The Bootstrap in Signal Processing

Abdelhak M Zoubir

Signal Processing Group
Technische Universität Darmstadt
E-Mail: zoubir@spg.tu-darmstadt.de
URL: www.spg.tu-darmstadt.de

Motivation

- Let the measurements $x = \{x_1, x_2, \dots, x_n\}$ be realisations of the random sample $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$, drawn from some *unspecified* F_θ.
- Let $\hat{\theta}(X)$ be an estimator of some parameter θ of F_θ, which could be the mean $(\mu = \theta)$ of F_θ.
- The goal is to find (statistical) characteristics of $\hat{\theta}(X)$. For example, $\hat{\theta}(X)$ is the average value of X, $\hat{\theta} = \hat{\mu} = 1/n \cdot \sum_{i=1}^{n} X_i$.

- If F_θ is *known* and $\hat{\theta}$ is computed from $X_1, X_2, \dots, X_n$, using a smooth function $s(X)$ that is relatively *simple*, one could *exactly* evaluate the distribution of $\hat{\theta}(X)$, otherwise approximate it for $n \to \infty$.

What if F_θ is unknown and asymptotic theory does not apply?

Contents

TECHNISCHE UNIVERSITÄT DARMSTADT

- Introduction
  - What can I use the bootstrap for?
  - The bootstrap idea
  - History of the bootstrap
- The independent data bootstrap
  - An example for independent data
  - An example of independent data bootstrap for dependent data
- The bootstrap principle for dependent data
  - The moving blocks bootstrap
- A real-life example of the bootstrap
  - Micro-Doppler Analysis
- Concluding remarks

Introduction (Cont'd)

- The term *bootstrap* is often associated with The Baron von Münchhausen.

Courtesy of Prof. Karl Heinrich Hofmann, TU Darmstadt and former Head of Group *Algebra, Geometry and Functional Analysis*

Introduction (Cont'd)

TECHNISCHE UNIVERSITÄT DARMSTADT

- This analogy may suggest that the bootstrap is able to perform the impossible.
- The bootstrap is not a magic technique that provides a panacea for all statistical inference problems!
- The bootstrap is a powerful tool that can substitute tedious or often impossible theoretical derivations with computational calculations [Efron (1979)].
- There are situations where the bootstrap fails and care is required [Young (1994)].

What Can I Use the Bootstrap for?

TECHNISCHE UNIVERSITÄT DARMSTADT

- *The bootstrap is a computational tool for statistical inference*.
- It can be used for:
  - Estimation of statistical characteristics such as *bias*, *variance*, *distribution function of estimators* and thus *confidence intervals*.
  - Hypothesis tests, for example for *signal detection*, and
  - *model selection*.
- When can I use the bootstrap?

When I know little about the statistics of the data and/or I have only a few data so that asymptotic theory does not hold.

Problem Statement

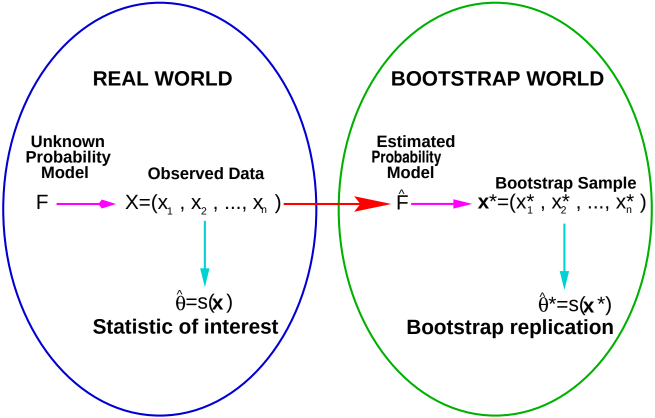
TECHNISCHE UNIVERSITÄT DARMSTADT

- Let the measurements $x = \{x_1, x_2, \dots, x_T\}$ be realisations of the random sample $\mathcal{X} = \{X_1, X_2, \dots, X_T\}$, drawn from some *unspecified F_θ*.
- Let $\hat{\theta}(X)$ be an estimator of some parameter θ of F_θ, which could be the spectral density $C_{XX}(\omega)$ of the stationary process X_t, $t = 0, \pm 1, \pm 2, \dots$.
- The goal is to find (statistical) characteristics of $\hat{\theta}(X)$.

- If F_θ is *known* and $\hat{\theta}$ is computed from $X_1, X_2, \dots, X_T$, using a smooth function $s(X)$ that is relatively *simple*, one could *exactly* evaluate the distribution of $\hat{\theta}(X)$.
- Alternatively, one may approximate the distribution of $\hat{\theta}(X)$, assuming $T \to \infty$.

If F_θ is unknown and asymptotic theory does not apply, the bootstrap provides the answer!

The Bootstrap Idea



Basic idea: simulate the probability mechanism of the real world by substituting the unknowns with estimates derived from the data.

A Brief History Recount

- It was by marrying the power of *Monte Carlo approximation* with an exceptionally broad view of the sort of problems that bootstrap methods might solve that Efron (1979) vaulted earlier resampling ideas into the realm of a universal statistical methodology.
- Arguably, the prehistory of the bootstrap encompasses pre-1979 developments of Monte Carlo methods for sampling.
- Important aspects of bootstrap roots lie in methods for spatial sampling in India in the 1920's [Hubback (1946), Mahalanobis (1946)], even before Quenouille's and Tukey's work on the jackknife [Quenouille (1949, 1956), Tukey (1958)].

The independent data bootstrap

The Bootstrap for Independent Data

Step 1. Conduct the experiment to obtain the random sample $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ and find the estimator $\hat{\theta}$ from $\mathcal{X}$.

Step 2. Construct the empirical distribution $\hat{F}_\theta$, which puts equal mass $1/n$ at each observation $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$.

Step 3. From the selected $\hat{F}_\theta$, draw a sample $\mathcal{X}^* = \{X_1^*, X_2^*, \dots, X_n^*\}$, called the bootstrap (re)sample.

Step 4. Approximate the distribution of $\hat{\theta}$ by the distribution of $\hat{\theta}^*$ derived from $\mathcal{X}^*$.

Remark: In the i.i.d. case, $\hat{F}_\theta$ does not need to be explicitly computed (it is the data itself) and bootstrap resamples are drawn as follows.

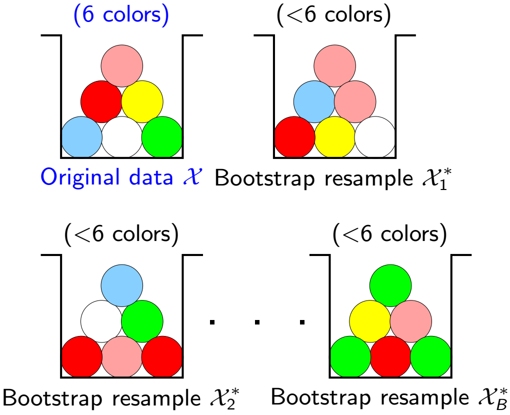
Resampling i.i.d. Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- A resample $\mathcal{X}^* = \{X_1^*, X_2^*, \dots, X_n^*\}$ is an unordered collection of n sample points drawn randomly from $\mathcal{X}$ with replacement, so that each X_i^* has probability n^{-1} of being equal to any one of the X_j's.
- That is, the X_i^*'s are i.i.d., conditional on the random sample $\mathcal{X}$, with this distribution [Hall (1992)].

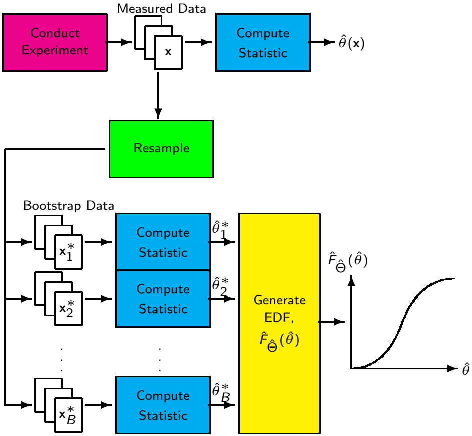
This means that $\mathcal{X}^*$ is likely to contain repeats. As an example, consider $n = 4$ and the collection $\mathcal{X}^* = \{0.5, -3.7, -3.7, 2.8\}$ which should not be mistaken for the set $\{0.5, -3.7, 2.8\}$ because the bootstrap sample has one repeat. Also, $\mathcal{X}^*$ is the same as $\{0.5, -3.7, 2.8, -3.7\}$, $\{-3.7, 0.5, 2.8, -3.7\}$, etc. because the order of elements in the resample plays no role [Hall (1992), Efron & Tibshirani (1993)].

Resampling i.i.d. Data



Implementation of the Bootstrap Procedure

TECHNISCHE UNIVERSITÄT DARMSTADT



Algorithm to compute the approximative distribution function of the estimator $\hat{\theta}$.

Example: Variance Estimation

Consider the problem of finding the variance $\sigma^2_{\hat{\theta}}$ of an estimator $\hat{\theta}$ of θ, based on the random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ from the unknown distribution F_θ.

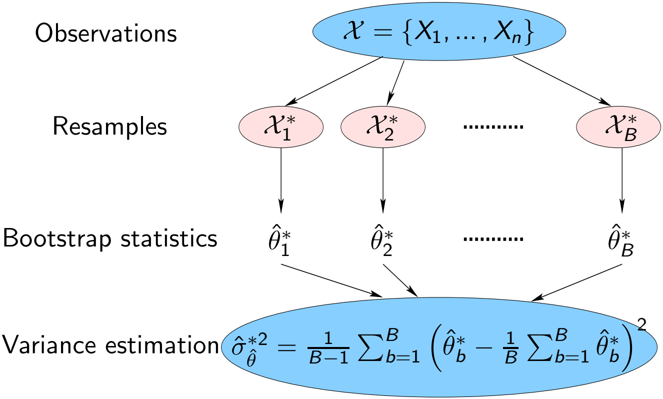
- If tractable, one may derive an analytic expression for $\sigma^2_{\hat{\theta}}$.
- For example, for $X_1, \ldots, X_n$ i.i.d. and $\theta = \mu$,

$$\boxed{\hat{\theta} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \sigma^2_{\hat{\mu}} = \frac{\sigma^2}{n}.}$$

- In general, the situation is more complex and one may use asymptotic arguments [Serfling (1980)] to compute an estimate $\hat{\sigma}^2_{\hat{\theta}}$ for $\sigma^2_{\hat{\theta}}$.
- The bootstrap can provide the answer in the finite sample size case.

Example: Estimation of the Variance of $\hat{\theta}$

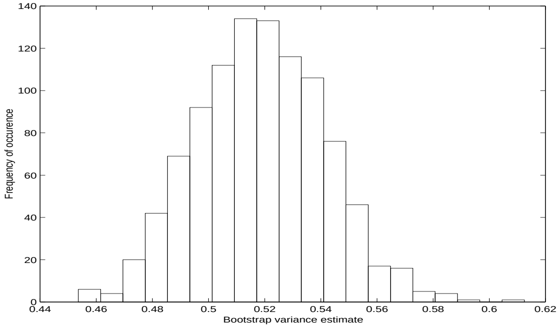
TECHNISCHE UNIVERSITÄT DARMSTADT



Example: Variance Estimation (Cont'd)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Estimate $\sigma_{\hat{\mu}}$ based on a random sample $\mathcal{X}$ of size $n = 50$ from $\mathcal{N}(10, 25)$.
- With $B = 25$, we find $\hat{\sigma}^2_{\hat{\mu}} = 0.49$ vs. the true $\sigma^2_{\hat{\mu}} = 0.5$.



Histogram of $\hat{\sigma}^{*2(1)}_{\hat{\mu}}, \hat{\sigma}^{*2(2)}_{\hat{\mu}}, \ldots, \hat{\sigma}^{*2(1000)}_{\hat{\mu}}$, based on $\mathcal{N}(10, 25)$, $n = 50$ and $B = 25$.

The independent data bootstrap for dependent data

An Example: Variance Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

Assume T observations x_t, $t = 1, \dots, T$, from

$$X_t + a \cdot X_{t-1} = Z_t \,,$$

where Z_t is i.i.d. noise with $\mathrm{E}[Z_t] = 0$, $c_{ZZ}(u) = \sigma_Z^2 \delta(u)$ and a such that $|a| < 1$.

Goal: Estimate the variance of $\hat{a}$.

An estimate for the variance of $\hat{a}$ exists asymptotically and for Z_t Gaussian:

- Detrend the data and fit the AR(1) model to x_t, $t = 1, \dots, T$.
- With $\hat{c}_{xx}(u) = 1/T \sum_{t=1}^{T-|u|} x_t x_{t+|u|}$ for $1 \leq |u| \leq T$, calculate the MLE of a, $\hat{a} = -\hat{c}_{xx}(1)/\hat{c}_{xx}(0)$, which has approximate variance $\hat{\sigma}_{\hat{a}}^2 = (1 - a^2)/T$.
- Under some regularity conditions an asymptotic formula for $\hat{\sigma}_{\hat{a}}^2$ can be found in the non-Gaussian case [Porat & Friedlander (1989)].

Example: The Bootstrap Variant

TECHNISCHE UNIVERSITÄT DARMSTADT

Step 1. *Calculate residuals*. With an estimate $\hat{a}$ of a, define the residuals $\hat{z}_t = x_t + \hat{a} \cdot x_{t-1}$ for $t = 2, \dots, T$.

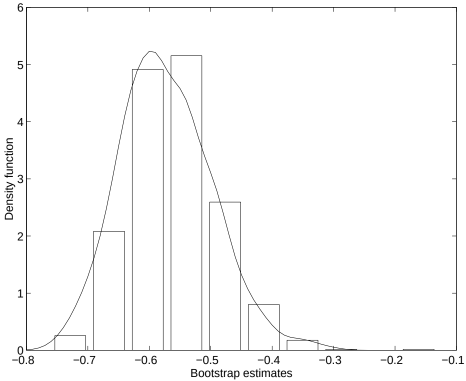
Step 2. *Resampling*. Create a bootstrap sample $x_1^*, x_2^*, \dots, x_T^*$ by sampling $\hat{z}_2^*, \hat{z}_3^*, \dots, \hat{z}_T^*$, with replacement, from the residuals $\hat{z}_2, \hat{z}_3, \dots, \hat{z}_T$, then letting $x_1^* = x_1$, and $x_t^* = -\hat{a}x_{t-1}^* + \hat{z}_t^*$, $t = 2, \dots, T$.

Step 3. *Calculate bootstrap estimates*. Centre the time series $x_1^*, x_2^*, \dots, x_T^*$, and compute $\hat{a}^*$ in the same way $\hat{a}$ was obtained but based on $x_1^*, x_2^*, \dots, x_T^*$.

Step 4. *Repetition*. Repeat Steps 2–3 B times, to obtain $\hat{a}_1^*, \hat{a}_2^*, \dots, \hat{a}_B^*$.

Step 5. *Variance estimation*. From $\hat{a}_1^*, \hat{a}_2^*, \dots, \hat{a}_B^*$, approximate the variance of $\hat{a}$ by $\hat{\sigma}_{\hat{a}}^{*2}$.

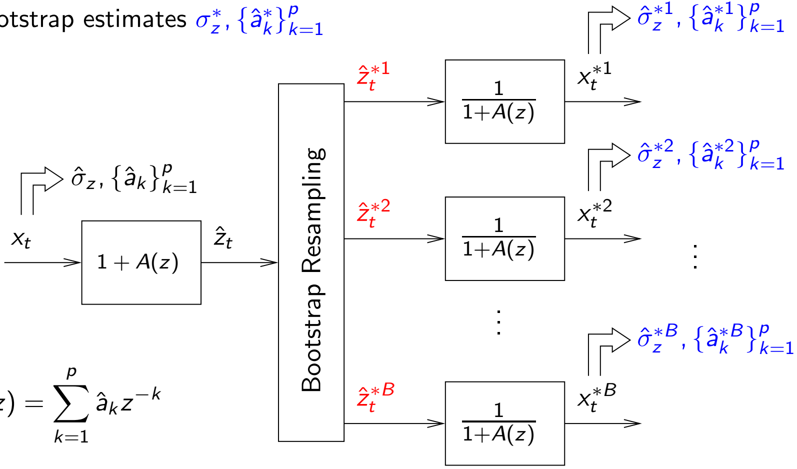
Example (Cont'd)



Histogram of $\hat{a}_1^*, \hat{a}_2^*, \ldots, \hat{a}_{1000}^*$ for $a = -0.6$, $T = 128$ and Z_t Gaussian. The MLE for a is $\hat{a} = -0.6351$ and $\hat{\sigma}_{\hat{a}} = 0.0707$. The bootstrap estimate is $\hat{\sigma}_{\hat{a}}^* = 0.0712$ as compared to $\hat{\sigma}_{\hat{a}} = 0.0694$ based on 1000 Monte Carlo simulations.

Bootstrap Resampling for AR(p) Models

- Resampling the residuals: $\hat{z}_t^*$
- Bootstrap estimates σ_z^*, $\{\hat{a}_k^*\}_{k=1}^p$



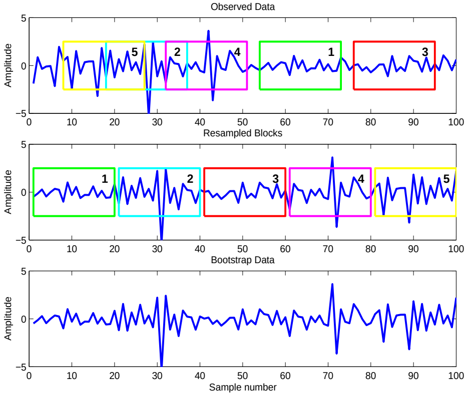
$$A(z) = \sum_{k=1}^{p} \hat{a}_k z^{-k}$$

The dependent data bootstrap

The Dependent Data Bootstrap

- If no plausible model such as AR is available, we could make the assumption of weak dependence.
- Strong mixing processes, i.e., *loosely speaking, processes for which observations far apart (in time) are almost independent* [Rosenblatt (1985)], for example, satisfy the weak dependence condition.
- The *moving blocks* bootstrap [Künsch (1989), Liu & Singh (1992), Politis & Romano (1992,1994)] has been proposed for weakly dependent data.
- Other variants of the moving blocks bootstrap exist. They include the *circular block bootstrap* [Politis & Romano (1992), Shao & Yu (1993)], the *stationary block bootstrap*[Politis & Romano (1994)] and the *taper block bootstrap* [Paparoditis & Politis (2001)].

The Moving Blocks Bootstrap



Randomly select blocks of the original data (top) and concatenate them together (centre) to form a resample (bottom). Here, the block size is $l = 20$ and $n = 100$.

TECHNISCHE UNIVERSITÄT DARMSTADT

A real-life example of the bootstrap

Micro-Doppler Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

- Doppler often arises in engineering applications due to the relative motion of an object with respect to the measurement system.
- If the motion is harmonic, for example due to vibration or rotation, the resulting signal can be well modeled by an FM process [Huang *et al.* (1990)] whose instantaneous angular frequency is

$$\boxed{\omega_i(t;\beta) = D\sin(\omega_m t + \phi)\,.}$$

- Estimation of the FM parameters may allow one to determine physical properties such as the angular velocity and displacement of the vibrational/rotational motion which can in turn be used for classification.

Objective: Estimate the FM parameters (D, ω_m, ϕ) along with a measure of accuracy, such as confidence intervals.

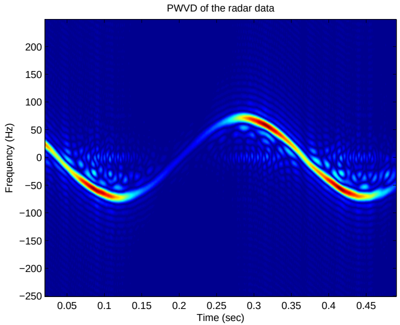
Micro-Doppler Analysis (Cont'd)

TECHNISCHE UNIVERSITÄT DARMSTADT

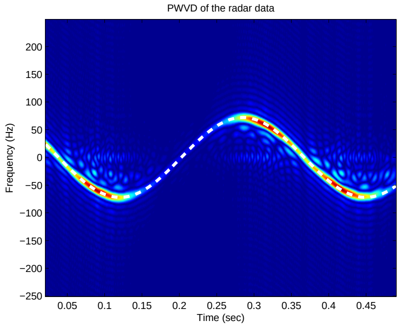
- The results shown here are based on an experimental radar system*, operating at carrier frequency $f_c = 919.82$ MHz.
- After demodulation, the in-phase and quadrature baseband channels are sampled at $f_s = 1$ kHz.
- The radar system is directed toward a spherical object, swinging with a pendulum motion, which results in a typical micro-Doppler signature.
- The estimation of the phase parameters is performed using a time-frequency Hough transform (TFHT) [Barbarossa & Lemoine (1996), Cirillo, Zoubir & Amin (2006)].
- The Pseudo Wigner-Ville Distribution (PWVD) of the observations is computed and shown below.

* Courtesy of Prof. Moeness Amin, Director of the Center of Advanced Communications, Villanova University, Philadelphia, USA

Micro-Doppler Analysis (Cont'd)



(a)

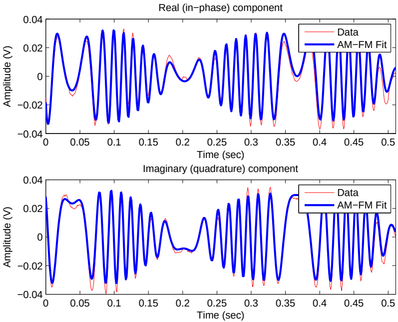


(b)

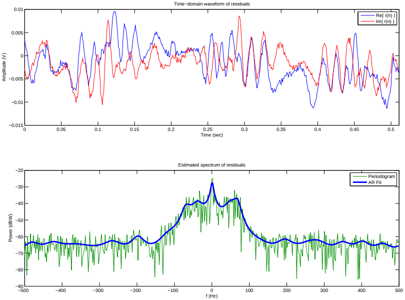
The PWVD of the radar data (a). The PWVD of the radar data and the micro-Doppler signature estimated using the TFHT (b).

Micro-Doppler Analysis (Cont'd)

TECHNISCHE
UNIVERSITÄT
DARMSTADT



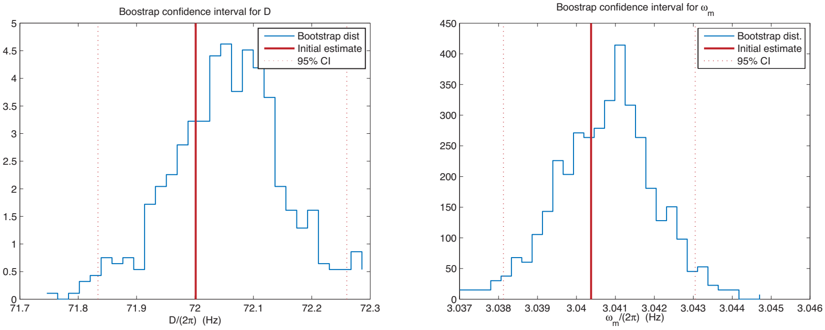
(c)



(d)

The real and imaginary components of the radar signal are with their estimated counterparts (c). The real and imaginary parts of the residuals and their spectral estimates (d).

Micro-Doppler Analysis (Cont'd)



(a) (b)

The bootstrap distributions and 95% confidence intervals for the FM parameters D (a) and ω_m (b) using $B = 500$.

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

- Many signal processing problems require the computation of quality measures for estimators.
- Most techniques available assume that the size of the available set of sample values is sufficiently large, so that "asymptotic" results can be applied.
- In many signal processing problems this assumption cannot be made, for example, because the process is non-stationary and only small portions of stationary data are considered.
- Bootstrap techniques are an alternative to asymptotic methods and provide good results for many practical problems.
- The current challenge in bootstrap technology is the application of robust techniques (measurements in the presence of outliers) in conjunction with the bootstrap.

The Baron von Münchhausen

Varney, Allen (2007), "The Extraordinary Adventures of Baron Münchhausen", in Lowder, James, Hobby Games: The Best 100, Green Ronin Publishing,

pp. 107109, ISBN 978-1-932442-96-0