Bootstraping Spectra

Momchil Nikolov

SIGNAL PROCESSING GROUP

1. Gutachten: Dr.-Ing. Michael Fauss
2. Gutachten: Prof. Dr.-Ing. A.M. Zoubir
zoubir@spg.tu-darmstadt.de

Outline

Motivation

Spectrum Estimation

Bootstrapping Method

Bootstrap Methods Applied in Signal Processing

Test of Frequency and Time Domain Bootstrap Methods

Conclusion

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

The are many different approaches to spectrum estimation in Signal Processing. The most common technique used for solving this problem is the asymptotic one, but this project seminar explores one method, that can be useful in case:

- the quantity of data is not abundant,
- or the data is non-Gaussian;

The Bootstrap approach

This method uses a re-sampling technique, that can generate numerous new samples from just one realization of a random process.

Outline

Spectral Estimation Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

The main goal of the spectral estimation problem is captured by the following informal expression:

Estimation Problem

From a finite record of a stationary data sequence, estimate how the total power is distributed over frequency

Signal

TECHNISCHE UNIVERSITÄT DARMSTADT

Let $X(t)$ be a signal, with properties:

- $X(t) \in \mathbb{R}$
- $t = 0, 1, 2, \ldots, T$
- assume $X(t)$ has finite energy, $\sum_{t=-\infty}^{\infty} |X(t)|^2 < \infty$.
- possesses a *discrete-time Fourier transform* (DTFT)
  $X(\omega) = \sum_{t=-\infty}^{\infty} X(t) e^{-j\omega t}$
- an energy spectral density: $C(\omega) = |X(\omega)|^2$.

True Spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

We will now assume that $X(t)$:

- is a sequence of random variables,
- with zero mean, $\mathbb{E}[X(t)] = 0$,
- covariance function, $c_{XX}(\tau) = E[X_0 X_\tau]$
- which implies that $X(t)$ is a second-order stationary sequence(WSS)

Let $X(t)$ have true spectral density:

$$C_{XX}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c_{XX}(\tau) e^{-j\omega\tau}, -\pi < \omega < \pi,$$

Kernel Estimate Spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

Given observations $x_1, \ldots, x_T$ we construct an estimator of the true spectrum $C_{XX}(\omega)$, as:

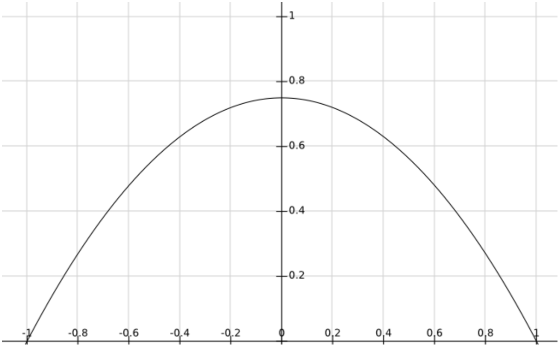
$$\hat{C}_{XX}(\omega; h) = \frac{1}{T \cdot h} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_{XX}(\omega_k)$$

I_{XX} denotes the periodogram, which is defined as:

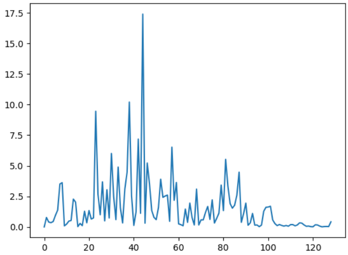
$$I_{XX}(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} X(t) e^{-j\omega t} \right|^2$$

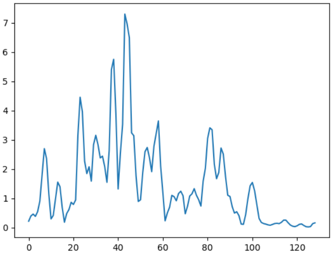
and is a candidate estimator for the spectrum $C_{XX}(\omega)$.

Kernel function



Smoothing Example





Mathematical Representation of Spectral Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

The spectral estimation problem can now be stated more formally as follows:

Estimation Problem

From a finite-length record $X_1, \ldots, X_T$ of a second-order stationary random process, determine an estimate $\hat{C}_{XX}(\omega)$ of its power spectral density $C_{XX}(\omega)$, for $\omega = [-\pi, \pi]$

Outline

Bootstrapping Method

Suppose that we have:

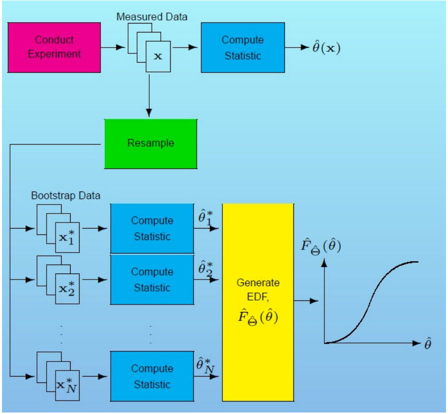
- a set of measured realizations $x = \{x_1, x_2, ..., x_n\}$,
- drawn from some unspecified distribution F_X
- task to construct estimator $\hat{\theta}$ of some parameter θ of F_X, e.g. mean;

Problem: Data is not Gaussian and realizations insufficient for asymptotic approach.

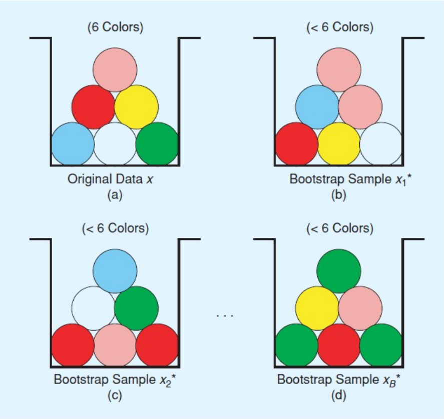
Bootstrap Solution

Substitution of the unknown distribution F_X by the empirical distribution of the data $\hat{F}_X$ with reuse of the original data through re-sampling.

Bootstrapping Procedure



Re-sampling



(6 Colors)
Original Data x
(a)
(< 6 Colors)
Bootstrap Sample x_1^*
(b)
(< 6 Colors)
Bootstrap Sample x_2^*
(c)
. . .
(< 6 Colors)
Bootstrap Sample x_B^*
(d)

Outline

Motivation

Spectrum Estimation

Bootstrapping Method

Bootstrap Methods Applied in Signal Processing
- Frequency domain residual based bootstrap
- Time domain tapered block bootstrap

Test of Frequency and Time Domain Bootstrap Methods

Conclusion

Frequency domain residual based bootstrap

Step 1

TECHNISCHE UNIVERSITÄT DARMSTADT

The residual based bootstrap procedure.

Step 1. *Centering.* Center $X_1, \ldots, X_T$ by subtracting the sample mean.

Step 2. *Initial Estimate.* Compute the periodogram according to Eq. (4). With an initial bandwidth $h_i > 0$, calculate $\hat{C}_{XX}(\omega; h_i)$ according to Eq. (3).

Step 3. *Compute and Rescale Residuals.* Calculate the residuals

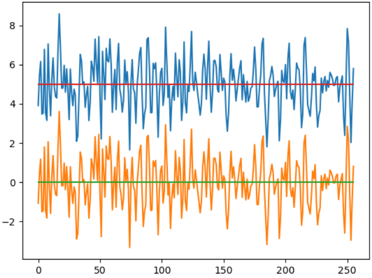
$$\hat{\varepsilon}_k = \frac{I_{XX}(\omega_k)}{\hat{C}_{XX}(\omega_k; h_i)}, \quad k = 1, \ldots, N$$

and rescale them to obtain

$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\varepsilon_{\cdot}}, \quad k = 1, \ldots, N, \quad \varepsilon_{\cdot} = \frac{1}{N}\sum_{k=1}^{N} \hat{\varepsilon}_k.$$

Step 4. *Bootstrap Residuals.* Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_N^*$ from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 1: Centering X



Frequency domain residual based bootstrap

Step 2

The residual based bootstrap procedure.

Step 1. *Centering.* Center $X_1, \ldots, X_T$ by subtracting the sample mean.

Step 2. *Initial Estimate.* Compute the periodogram according to Eq. (4). With an initial bandwidth $h_i > 0$, calculate $\hat{C}_{XX}(\omega; h_i)$ according to Eq. (3).

Step 3. *Compute and Rescale Residuals.* Calculate the residuals

$$\hat{\varepsilon}_k = \frac{I_{XX}(\omega_k)}{\hat{C}_{XX}(\omega_k; h_i)}, \quad k = 1, \ldots, N$$

and rescale them to obtain

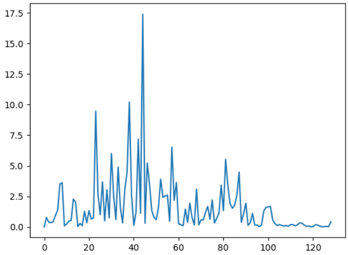
$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\varepsilon.}, \quad k = 1, \ldots, N, \quad \varepsilon. = \frac{1}{N}\sum_{k=1}^{N} \hat{\varepsilon}_k.$$

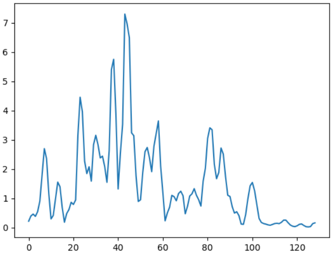
Step 4. *Bootstrap Residuals.* Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_N^*$ from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 2:

- $I_{XX}(\omega) = \dfrac{1}{N}\left|\sum_{t=1}^{N} X(t) e^{-j\omega t}\right|^2$
- $\hat{C}_{XX}(\omega; h_i) = \dfrac{1}{T \cdot h}\sum_{k=-N}^{N} K\left(\dfrac{\omega - \omega_k}{h}\right) I_{XX}(\omega_k)$

Periodogram and Kernel Estimate





Frequency domain residual based bootstrap

Step 3

TECHNISCHE
UNIVERSITÄT
DARMSTADT

The residual based bootstrap procedure.

Step 1. *Centering.* Center $X_1, \ldots, X_T$ by subtracting the sample mean.

Step 2. *Initial Estimate.* Compute the periodogram according to Eq. (4). With an initial bandwidth $h_i > 0$, calculate $\hat{C}_{XX}(\omega; h_i)$ according to Eq. (3).

Step 3. *Compute and Rescale Residuals.* Calculate the residuals

$$\hat{\varepsilon}_k = \frac{I_{XX}(\omega_k)}{\hat{C}_{XX}(\omega_k; h_i)}, \quad k = 1, \ldots, N$$

and rescale them to obtain

$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\varepsilon.}, \quad k = 1, \ldots, N, \quad \varepsilon. = \frac{1}{N}\sum_{k=1}^{N} \hat{\varepsilon}_k.$$

Step 4. *Bootstrap Residuals.* Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_N^*$ from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 3:

- compute residuals $\hat{\varepsilon}_k$
- normalize $\hat{\varepsilon}_k$, by division with the residual mean $\varepsilon = \frac{1}{N}\sum_{t=1}^{N} \hat{\varepsilon}_k$
- re-scaled real valued residuals $\tilde{\varepsilon}_k$ are approximately i.i.d. for N sufficiently large

Frequency domain residual based bootstrap

Step 4

The residual based bootstrap procedure.

Step 1. *Centering.* Center $X_1, \ldots, X_T$ by subtracting the sample mean.

Step 2. *Initial Estimate.* Compute the periodogram according to Eq. (4). With an initial bandwidth $h_i > 0$, calculate $\hat{C}_{XX}(\omega; h_i)$ according to Eq. (3).

Step 3. *Compute and Rescale Residuals.* Calculate the residuals

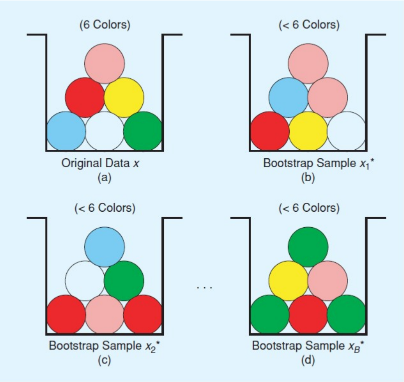
$$\hat{\varepsilon}_k = \frac{I_{XX}(\omega_k)}{\hat{C}_{XX}(\omega_k; h_i)}, \quad k = 1, \ldots, N$$

and rescale them to obtain

$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\bar{\varepsilon}_{\cdot}}, \quad k = 1, \ldots, N, \quad \bar{\varepsilon}_{\cdot} = \frac{1}{N} \sum_{k=1}^{N} \hat{\varepsilon}_k.$$

Step 4. *Bootstrap Residuals.* Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_N^*$ from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 4: Draw $\tilde{\varepsilon}_k^*$ from $\tilde{\varepsilon}_k$



SPG

Frequency domain residual based bootstrap

Step 5

Step 5. *Bootstrap Estimate*. Choose a resampling kernel bandwidth g and define the bootstrap periodogram as
$I^*_{XX}(\omega_k) = I^*_{XX}(-\omega_k) = \hat{C}_{XX}(\omega_k; g)\tilde{\varepsilon}^*_k,$

setting $I^*_{XX}(0) = 0$. Then find the bootstrap kernel spectral density estimate
$\hat{C}^*_{XX}(\omega; h; g) = \frac{1}{T \cdot h} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I^*_{XX}(\omega_k).$

Step6. *Repetition*. Repeat Steps 4 and 5 a large number of times.

Step 7. *Confidence Interval Estimation*. With an arbitrary α find c^*_U such that

$$\text{Prob}^*\left(\sqrt{T \cdot h}\,\frac{\hat{C}^*_{XX}(\omega; h; g) - \hat{C}_{XX}(\omega; g)}{\hat{C}_{XX}(\omega; g)} \le c^*_U\right) = \frac{\alpha}{2}$$

and set the upper bound of the $100(1-\alpha)\%$ confidence interval for $C_{XX}(\omega)$ as $\hat{C}_{XX}(\omega; h)/(1 + c^*_U(T \cdot h)^{-1/2})$. Herein, $\text{Prob}^*(\cdot)$ denotes probability conditioned on the measured data. Analogously using a c^*_L, compute the lower bound of the confidence interval.

Step 5:

- multiply $\tilde{\varepsilon}^*$ with our initial spectral estimate $\hat{C}_{XX}(\omega_k)$ in order to get the bootstrap periodogram $I^*{}_{XX}(\omega_k)$
- set the $I^*{}_{XX}(0) = 0$
- estimate the spectrum $\hat{C}^*{}_{XX}$ with $I^*{}_{XX}(\omega_k)$

Frequency domain residual based bootstrap

Step 7

Step 5. *Bootstrap Estimate*. Choose a resampling kernel bandwidth g and define the bootstrap periodogram as

$$I^*_{XX}(\omega_k) = I^*_{XX}(-\omega_k) = \hat{C}_{XX}(\omega_k; g)\tilde{\varepsilon}^*_k,$$

setting $I^*_{XX}(0) = 0$. Then find the bootstrap kernel spectral density estimate

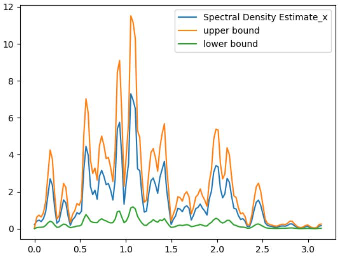
$$\hat{C}^*_{XX}(\omega; h; g) = \frac{1}{T \cdot h} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I^*_{XX}(\omega_k).$$

Step6. *Repetition*. Repeat Steps 4 and 5 a large number of times.

Step 7. *Confidence Interval Estimation*. With an arbitrary α find c^*_U such that

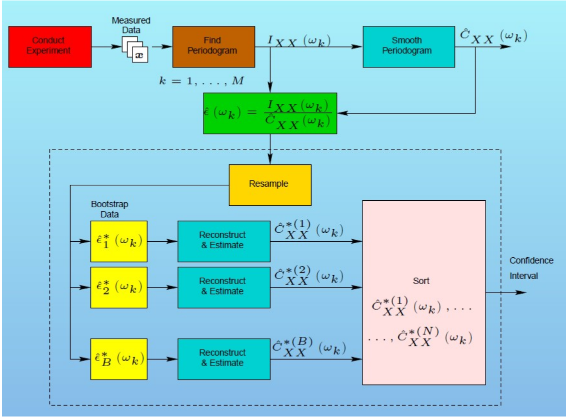
$$\text{Prob}^*\left(\sqrt{T \cdot h}\,\frac{\hat{C}^*_{XX}(\omega; h; g) - \hat{C}_{XX}(\omega; g)}{\hat{C}_{XX}(\omega; g)} \leq c^*_U\right) = \frac{\alpha}{2}$$

and set the upper bound of the $100(1-\alpha)\%$ confidence interval for $C_{XX}(\omega)$ as $\hat{C}_{XX}(\omega; h)/(1 + c^*_U(T \cdot h)^{-1/2})$. Herein, $\text{Prob}^*(\cdot)$ denotes probability conditioned on the measured data. Analogously using a c^*_L, compute the lower bound of the confidence interval.



Frequency domain residual based bootstrap

Full process



Outline

Time domain tapered block bootstrap

Step 1 & Step 2

The tapered block bootstrap procedure.

Step 1.	*Centering*. Center $X_1, \ldots, X_T$ by removing the sample mean.
Step 2.	*Spectrum Estimation*. Calculate the periodogram $I_{XX}(\omega_k)$ from $X_1, \ldots, X_T$ according to Eq. (4) and the corresponding kernel spectral density estimate $\hat{C}_{XX}(\omega_k)$, $k = 1, \ldots, N$, as in Eq. (3).
Step 3.	*Set the Bootstrap Parameters*. Choose a block size b, where b is a positive integer less than T and define $k = \lfloor T/b \rfloor$ as the number of blocks to draw from, where $\lfloor \cdot \rfloor$ denotes the integer part.
Step 4.	*Draw Bootstrap Resamples*. Let the integers $i_0, i_1, \ldots, i_{k-1}$ be drawn independently and identically distributed with distribution uniform on the set $\{1, 2, \ldots, T-b+1\}$. Then for $m = 0, 1, \ldots, k-1$, compute $X^*_{mb+j} = w_b(j) \frac{\sqrt{b}}{\Vert w_b \Vert_2} X_{i_m+j-1}, \quad j = 1, 2, \ldots, b,$

Step 1 & Step 2 the same as the frequency domain residual-based bootstrap

Time domain tapered block bootstrap

Step 3

The tapered block bootstrap procedure.

Step 1. *Centering*. Center $X_1, \ldots, X_T$ by removing the sample mean.

Step 2. *Spectrum Estimation*. Calculate the periodogram $I_{XX}(\omega_k)$ from $X_1, \ldots, X_T$ according to Eq. (4) and the corresponding kernel spectral density estimate $\hat{C}_{XX}(\omega_k)$, $k = 1, \ldots, N$, as in Eq. (3).

Step 3. *Set the Bootstrap Parameters*. Choose a block size b, where b is a positive integer less than T and define $k = \lfloor T/b \rfloor$ as the number of blocks to draw from, where $\lfloor \cdot \rfloor$ denotes the integer part.

Step 4. *Draw Bootstrap Resamples*. Let the integers $i_0, i_1, \ldots, i_{k-1}$ be drawn independently and identically distributed with distribution uniform on the set $\{1, 2, \ldots, T-b+1\}$. Then for $m = 0, 1, \ldots, k-1$, compute

$$X^*_{mb+j} = w_b(j) \frac{\sqrt{b}}{\|w_b\|_2} X_{i_m+j-1}, \quad j = 1, 2, \ldots, b,$$

Step 3:

- Choose block size b
- define number of blocks to draw as $k = \lfloor T/b \rfloor$
- Note: in most cases $kb \neq T \Rightarrow$ draw one more block and truncate

Time domain tapered block bootstrap

Step 4

The tapered block bootstrap procedure.

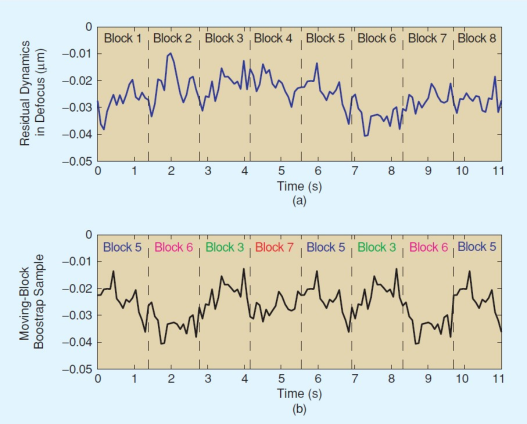
Step 1.	*Centering*. Center $X_1, \ldots, X_T$ by removing the sample mean.
Step 2.	*Spectrum Estimation*. Calculate the periodogram $I_{XX}(\omega_k)$ from $X_1, \ldots, X_T$ according to Eq. (4) and the corresponding kernel spectral density estimate $\hat{C}_{XX}(\omega_k)$, $k = 1, \ldots, N$, as in Eq. (3).
Step 3.	*Set the Bootstrap Parameters*. Choose a block size b, where b is a positive integer less than T and define $k = \lfloor T/b \rfloor$ as the number of blocks to draw from, where $\lfloor \cdot \rfloor$ denotes the integer part.
Step 4.	*Draw Bootstrap Resamples*. Let the integers $i_0, i_1, \ldots, i_{k-1}$ be drawn independently and identically distributed with distribution uniform on the set $\{1, 2, \ldots, T-b+1\}$. Then for $m = 0, 1, \ldots, k-1$, compute $X^*_{mb+j} = w_b(j) \frac{\sqrt{b}}{\lVert w_b \rVert_2} X_{i_m+j-1}, \quad j = 1, 2, \ldots, b,$

Step 4:

- Choose integers i_k as starting point of drawn block
- make sure $i_{max} = T - b + 1$
- Glue blocks together to form bootstrap sample X^* of length l
- truncate, to secure $l = kb = T$

Time domain tapered block bootstrap

Example Block Draw



Time domain tapered block bootstrap

Step 5

TECHNISCHE UNIVERSITÄT DARMSTADT

which yields the pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$ of size $l = k \cdot b$. Herein w_b denotes the tapering window defined in Eq. (9) and $\|w_b\|_2 = (\sum_{t=1}^{b} w_b(t)^2)^{1/2}$.

Step 5. *Construct Bootstrap Estimates*. Center the pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$ and find the bootstrap periodogram $I_{XX}^*(\omega_k)$ for $-N^* < k < N^*$, computed as in Eq. (4), replacing T by l and $X_1, \ldots, X_T$ by $X_1^*, X_2^*, \ldots, X_l^*$, where $N^* = \lfloor l/2 \rfloor$ and $I_{XX}^*(0) = 0$. Then find the corresponding bootstrap kernel spectral density estimate $\hat{C}_{XX}^*(\omega)$ as in Eq. (3), with l replacing T.

Step 6. *Repetition*. Repeat Steps 4 and 5 a large number of times.

Step 7. *Confidence Interval Estimation*. Find confidence bounds as in Step 7 of Table 1, replacing T by l.

Step 5:

- center X^*
- compute $I^*{}_{XX}(\omega_k)$
- set the $I^*{}_{XX}(0) = 0$
- estimate the spectrum $\hat{C}^*{}_{XX}$ with $I^*{}_{XX}(\omega_k)$

Outline

Test Model

TECHNISCHE UNIVERSITÄT DARMSTADT

We will use a AR(5) model:

$$X(t) = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + N_t,$$

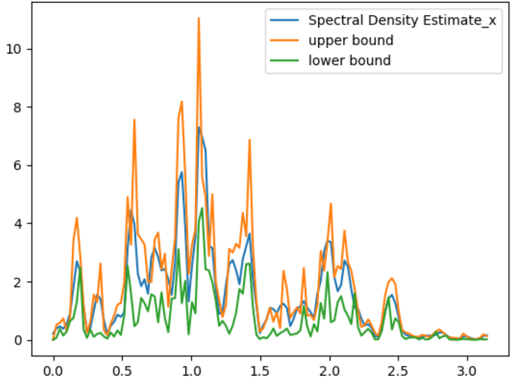
with:

- N_t are independently and uniformly distributed variables on the interval $[0, 1]$
- sample size of $T = (256, 128)$

Outline

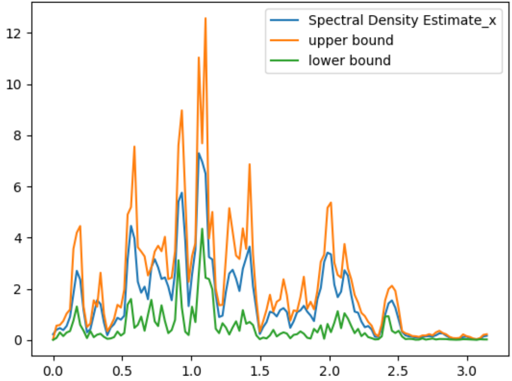
Frequency Domain Residual based Bootstrap

$T = 256$, $B = 3$



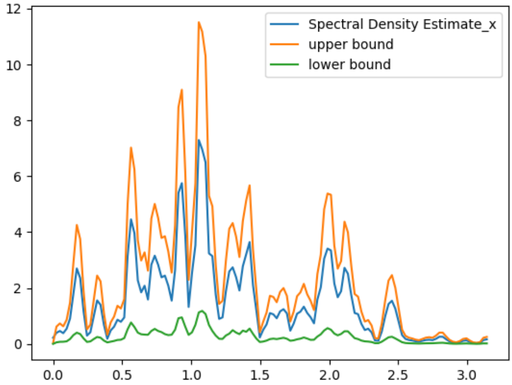
Frequency Domain Residual based Bootstrap

$T = 256$, $B = 6$



Frequency Domain Residual based Bootstrap

$T = 256$, $B = 1000$



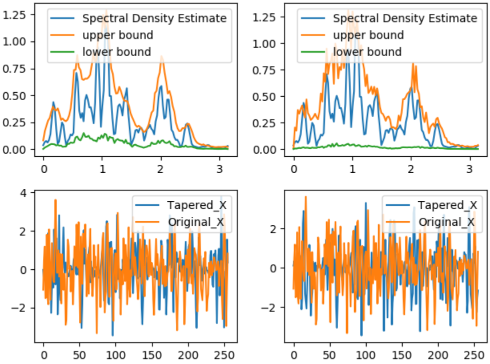
Confidence interval

Frequency Domain Residual based Bootstrap

T = 256						
bootstrap samples	1	3	6	10	100	1000
coverage probability in %	50	83.24	95.33	98.81	99.51	99.61
T = 128						
bootstrap samples	6	50	55	60	100	1000
coverage probability in %	69.38	94.0	95.84	96.76	98.61	99.23

Tabelle: Frequency Domain Residual based Bootstrap: Coverage probability for different number of bootstrap samples and length of X_t

Time domain tapered block bootstrap



Time domain tapered block bootstrap

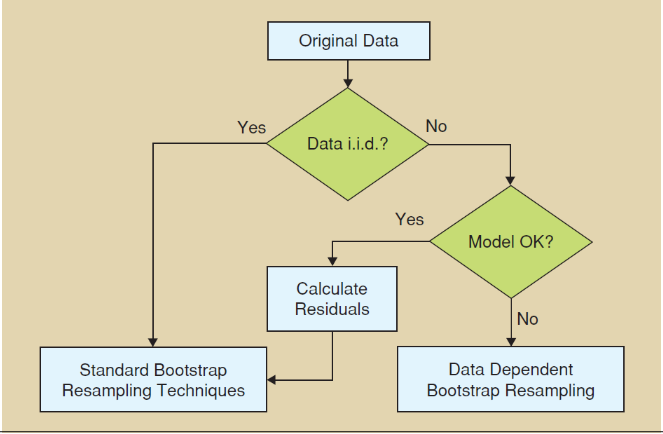
T = 256						
bootstrap samples	1	3	6	10	100	1000
coverage probability in %	50	67.32	74.88	80.71	96.67	99.32
T = 128						
bootstrap samples	6	11	16	21	46	1000
coverage probability in %	89.61	93.40	94.98	95.90	97.57	99.15

Tabelle: Time domain tapered block bootstrap: Coverage probability for different number of bootstrap samples and length of X_t

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

Conclusion



Original Data
Data i.i.d.?
Yes
No
Model OK?
Yes
No
Calculate Residuals
Standard Bootstrap Resampling Techniques
Data Dependent Bootstrap Resampling

Thank you for your attention!