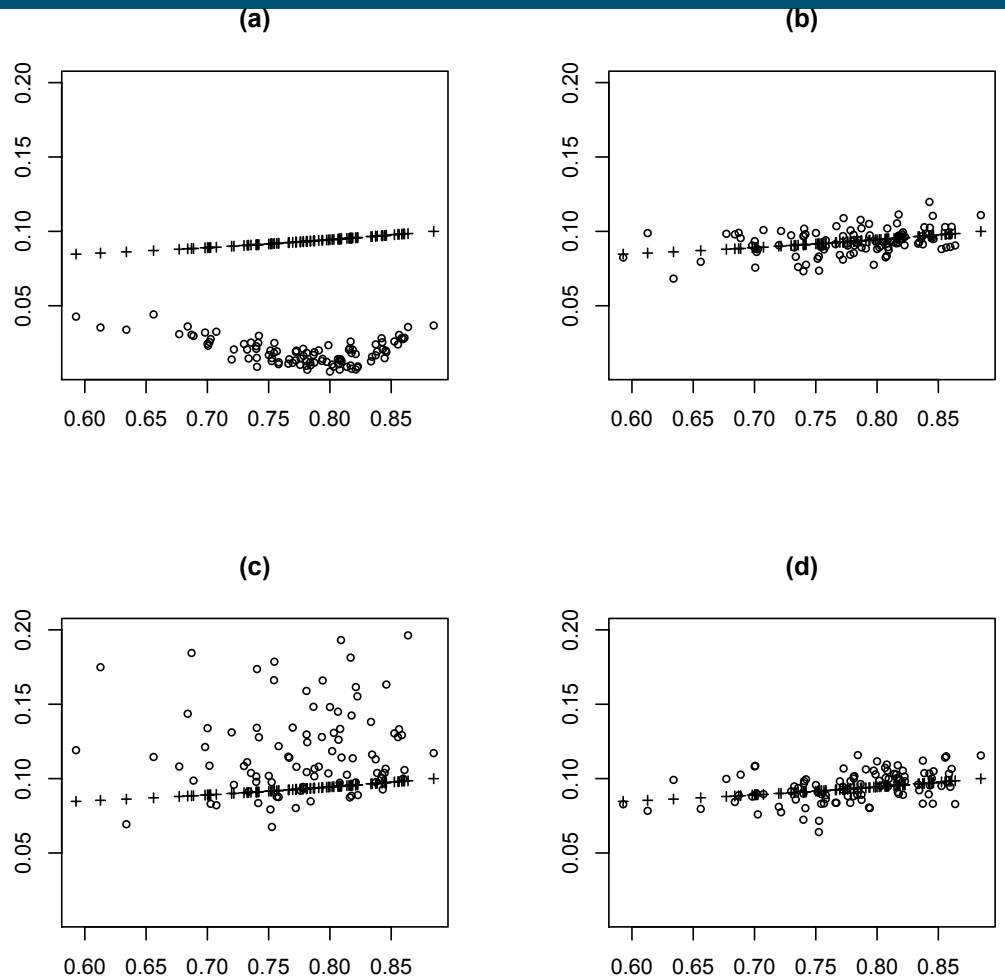
with $\phi = 0.8$ and ε_t i.i.d. $\mathcal{N}(0, 1)$ innovations.

The goal is to estimate the distribution of $S_n = \sqrt{n}(\widetilde{\phi}_n - \phi)$ by using each one of the aforementioned four different bootstrap procedures.

We also compare the performance of the different bootstrap methods with the approximation of the distribution of interest provided by the asymptotic distribution.

Recall that distribution is Gaussian with mean zero and estimated variance $(1 - (\widetilde{\phi}_n)^2)$. In order to judge the quality of these estimators we calculate the empirical d_1-distance, see (31), between the estimated exact distribution of S_n (estimated using 10,000 replications), the asymptotic Gaussian distribution and the distribution of the estimates obtained using each of the four different bootstrap proposals.
$R = 100$ replications have been considered and all bootstrap estimators are based on $B = 1,000$ bootstrap samples. The results obtained are presented in the following figure.

A Numerical Comparison



Empirical d_1-distances (vertical axis) against the estimated parameter $\widetilde{\phi}_n^{(r)}$, $r = 1, 2, \ldots, R = 100$, (horizontal axis). The crosses (+) refer to the distance of the asymptotic and the circles (o) to the distance of the bootstrap distribution to the estimated exact distribution of $\sqrt{n}(\widetilde{\phi}_n - \phi)$. Panel (a) refers to the model-based autoregressive bootstrap, (b) to the regression-type bootstrap, (c) to the bootstrap of segments and (d) to the regression-type bootstrap with wild innovations.

The findings of the numerical study show some striking differences in the finite sample performance of the four different bootstrap methods considered.

In particular, it is seen in the figure that, apart from the model-based autoregressive bootstrap, no one of the other three methods seem to outperform the asymptotic Gaussian approximation.

A typical assumption in the literature is, that the underlying process is a linear AR(∞)-process, that is X_t is generated as

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t \,,\; t \in \mathbb{Z} \,, \tag{33}$$

with respect to i.i.d. innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$ having mean zero, variance $\mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2 \in (0, \infty)$, fourth-order moment $\mathrm{E}(\varepsilon_t^4) < \infty$ and absolutely summable coefficients ϕ_j, $j = 1, 2, \ldots$, with $\phi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for all $|z| \leq 1$.

However, as we will see, the applicability of the AR-sieve bootstrap is not restricted to this particular and linear class of processes.
In fact, the autoregressive fit can be thought of as an approximation of the unknown dependence structure of an underlying general stationary process.

A theoretical justification of such a view of an autoregressive fit is offered by the basic fact that any stationary stochastic process, the second order structure of which satisfies certain (general) properties, obeys a so-called **autoregressive representation**.

By this we mean a representation as in (34) but with the i.i.d. white noise $\{\varepsilon_t, t \in \mathbb{Z}\}$ replaced by a general white noise $\{e_t, t \in \mathbb{Z}\}$ which may be only uncorrelated and not i.i.d.:

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + e_t \,,\; t \in \mathbb{Z}\,, \tag{34}$$

Suppose we are interested in the statistics $T_n = T_n(X_1, X_2, \ldots, X_n)$, for example a generalized mean, for which we assume that $S_n = c_n(T_n - \theta)$, for some θ, possesses a non-degenerate limit distribution.

Under the following Assumption

Let $\{X_t, t \in \mathbb{Z}\}$ be a (weakly) stationary purely non-deterministic and centered time series. The autocovariance function $\gamma(\cdot)$ of $\{X_t\}$ fullfils $\sum_{k\in\mathbb{Z}}(1+|k|)^r|\gamma(k)| < \infty$ for some $r \geq 0$ to be specified in the respective results later on. The spectral density of $\{X_t, t \in \mathbb{Z}\}$,

$$f(\lambda) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\gamma(k)\mathrm{e}^{-\mathrm{i}k\lambda}\ ,\ \lambda \in (-\pi, \pi]\,,$$

fulfills $f(\lambda) \geq c_f$ for all frequencies λ for some $c_f > 0$

we can show the following result.

Theorem 20

Let $\{X_t, t \in \mathbb{Z}\}$ fulfill the above Assumption for some $r \geq 0$. Then there exist uniquely determined autoregressive (AR) coefficients $\{a_k, k \in \mathbb{N}\}$ and an uncorrelated white noise process $\{e_t, t \in \mathbb{Z}\}$, such that

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + e_t \ , \ t \in \mathbb{Z} \ , \tag{35}$$

and such that $\sum_{k=1}^{\infty} a_k X_{t-k}$ represents the L^2-projection of X_t onto the infinite past $\overline{span}\{X_{t-k} : k = 1, 2, \ldots\}$.

The uncorrelated white noise process $\{e_t, t \in \mathbb{Z}\}$, which consists of the uniquely determined residuals of this projection, is called the innovation process of $\{X_t, t \in \mathbb{Z}\}$.
The coefficients in (35) fulfill the summability $\sum_{k=1}^{\infty}(1+k)^r |a_k| < \infty$.
The power series $A(z) := 1 - \sum_{k=1}^{\infty} a_k z^k$ can be defined and has no zeroes within the closed unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ and therefore is always invertible with $B(z) = 1/A(z)$ for all $|z| \leq 1$.

Corollary 21

Under the same assumptions as in Theorem 20, there are uniquely determined moving average (MA) coefficients $(b_k, k \in \mathbb{N})$ *such that for the innovation process* $\{e_t, t \in \mathbb{Z}\}$ *of* $\{X_t, t \in \mathbb{Z}\}$ *the representation*

$$X_t = e_t + \sum_{k=1}^{\infty} b_k e_{t-k} \ , \ \ t \in \mathbb{Z} \ , \tag{36}$$

is valid. This is the so-called Wold representation.

The MA coefficients fulfill

$$\sum_{k=1}^{\infty} (1+k)^r |b_k| < \infty \, . \tag{37}$$

Theorem 22

Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary process (with autocovariance function γ) which fulfills the above Assumption for an $r > 0$. Let $(a_k(p) : k = 1, \ldots, p)$ be its finite predictor coefficients as defined as

$$(a_1(p), \ldots, a_p(p)) := \mathit{argmin}_{c_1,\ldots,c_p} \mathrm{E}\big((X_t - \sum_{k=1}^{p} c_k X_{t-k})^2\big)\,. \tag{38}$$

and let $(a_k, k \in \mathbb{N})$ be the AR coefficients from Theorem 20. Let $K := \sum_{k=-\infty}^{\infty} |\gamma(k)|$. Then, for all $s \geq 0$ with $s + 1/2 \leq r$ and for all $p \in \mathbb{N}$, it holds true that,

$$\sum_{k=1}^{p} (1+k)^s |a_k(p) - a_k| \leq \frac{K}{2\pi c_f} \sum_{k=p+1}^{\infty} (1+k)^{s+1/2} |a_k|\,, \tag{39}$$

where $c_f > 0$ is defined in the above Assumption. Due to Theorem 20 the right hand side converges to zero as p tends to infinity.

Step 1: Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a p-th order autoregressive model to $X_1, X_2, \ldots, X_n$. Denote by $\widehat{a}_n(p) = (\widehat{a}_k(p), k = 1, 2, \ldots, p)$ the autoregressive parameter estimates which are assumed to fulfill an assumption given below.

Step 2: Let $\widetilde{e}_t(p) = X_t - \sum_{j=1}^{p} \widehat{a}_j(p) X_{t-j}$, $t = p+1, p+2, \ldots, n$, be the residuals of the autoregressive fit and denote by $\widehat{F}_{n,e}$ the empirical distribution function of the centered residuals $\widehat{e}_t(p) = \widetilde{e}_t(p) - \overline{e}_n$, where $\overline{e}_n = (n-p)^{-1} \sum_{t=p+1}^{n} \widetilde{e}_t(p)$. Let $X_1^*, X_2^*, \ldots, X_n^*$ be a pseudo time series stemming from the process $\{X_t^*, t \in \mathbb{Z}\}$, where $X_t^* = \sum_{j=1}^{p} \widehat{a}_j(p) X_{t-j}^* + e_t^*$ and the e_t^*'s are independent random variables each having distribution $\widehat{F}_{n,e}$.

Step 3: Let $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^*)$ be the same statistic as T_n based on $X_1^*, X_2^*, \ldots, X_n^*$ and θ^* the analogue of θ associated with the bootstrap process $\{X_t^*, t \in \mathbb{Z}\}$. The AR-sieve bootstrap approximation of S_n is then given by $S_n^* = c_n(T_n^* - \theta^*)$.

Assume for the following that parameter estimates $\widehat{a}_n(p) = (\widehat{a}_k(p), k = 1, \ldots, p)$ of an autoregressive fit of order p exist, and fulfill for increasing $p = p(n)$ the following consistency condition

$$p(n)^2 \sum_{k=1}^{p(n)} |\widehat{a}_k(p(n)) - a_k(p(n))| = \mathcal{O}_{P_n}(1)\,,$$

as $n \to \infty$. Here $a(p) = (a_k(p), k = 1, \ldots, p)$ is defined as in (38).

From Theorem 20 we know that if $\{X_t, t \in \mathbb{Z}\}$ satisfies

> $\{X_t, t \in \mathbb{Z}\}$ is a zero mean, strictly stationary and purely non-deterministic stochastic process. The autocovariance function γ satisfies $\sum_{h=0}^{\infty}(1+h)^r|\gamma(h)| < \infty$ for some $r \geq 0$ specified in the respective results and the spectral density f is bounded away from zero by a constant $c_f > 0$. Furthermore $\mathrm{E}(X_t^4) < \infty$.

Then X_t possesses an autoregressive representation with an uncorrelated white noise process $\{e_t, t \in \mathbb{Z}\}$, i.e.,

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + e_t, \quad t \in \mathbb{Z},$$

cf. equation (35). Because of strict stationarity of $\{X_t, t \in \mathbb{Z}\}$ we have that the process $\{e_t, t \in \mathbb{Z}\}$ is strictly stationary as well and therefore that the marginal distribution $\mathcal{L}(e_t)$ of e_t does not depend on t.

The following theorem is the main result for the AR-sieve bootstrap and gives a clear answer to the question of interest.

To state this theorem, we define the following *companion* autoregressive process $\{\widetilde{X}_t, t \in \mathbb{Z}\}$, where $\widetilde{X}_t$ is generated as follows

$$\widetilde{X}_t = \sum_{j=1}^{\infty} a_j \widetilde{X}_{t-j} + \widetilde{\varepsilon}_t, \qquad t \in \mathbb{Z}\,. \tag{40}$$

Here, the coefficients $(a_j, j \in \mathbb{N})$ are identical to the coefficients in (35) and $\{\widetilde{\varepsilon}_t, t \in \mathbb{Z}\}$ consists of i.i.d. random variables whose marginal distribution is identical to that of e_t from (35), i.e., $\mathcal{L}(\widetilde{\varepsilon}_t) = \mathcal{L}(e_t)$.

We will see that the bootstrap process $\{X_t^*, t \in \mathbb{Z}\}$, with X_t^* defined as in Step 2 of the AR-sieve Bootstrap Proposal, asymptotically mimics the dependence structure of the companion process $\{\widetilde{X}_t, t \in \mathbb{Z}\}$.

It is easily seen that all second order properties of the processes $\{\widetilde{X}_t, t \in \mathbb{Z}\}$ and $\{X_t, t \in \mathbb{Z}\}$, like autocovariance function and spectral density, coincide, while the probabilistic characteristics beyond second order quantities of both stationary processes are not necessarily the same.

This latter statement is due to the differences in the stochastic properties of the innovation processes $\{e_t\}$ and $\{\widetilde{\varepsilon}_t\}$.

Recall that $\{e_t, t \in \mathbb{Z}\}$ in (35) is an uncorrelated white noise process while $\{\widetilde{\varepsilon}_t, t \in \mathbb{Z}\}$ in (40) is an i.i.d. white noise process.

Now, we focus on generalized mean statistics T_n and let $\widetilde{T}_n$ be the same statistic as T_n but with X_t replaced by $\widetilde{X}_t$, i.e.,

$$\widetilde{T}_n = f\Big(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g(\widetilde{X}_t, \ldots, \widetilde{X}_{t+m-1})\Big), \tag{41}$$

The main message of the following theorem is that the AR-sieve bootstrap applied to a time series $X_1, \ldots, X_n$ in order to approximate the distribution of a generalized mean statistic T_n will generally lead to an asymptotically consistent estimation of the distribution of the statistic $\widetilde{T}_n$.

This implies that for the class of generalized mean statistics, the AR-sieve bootstrap works **if and only if** the limiting distributions of T_n and of $\widetilde{T}_n$ are identical.

Theorem 23

Under appropriate assumptions and the moment condition $\mathrm{E}(e_t^{2(h+2)}) < \infty$ *cf. Theorem 20 for the special uncorrelated white noise process* $\{e_t, t \in \mathbb{Z}\}$ *explained there). Moreover, assume that* $p(n) = o((n/\log n)^{1/4})$ *for the order* $p(n)$ *of the autoregression fitted to the data and that*

(i) *The empirical distribution function* $F_{n,e}$ *of the random variables* $e_1, \ldots, e_n$, *where* $e_t = X_t - \sum_{k=1}^{\infty} a_k X_{t-k}$ *converges weakly to the distribution function* F_e *of* $\mathcal{L}(e_1)$.

(ii) $n^{-1} \sum_{t=1}^{n} e_t^r \xrightarrow{P_n} \mathrm{E}(e_1^r)$ *for all* $r \leq 2(h+2)$.

Then,

$$d_K\left(\mathcal{L}(\sqrt{n}(T_n^* - f(\theta^*))|\underline{X}_n), \mathcal{L}(\sqrt{n}(\widetilde{T}_n - f(\widetilde{\theta})))\right) \to 0 \tag{42}$$

in P_n*-probability as* $n \to \infty$. *Here* $\theta^* = \mathrm{E}^*(g(X_t^*, \ldots, X_{t+m-1}^*))$, *and* $\widetilde{\theta} = \mathrm{E}(g(\widetilde{X}_t, \ldots, \widetilde{X}_{t+m-1}))$.

Assumptions for the Theorem

Differentiability assumptions for the functions f and g appearing in the definition of the generalized mean statistics

$$T_n = f\Big(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g\big(X_t, X_{t+1}, \ldots, X_{t+m-1}\big)\Big). \tag{43}$$

$\{X_t, t \in \mathbb{Z}\}$ is a zero mean, strictly stationary and purely non-deterministic stochastic process. The autocovariance function γ satisfies $\sum_{h=0}^{\infty}(1+h)^r|\gamma(h)| < \infty$ for some $r \geq 0$ specified in the respective results and the spectral density f is bounded away from zero by a constant $c_f > 0$. Furthermore $\mathrm{E}(X_t^4) < \infty$.

Assume that parameter estimates $\widehat{a}_n(p) = (\widehat{a}_k(p), k = 1, \ldots, p)$ of an autoregressive fit of order p exist, and fulfill for increasing $p = p(n)$ the following consistency condition

$$p(n)^2 \sum_{k=1}^{p(n)} |\widehat{a}_k(p(n)) - a_k(p(n))| = \mathcal{O}_{P_n}(1) \; n \to \infty,$$

Here $a(p) = (a_k(p), k = 1, \ldots, p)$ defined as in (38).

Theorem 23 can be used as a criterion for checking validity of AR-sieve bootstrap.

It is only necessary to check whether the asymptotic distributions of $T_n = T_n(X_1, \ldots, X_n)$ based on the observed time series is identical to the asymptotic distribution of the statistic $\widetilde{T}_n = T_n(\widetilde{X}_1, \ldots, \widetilde{X}_n)$ based on fictitious observations $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ stemming from the companion process $\{\widetilde{X}_t, t \in \mathbb{Z}\}$.

If and only if this is the case, the AR-sieve bootstrap works asymptotically.

This message of Theorem 23, which is illustrated in the following by means of some specific examples.

Consider the sample mean $T_n = \overline{X}_n \equiv n^{-1} \sum_{t=1}^{n} X_t$.
Under mild regularity conditions:

$$\sqrt{n}\,(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sum_{h=-\infty}^{\infty} \gamma(h)) \text{ as } n \to \infty.$$

The asymptotic distribution depends only on second order properties of $\{X_t, t \in \mathbb{Z}\}$. Since the companion process $\{\widetilde{X}_t, t \in \mathbb{Z}\}$ has the same second order properties as $\{X_t, t \in \mathbb{Z}\}$ we immediately get by Theorem 23 that the AR-sieve bootstrap asymptotically works in the case of the sample mean for general stationary time series possessing a bounded and strictly positive spectral density.

Let $T_n = n^{-1}\sum_{t=1}^{n-h} X_t X_{t+h}$. It can be shown (under assumptions) that

$$\sqrt{n}(T_n - \gamma(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2), \text{ as } n \to \infty,$$

where

$$\begin{aligned}\tau^2 &= 2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \cos(\lambda_1 h)\cos(\lambda_2 h) f_4(\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2 \\ &\quad + 4\pi \int_{-\pi}^{\pi} \cos^2(\lambda h)\, f^2(\lambda) d\lambda\,.\end{aligned}$$

and $f_4(\lambda_1, \lambda_2, \lambda_3) = \dfrac{1}{(2\pi)^3} \sum_{h_1, h_2, h_3 = -\infty}^{\infty} \operatorname{cum}(X_0, X_{h_1}, X_{h_2}, X_{h_3}) \mathrm{e}^{-\mathrm{i}\sum_{r=1}^{3}\lambda_r h_r}$,

is the fourth-order cumulant spectrum.

For the companion process $\{\widetilde{X}_t, t \in \mathbb{Z}\}$: $\sqrt{n}(\widetilde{T}_n - \gamma(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \widetilde{\tau}^2)$ with

$$\widetilde{\tau}^2 = \Big(\frac{\mathrm{E}\,(\widetilde{\varepsilon}_1^4)}{(\mathrm{E}\,(\widetilde{\varepsilon}_1^2))^2} - 3\Big)\gamma(h)^2 + 4\pi \int_{-\pi}^{\pi} \cos^2(\lambda h)\, f^2(\lambda) d\lambda \qquad (44)$$

The variances τ^2 and $\widetilde{\tau}^2$ do not coincide in general!

Interestingly enough, even if the underlying process $\{X_t, t \in \mathbb{Z}\}$ is a linear time series, i.e., satisfies a MA(∞) representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \tag{45}$$

with i.i.d. random variables $\{\varepsilon_t, t \in \mathbb{Z}\}$, the AR-sieve bootstrap in general fails for sample autocovariances.

But: The AR-Sieve Bootstrap works for sample autocovariances of *Gaussian linear time series*!

Moreover: AR-Sieve Bootstrap works for sample autocorrelations of *linear time series*, because the asymptotic distribution for sample autocorrelations of linear time series only depends on the autocovariance function (so-called *Bartlett-formula*, cf. Brockwell and Davis, Theorem 7.2.1.

Let us consider time series of length $n = 100$ generated from a first order autoregressive moving average (ARMA) model,

$$X_t = \phi X_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}. \tag{46}$$

Here $\phi = 0.8$ and $\{\varepsilon_t, t \in \mathbb{Z}\}$ an i.i.d. process with mean zero and unit variance. Two values of the parameter θ are considered, $\theta = 0.5$ and $\theta = 2$. Two different distributions for the i.i.d. innovations: standard Gaussian and a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$.

Also two statistics are considered, the sample autocovariance at lag one, $\widehat{\gamma}(1)$, and the sample autocorrelation at the same lag, $\widehat{\rho}(1)$. The AR-sieve bootstrap approximations of the distributions of

$$S_{1,n} = \sqrt{n}\,(\widehat{\gamma}(1) - \gamma(1)) \text{ and } S_{2,n} = \sqrt{n}\,(\widehat{\rho}(1) - \rho(1)),$$

have been calculated using $B = 1,000$ bootstrap replications and for three choices of the order p of the autoregressive fit:
$p = 2$, $p = 4$ and, p based on the AIC criterion.

AR-sieve compared with asymptotic normal approximation.

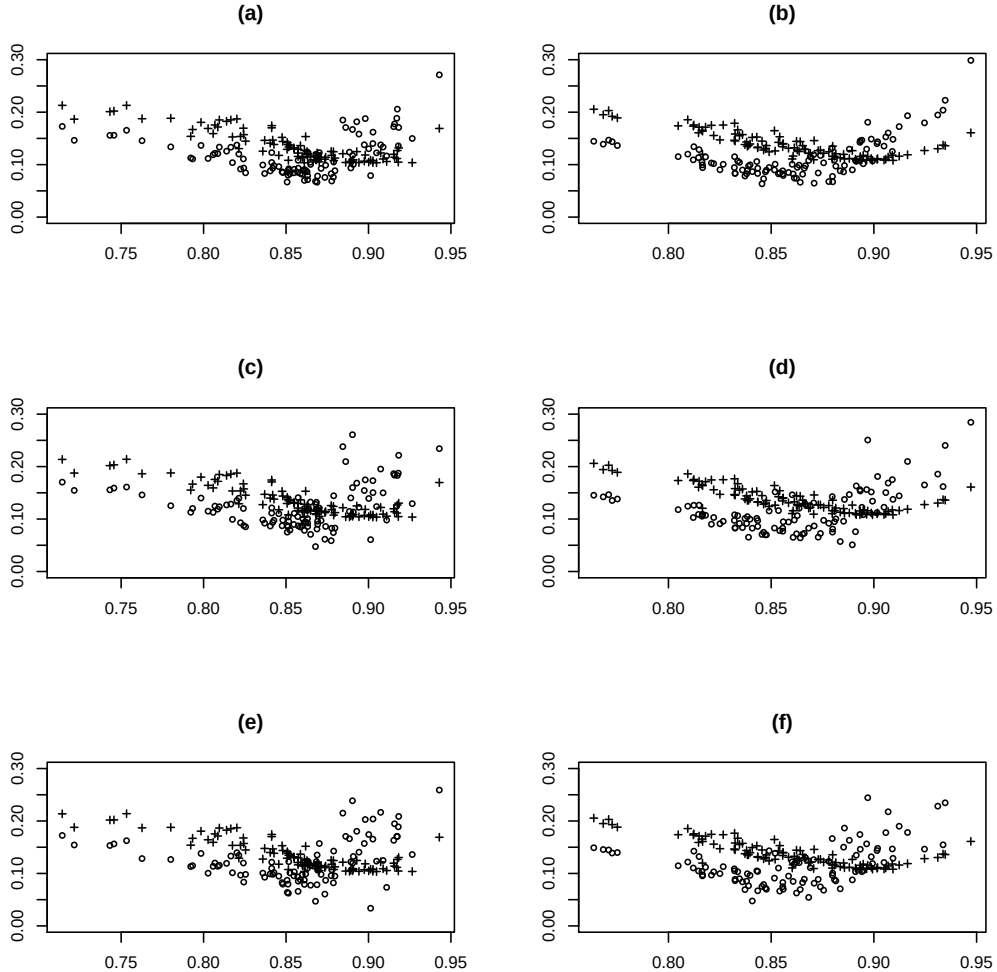
To assess the overall performance of the different methods in estimating the distributions of interest, we calculated the empirical d_1-distances over $R = 100$ replications between the bootstrap (respectively the asymptotic Gaussian) distribution and the estimated exact distribution of interest.

The latter has been obtained using 1,000 replications of model (46). See equation (31) for the definition of the empirical d_1-distance.
The results obtained in the case $\theta = 0.5$ are shown in the following figures for the statistic $\widehat{\gamma}(1)$ and for the statistic $\widehat{\rho}(1)$.

As it is seen from the corresponding exhibits, the AR-sieve bootstrap clearly outperforms the asymptotic Gaussian approximation only for the statistic $\widehat{\rho}(1)$.

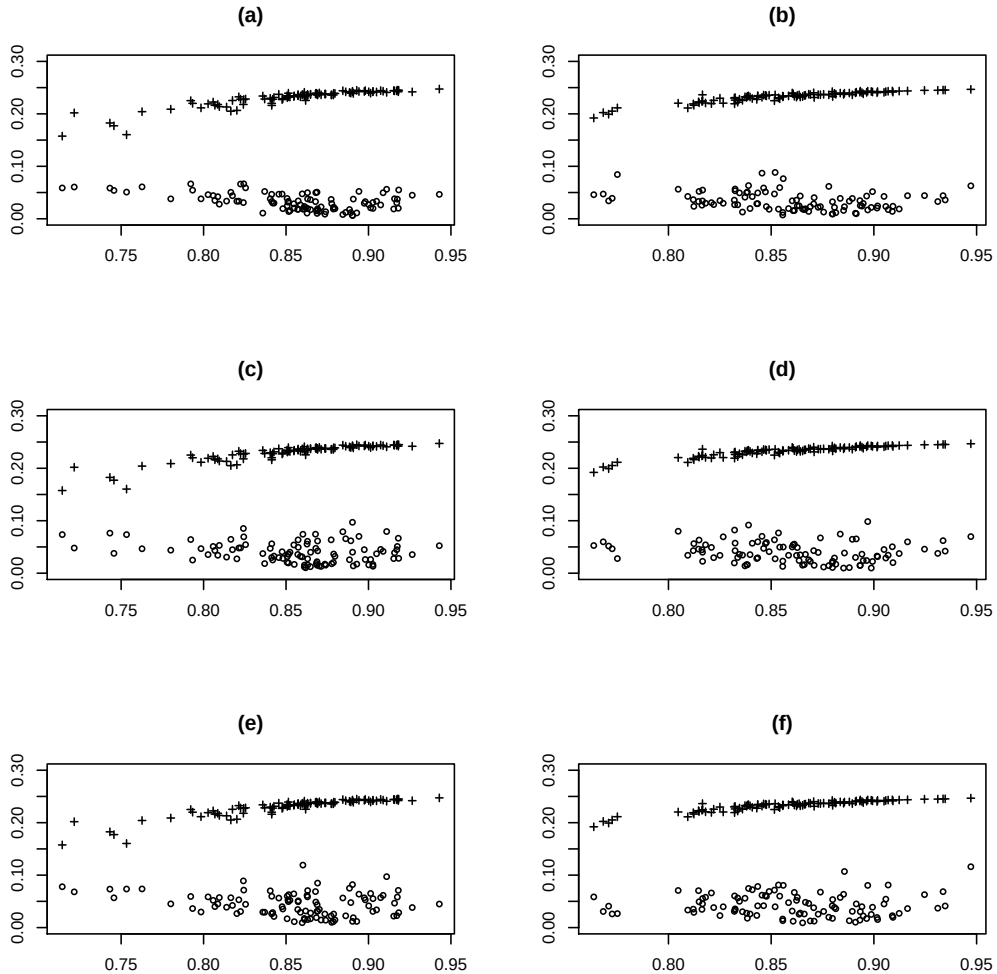
AR-sieve bootstrap seem not to be very sensitive with respect to the choice of the autoregressive order p used.

Finite Sample Properties of AR-Sieve Bootstrap



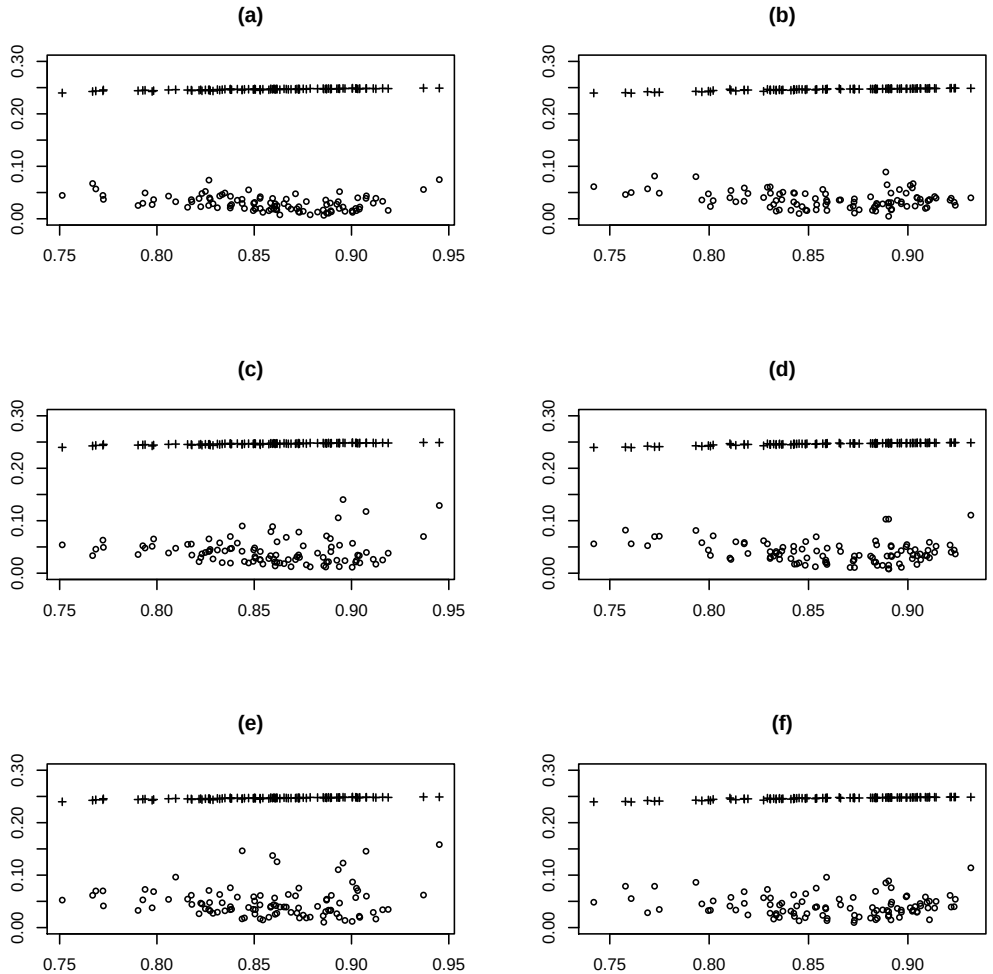
Estimation of the distribution of $\sqrt{n}\,(\widehat{\gamma}(1) - \gamma(1))$ when $\phi = 0.8$ and $\theta = 0.5$: Empirical d_1-distances (vertical axis) against the estimated parameter $\widehat{\rho}(1)$ (horizontal axis). The crosses (+) refer to the distance of the asymptotic and the circles (o) to the distance of the AR-sieve bootstrap distribution both to the estimated exact distribution. The plots in the first row are those obtained for $p = 2$, in the second row for $p = 4$ and in the third row for p selected according to AIC. The plots in the first column are for Gaussian while in the second column for uniformly distributed innovations.

Finite Sample Properties of AR-Sieve Bootstrap



Estimates of the distribution of $\sqrt{n}\,(\widehat{\rho}(1) - \rho(1))$ when $\phi = 0.8$ and $\theta = 0.5$: Empirical d_1-distances (vertical axis) against the estimated parameter $\widehat{\rho}(1)$ (horizontal axis). The crosses (+) refer to the distance of the asymptotic and the circles (o) to the distance of the AR-sieve bootstrap distribution both to the estimated exact distribution. The plots in the first row are those obtained for $p = 2$, in the second row for $p = 4$ and in the third row for p selected according to AIC. The plots in the first column are for Gaussian while in the second column for uniform innovations.

Finite Sample Properties of AR-Sieve Bootstrap



Estimates of the distribution of $\sqrt{n}\,(\widehat{\rho}(1) - \rho(1))$ when $\phi = 0.8$ and $\theta = 2.0$: Empirical d_1-distances (vertical axis) against the estimated parameter $\widehat{\rho}(1)$ (horizontal axis). The crosses (+) refer to the distance of the asymptotic and the circles (o) to the distance of the AR-sieve bootstrap distribution both to the estimated exact distribution. The plots in the first row are those obtained for $p = 2$, in the second row for $p = 4$ and in the third row for p selected according to AIC. The plots in the first column are for Gaussian while in the second column for uniform innovations.

Next we discuss fully nonparametric bootstrap methods for time series. The different proposals that will be discussed are variations of the simple idea to generate a new set of pseudo random variables $X_1^*, X_2^*, \ldots, X_n^*$, by just resampling segments or blocks of consecutive observations of the time series at hand instead of resampling single observations, as this is done in the case of an i.i.d. sample $X_1, X_2, \ldots, X_n$.

Bootstrapping of blocks of observations, is a method capable to extend the original idea of Efron (1979) of *drawing with replacement*, to temporal dependent observations.

The main idea of all block bootstrap procedures consists of dividing the data into blocks of consecutive observations of length ℓ,

$$(X_t, X_{t+1}, \ldots, X_{t+\ell-1}),$$

say, and sampling the blocks randomly with replacement from the set of all possible blocks.

The differences between the various block bootstrap methods proposed depend on whether one considers overlapping or non overlapping blocks and whether the length ℓ of the blocks is allowed to be random or not.

The nonparametric bootstrap procedure we consider is the so-called moving block bootstrap (MBB) which takes overlapping blocks.

This bootstrap procedure works by resampling blocks of length ℓ, $\ell \in \mathbb{N}$, $1 \leq \ell \ll n$, from the set

$$\{(X_1, \ldots, X_\ell), (X_2, \ldots, X_{\ell+1}), \ldots, (X_{n-\ell+1}, \ldots, X_n)\},$$

and merging them together in the order selected to form a new pseudo-time series.

Let θ be the parameter of interest and $T_n = T(X_1, \ldots, X_n)$ a $\sqrt{n}$-consistent estimator of this parameter.

As usual we are interested in the distribution of $S_n = \sqrt{n}(T_n - \theta)$.

Step 1: Select a block length $1 \leq \ell = \ell(n) < n$.

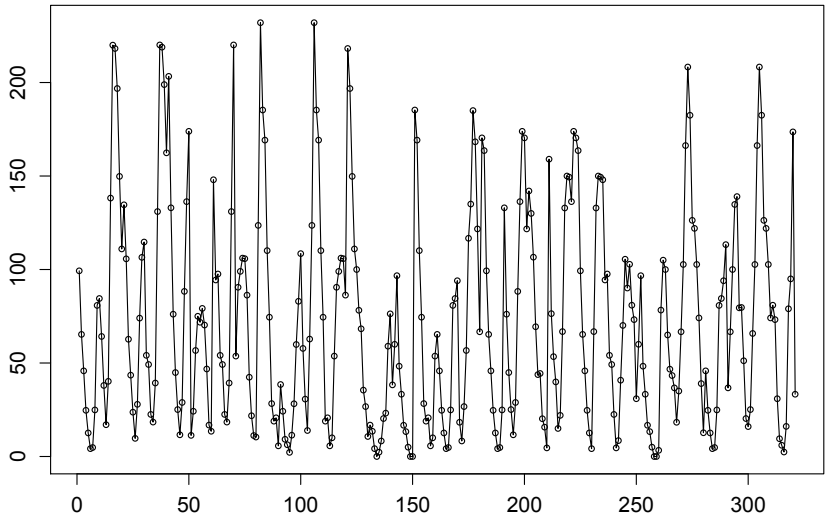
Step 2: Generate i.i.d. random variables $I_1, I_2, \ldots,$ each having a uniform distribution on the set $\{1, 2, \ldots, n - \ell + 1\}$.

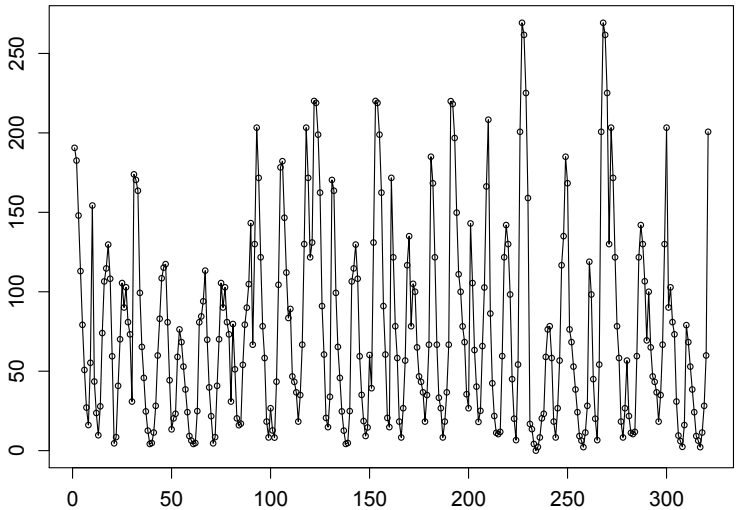
Step 3: $(X_{I_j}, X_{I_j+1}, \ldots, X_{I_j+\ell-1})$, $j = 1, 2, \ldots$: The resampled blocks.

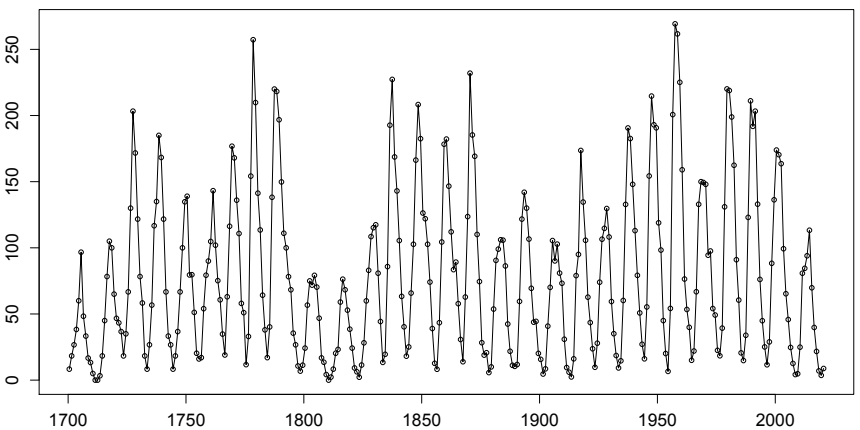
Step 4: The bootstrap sample $X_1^*, \ldots, X_n^*$ is obtained by merging the resampled blocks, that is, by laying them one after the other together in the order selected, and possibly discarding some last observations in order to obtain a sample of length n.

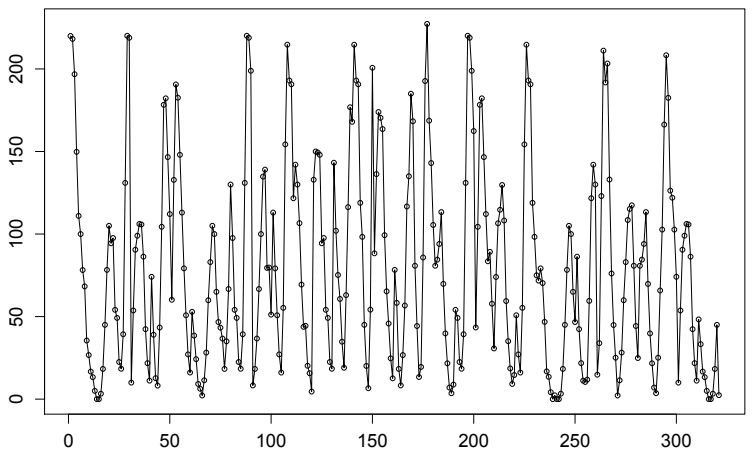
Step 5: Let $T_n^* = T(X_1^*, \ldots, X_n^*)$ be the bootstrap analogue of $T_n = T(X_1, X_2, \ldots, X_n)$ and consider the distribution of $S_n^* = \sqrt{n}(T_n^* - \theta^*)$, where the centering sequence θ^* stays unspecified for the moment. Note that in some applications we may let $\theta^* = T_n$. We then approximate the distribution of $S_n = \sqrt{n}(T_n - \theta)$ by the distribution of $S_n^* = \sqrt{n}(T_n^* - \theta^*)$.

Sunspot Numbers









Sunspot Numbers plus 3 block bootstrap time series with block size $\ell = 10$.

For the sample mean $T_n = \overline{X}_n$ the distribution which we intend to approximate is that of $\sqrt{n}(\overline{X}_n - \mu)$, where μ denotes the mean of the underlying, strictly stationary process $\{X_t, t \in \mathbb{Z}\}$. Under quite mild additional assumptions, it is well known that,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sum_{h=-\infty}^{\infty} \gamma(h)\big), \tag{47}$$

as $n \to \infty$.

As a moving block bootstrap approximation we consider $\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*))$, where $\overline{X}_n^*$ denotes the sample mean of $X_1^*, \ldots, X_n^*$ which are generated according to the Moving Bootstrap Proposal.

$\mathrm{E}^*(\overline{X}_n^*)$ will be clarified later.

We will show that in P_n-probability,

$$\mathcal{L}\big(\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*))|\underline{X}_n\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sum_{h=-\infty}^{\infty} \gamma(h)\big) \text{ as } n \to \infty.$$

To show that the Moving Block Bootstrap works for the sample mean we assume:

(i) $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$

(ii) $\ell \to \infty$ and $\ell/\sqrt{n} \to 0$ as $n \to \infty$

(iii) $\sum_{h=-(\ell-1)}^{\ell-1} \left(1 - \frac{|h|}{\ell}\right)\widehat{\gamma}(h) \xrightarrow{P_n} \sum_{h=-\infty}^{\infty} \gamma(h)$.

To understand the real nature of $T_n^* = \overline{X}_n^*$, write it in the following form

$$T_n^* = \frac{1}{n}\sum_{t=1}^{n} X_t^* = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{\ell}\sum_{t=1}^{\ell} X_{I_j+t-1}\right). \tag{48}$$

Observe that, given the time series $X_1, X_2, \ldots, X_n$, the bootstrap pseudo random variable T_n^* is a sum of the $k, k \cdot \ell = n$, i.i.d. random variables $\ell^{-1}\sum_{t=1}^{\ell} X_{I_j+t-1}$, because the random variables I_j are i.i.d.

It is then an easy task (left as exercise) to verify that

$$\begin{aligned} \mathrm{E}^*(T_n^*) &= \frac{1}{\ell}\sum_{t=1}^{\ell}\mathrm{E}^*(X_{I_1+t-1}) \\ &= \frac{1}{n-\ell+1}\Big(\sum_{t=1}^{\ell-1}\frac{t}{\ell}X_t + \sum_{t=\ell}^{n-\ell+1}X_t + \sum_{t=n-\ell+2}^{n}\frac{n-t+1}{\ell}X_t\Big) \\ &= \sum_{t=1}^{n}\omega_{t,n}X_t\,, \end{aligned} \tag{49}$$

with an obvious definition of the weights $\omega_{t,n}$, $\sum_{t=1}^{n}\omega_{t,n}=1$.

To proceed, consider (recall that $n = k\,\ell$)

$$\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*)) = \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{1}{\sqrt{\ell}}\sum_{t=1}^{\ell}(X_{I_j+t-1} - \mathrm{E}^*(X_{I_j+t-1})) = \frac{1}{\sqrt{k}}\sum_{j=1}^{k}V_{I_j,n}\,,$$

with an obvious notation for the i.i.d. random variables $V_{I_j,n}$.

We can calculate explicitely $\frac{1}{\ell}\sum_{t=1}^{\ell} \mathrm{E}^*(X_{I_1+t-1}) = \overline{X}_n + \mathcal{O}_{P_n}(\ell/n)$ and

$$\mathrm{Var}^*\left(\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*))\right) = \mathrm{E}^*\left(V_{I_1,n}^2\right)$$

$$= \sum_{h=-(\ell-1)}^{\ell-1} \frac{1}{\ell} \sum_{t=1}^{\ell-|h|} \frac{1}{n-\ell+1} \sum_{s=1}^{n-\ell+1} (X_{s+t+|h|-1} - \overline{X}_n)(X_{s+t-1} - \overline{X}_n) + \mathcal{O}_{P_n}\left(\frac{\ell^3}{n^2}\right).$$

Direct but tedious calculations show that the last expression, up to asymptotically vanishing terms, equals

$$\sum_{h=-(\ell-1)}^{\ell-1} \frac{1}{\ell} \sum_{t=1}^{\ell-|h|} \frac{1}{n} \sum_{r=1}^{n-|h|} (X_r - \overline{X}_n)(X_{r+|h|} - \overline{X}_n)$$

$$= \sum_{h=-(\ell-1)}^{\ell-1} \left(1 - \frac{|h|}{\ell}\right) \frac{1}{n} \sum_{r=1}^{n-|h|} (X_r - \overline{X}_n)(X_{r+|h|} - \overline{X}_n)$$

$$= \sum_{h=-(\ell-1)}^{\ell-1} \left(1 - \frac{|h|}{\ell}\right) \widehat{\gamma}(h) = \sum_{h=-\infty}^{\infty} \gamma(h) + o_{P_n}(1)\,.$$

It remains to show that $\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*)) = k^{-1/2}\sum_{j=1}^k V_{I_j,n}$ is asymptotically normal.

But, in the "bootstrap world" $\sum_{j=1}^k V_{I_j,n}$ is a sum of i.i.d. random variables!!

That is verifying Lindeberg's condition suffices to show that

$$\mathcal{L}\left(\sqrt{n}(\overline{X}_n^* - \mathrm{E}^*(\overline{X}_n^*))\middle|\underline{X}_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sum_{h=-\infty}^{\infty}\gamma(h)) \quad \text{in } P_n\text{-probability},$$

as $n \to \infty$, which establishes asymptotic validity of the moving block bootstrap for the sample mean.

Take home message: Moving Block Bootstrap for sample means only works, if in the real world asymptotic normality holds for $\overline{X}_n$.

Suppose that the time series $\{X_t, t \in \mathbb{Z}\}$ is strictly stationary, that the assumption and that $X_1^*, X_2^*, \ldots, X_n^*$ is generated as in the Moving Block Bootstrap Proposal where $\ell \to \infty$ and $\ell/n \to 0$ as $n \to \infty$.

Then for

$$M_n = (n-m+1)^{-1} \sum_{t=1}^{n-m+1} \underbrace{g(X_t, X_{t+1}, \ldots, X_{t+m-1})}_{=:Y_t},$$

under the assumption

$$\sum_{h_1,h_2,h_3 \in \mathbb{Z}} \left|\mathrm{cum}(Y_0^c, Y_{h_1}^c, Y_{h_2}^c, Y_{h_3}^c)\right| < \infty,$$

and if M_n is asymptotic normal the **Moving Block Bootstrap works**, that is

$$d_2\Big(\mathcal{L}\big(\sqrt{n}(M_n^* - \theta^*)\big|\underline{X}_n\big), \mathcal{L}\big(\sqrt{n}(M_n - \theta)\big)\Big) \to 0$$

in P_n-probability, where $\theta^* = \mathrm{E}^*(M_n^*)$.

Bootstrap methods that cannot be covered in this lecture include, for example,

(i) The Stationary Bootstrap (cf. Politis and Romano (1994)).

(ii) The Tapered Block Bootstrap (cf. Paparoditis and Politis (2002)).

(iii) The Dependent Wild Bootstrap (Shao (2010)).

(iv) Subsampling (cf. Politis, Romano and Wolf (1999) and (2001)).

(v) The huge class Frequency Domain Bootstrap Methods (cf. a series of papers K. and Paparoditis)

(vi) The Linear Process Bootstrap (cf. McMurry and Politis (2010)).

(vii) The Markovian Bootstrap (cf. several papers of Paparoditis and Politis, Horowitz, Bertail and co-authors).

(viii) Bootstrap for nonparametric estimators (cf. Neumann and K. (1998) and others).

(ix) Bootstrapping locally stationary processes (K. and Paparoditis (2015)).

Thank you very much for attending !

References

References

Most of the presented material is part of

- Kreiss, J.-P. and Paparoditis, E. (2026). *Bootstrap for Time Series: Theory and Applications*. Springer, New York. Forthcoming.
- Bertail, P. and Clémençon, S. (2006). Regenerative block bootstrap for Markov chains. *Bernoulli* **12**, 689–712.
- Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196–1217.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Second Edition. Springer, New York.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* **7**, 1–16.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansions*. Springer, New York.
- Horowitz, J. L. (2003). Bootstrap Methods for Markov Processes. *Econometrica* **71**, 1049–1082.

References

- Kreiss, J.-P. (1992). Bootstrap procedures for AR(∞)-processes. In: *Lect. Notes in Economics & Mathem. Systems No. 376. (Proc. Bootstrapping and Related Techniques, Trier).* Springer, Heidelberg, 107–113.
- Kreiss, J.-P., Paparoditis, E. and Politis, D.N. (2011). The range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics* **39**, 2103–2130.
- Kreiss, J.-P. and Paparoditis, E. (2015). Bootstrapping locally stationary processes. *Journal of the Royal Statistical Society, Series B* **77**, 267–290.
- Kreiss, J.-P. and Paparoditis, E. (2023). Bootstrapping Whittle Estimators. *Biometrika* **110**, 499–518.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**, 1217–1241.
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York.
- McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis* **31**, 471–482.

References

- Meyer, M., Jentsch, C. and Kreiss, J.-P. (2017). Baxter's inequality and sieve bootstrap for random fields. *Bernoulli* **23**, 2988–3020.
- Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics* **26**, 1570–1613.
- Paparoditis, E. and Politis, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal* **5**, 131–148.
- Politis, D. N. and Romano, J. P. (1994a). The stationary bootstrap. *Journal of the American Statistical Association* **89**, 1303–1313.
- Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* Springer, New York.
- Politis, D. N., Romano, J. P. and Wolf, M. (2001). On the asymptotic theory of subsampling. *Statistica Sinica* **11**, 1105–1124.
- Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association* **489**, 218–235.

Backup material

- Subsampling
- Some Remarks on Second-Order Properties of Bootstrap

Subsampling

Subsampling, introduced by Politis and Romano (1994), is a general resampling procedure which can be applied to a variety of situations and especially to those in which bootstrap procedures do not work.

To motivate the basic idea behind subsampling, let us consider the following example. Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary process with $\mathrm{E}(X_t) = \mu$ and $\mathrm{E}(X_t^2) < \infty$. Assume that $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sum_{h=-\infty}^{\infty} \gamma_X(h))$, as $n \to \infty$. We know that in this case and under standard conditions on the block length ℓ, the MBB procedure works for the sample mean.

Suppose now that we are interested in the distribution of $g(\overline{X}_n)$, where $g : \mathbb{R} \to \mathbb{R}$ is a three times continuously differentiable function with $g'(\mu) = 0$ and $g''(\mu) \neq 0$. Cf. Example 13.
It easily can be seen that

$$S_n = n\big(g(\overline{X}_n) - g(\mu)\big) = \frac{1}{2} n(\overline{X}_n - \mu)^2 g''(\mu) + \mathcal{O}_{P_n}(n^{-1/2}).$$

And also

$$n\big(g(\overline{X}_n) - g(\mu)\big) \xrightarrow{\mathcal{D}} (1/2)g''(\mu)Z^2 \sum_{h=-\infty}^{\infty} \gamma_X(h)$$

follows, with $Z \sim \mathcal{N}(0,1)$, as $n \to \infty$.

The Moving Block Bootstrap does not work for S_n. The reason for this failure is the fact that although $g'(\mu) = 0$ we typically have $g'(\overline{X}_n) \neq 0$ and the convergence $g'(\overline{X}_n) \xrightarrow{P} g'(\mu) = 0$ as $n \to \infty$ is not fast enough.

The subsampling procedure proposes to calculate S_n on subsamples of length b, $b < n$, of the time series and to approximate the distribution of the random variable S_n by the empirical distribution of the same random variable calculated over all subsamples.

To elaborate, let

$$\overline{X}_{t,b} = \frac{1}{b}\sum_{s=1}^{b} X_{t+s-1}$$

be the sample mean of the subsample $X_t, X_{t+1}, \ldots, X_{t+b-1}$, then

$$b \cdot (g(\overline{X}_{t,b}) - g(\overline{X}_n))$$

is the analogue of S_n based on the subsample.

Clearly there are $n - b + 1$ such subsamples for which this analogue can be calculated.

The subsampling procedure proposes to approximate for every $x \in \mathbb{R}$ the distribution function $P(S_n \leq x)$ by the empirical distribution function

$$\widehat{F}_{sub,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbb{1}_{(-\infty,x]}\big(b \cdot (g(\overline{X}_{t,b}) - g(\overline{X}_n))\big).$$

We now proceed with a more general discussion of the subsampling procedure. Suppose that we are interested in the distribution of

$$S_n = r_n \cdot (T_n - \theta),$$

where $(r_n, n \in \mathbb{N})$ is some positive sequence increasing to infinity as n increases and $T_n = T(X_1, X_2, \ldots . X_n)$ is a statistic of interest.

The only requirement we will impose on S_n, and which constitutes the universality of subsampling, is the following.

$(*)$ S_n converges weakly to a random variable S as $n \to \infty$.

S does not have to be Gaussian!

To estimate the distribution function $P(S_n \leq x),\ x \in \mathbb{R}$, the subsampling proposal suggests to use the empirical distribution function

$$F_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbb{1}_{(-\infty,x]}\big(r_b \cdot (T_{t,b} - T_n)\big),$$

where r_b is the same normalizing sequence as r_n but adapted to the subsample size b.

The rationale behind this proposal is that $Y_{t,b}$ is for any $t \in \{1, 2, \ldots, n-b+1\}$ a time series of length b stemming from $\{X_t, t \in \mathbb{Z}\}$.

Hence all $r_b \cdot (T_{t,b} - \theta),\ t = 1, 2, \ldots, n-b+1$, have the same distribution which differ from that of $r_n(T_n - \theta)$ solely by the fact that the latter is based on a time series of length n while the former is based on a time series of length b.

One can show the following result.

Theorem 24

Suppose that $\{X_t, t \in \mathbb{Z}\}$ *is a strictly stationary process, that assumption* $(*)$ *applies. and that* $r_b/r_n \to 0$, $b/n \to 0$ *and* $b \to \infty$ *as* $n \to \infty$.
For $x \in \mathbb{R}$ *consider* $S_{t,b}(x) = \mathbb{1}_{(-\infty,x]}\big(r_b \cdot (T_{t,b} - \theta)\big)$ *and assume that*

$$\frac{1}{n-b+1} \sum_{h=1}^{n-b+1} \big|\mathrm{Cov}(S_{1,b}(x), S_{1+h,b}(x))\big| \to 0,$$

as $n \to \infty$.

Then, for any continuity point $x \in \mathbb{R}$ *of* $P(S \leq x)$, *we have that*

$$F_{n,b}(x) \xrightarrow{P} P(S \leq x), \text{ as } n \to \infty.$$

For m-dependent time series, the covariance condition of the above theorem is obviously satisfied. In more general cases, assumptions about the dependence structure of the time series are required to ensure this.

Second-Order Properties of Bootstrap

A well-known result on the rate of convergence of $P\{\sqrt{n}(\overline{X}_n - \mu)/\sigma \leq x\}$ towards the cumulative distribution function $\Phi(x)$ of the standard normal distribution is given by the **Theorem of Berry-Esseen**.
This result states that for such i.i.d. random variables under the additional assumption $\mathrm{E}|X_1|^3 < \infty$ the following bound for all $n \in \mathbb{N}$ applies,

$$\sup_{x \in \mathbb{R}} \left| P\{\sqrt{n}(\overline{X}_n - \mu)/\sigma \leq x\} - \Phi(x) \right| \leq \frac{C \cdot \mathrm{E}|X_1 - \mu|^3}{\sigma^3 \sqrt{n}}.$$

Here C is a universal constant which does not depend on the distribution of the random variables X_j. It is known that $C < 0.5$.

We will see in the following that under some mild assumptions the bootstrap can improve the above Berry-Esseen rate of $n^{-1/2}$.

To elaborate: The characteristic function φ_n of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ can be easily calculated, namely

$$\varphi_n(t) = \big(\varphi_Z(t/\sqrt{n})\big)^n,$$

where φ_Z denotes the characteristic function of $Z = (X_1 - \mu)/\sigma$.

A formal power series expansion leads because of because $d^j\varphi_Z(t)/dt^j|_{t=0} = \mathrm{i}^j\,\mathrm{E}Z^j$, $j \in \mathbb{N}_0$, to

$$\log(\varphi_Z(t)) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s+1}\Big(-\frac{1}{2}t^2 + \sum_{j=3}^{\infty}\frac{(\mathrm{i}t)^j}{j!}\mathrm{E}(Z^j)\Big)^{s+1}.$$

For the last identity recall that $\mathrm{E}Z = 0$ and $\mathrm{E}Z^2 = 1$.

By rearranging the terms according to powers of t, the following result is obtained for appropriately chosen coefficients $\kappa_r, r \in \mathbb{N}$,

$$\log(\varphi_Z(t)) = \sum_{r=2}^{\infty}\kappa_r\frac{(\mathrm{i}t)^r}{r!}.$$

For a random variable Z the function $\log(\varphi_Z)$ is denoted the *cumulant generating function* and the coefficients κ_r are referred to as the *cumulants* of Z of the r-th order .

A careful comparison of the last two equations gives for standardized Z

$$\kappa_1 = 0,\ \kappa_2 = 1,\ \kappa_3 = \mathrm{E}Z^3,\ \kappa_4 = \mathrm{E}Z^4 - 3,\ \kappa_5 = \mathrm{E}Z^5 - 10\,\mathrm{E}Z^3, \ldots .$$

For the characteristic function φ_n of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ this leads to

$$\begin{aligned}\varphi_n(t) &= \exp\left(n \log \varphi_Z(t/\sqrt{n})\right) \\ &= \mathrm{e}^{-t^2/2}\left(1 + \sum_{j=1}^{k-2} n^{-j/2} r_j(\mathrm{i}t)\right)\left(1 + o(n^{-(k-2)/2})\right),\end{aligned}$$

with suitably chosen polynomials $r_j(z)$.

It turns out that the $r_j(z)$ are of degree $3j$ and that they depend on the cumulants $\kappa_3, \kappa_4, \ldots$, but they do not depend on n.

A careful inspection gives for instance,

$$r_1 = \frac{1}{6}\kappa_3 z^3 \quad \text{and} \quad r_2 = \frac{1}{24}\kappa_4 z^4 + \frac{1}{72}\kappa_3 z^6.$$

By inversion of characteristic functions (cf. Theorem 6.2.1 in Chung (2001)) we finally obtain for $L_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ under the simplifying assumptions that $P\{L_n = a\} = P\{L_n = b\} = 0, a < b$, and $\int |\varphi_n(t)|dt < \infty$, that,

$$P\{L_n \in (a, b]\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\varphi_n(t)dt.$$

This leads to the so-called *Edgeworth expansion* of L_n

$$\begin{aligned} P\{L_n \in (a, b]\} =& \sum_{j=1}^{k-2} n^{-j/2} \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\mathrm{e}^{-t^2/2} r_j(\mathrm{i}t)dt + o(n^{-(k-2)/2}) \\ =& \Phi(b) - \Phi(a) + \sum_{j=1}^{k-2} n^{-j/2}\phi(x)p_j(x)\Big|_a^b + o(n^{-(k-2)/2}), \end{aligned}$$

where the p_j are polynomials of degree $3j - 1$.

A rigorous statement with proof of such an Edgeworth expansion is given as a special case in Theorem 2.2 of Hall (1992).
Besides the assumption that $\mathrm{E}|X|^k < \infty$ a regularity condition is

$$\limsup_{|t|\to\infty} |\varphi_X(t)| < 1\,,$$

where φ_X denotes the characteristic function of X. This assumption, called *Cramér's condition*, applies if the distribution of X possesses a Lebesgue density.
For the case $k \geq 4$ one obtains

$$P\{\sqrt{n}(\overline{X}_n - \mu)/\sigma \leq x\} = \Phi(x) - n^{-1/2}\frac{\kappa_3}{6}(x^2-1)\phi(x) + \mathcal{O}(n^{-1}),$$

where $\kappa_3 = \mathrm{E}Z^3 = \mathrm{E}(X_1-\mu)^3/\sigma^3$.

$n^{-1/2}\frac{\kappa_3}{6}(x^2-1)\phi(x)$: First-order correction term after $\Phi(x)$.

Bootstrap typically implicitly uses $\sqrt{n}$-consistent estimators of third or even higher order moments. If an Edgeworth expansion also would hold for the bootstrap distribution of the standardized sample mean, a comparison of both Edgeworth expansions will lead to an approximation rate of the bootstrap of order $\mathcal{O}(n^{-1})$ in P_n-probability.

This means that the bootstrap would outperform the asymptotic normal distribution, a property which is coined as *second-order efficiency* of the bootstrap.

But it is important to note that such an effect only can happen if *standardized* statistics are considered for bootstrap proposals!

It can also be shown that the same effect appears when bootstrapping *studentized* statistics. We refer to the monograph Hall (1992).