The Bootstrap: A Powerful Tool for Statistical Signal Processing with Small Sample Sets

A.M. Zoubir*
Australian Telecommunications Research Institute
School of Electrical & Computer Engineering
Curtin University of Technology
GPO Box U 1987, Perth, WA 6001, Australia
e-mail: zoubir@ieee.org
URL: www.atri.curtin.edu.au/csp

*Formerly with the Queensland University of Technology

Outline

Outline (Cont'd)

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimates
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

Outline (Cont'd)

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Outline (Cont'd)

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Introduction

- Most techniques for computing variances of parameter estimators or for setting confidence intervals assume a large sample size [Bhattacharya & Rao (1976)].
- In many signal processing problems large sample methods are inapplicable [Fisher & Hall (1991)].
- The bootstrap was introduced [Efron (1979)] to calculate confidence intervals for parameters when few data are available.
- The bootstrap has been shown to solve many other problems which would be too complicated for traditional statistical analysis [Hall (1992), Efron & Tibshirani (1993), Shao & Tu (1995)].

Introduction (Cont'd)

- The bootstrap does with a computer what the experimenter would do in practice, if it were possible.
  1. The observations are randomly re-assigned, and the estimates re-computed.
  2. These assignments and re-computations are done many times and treated as repeated experiments.
- In an era of exponentially increasing computational power, computer-intensive methods such as the bootstrap are affordable.

Introduction (Cont'd)

Applications of the bootstrap to real-life problems have been reported in (see also special session at ICASSP-94 [64])

- radar signal processing [Nagaoka & Amai (1990,1991)],
- sonar signal processing [Krolik *et al.* (1991), Böhme & Maiwald (1994), Reid *et al.* (1996)],
- geophysics [Fisher & Hall (1989,1990,1991), Tauxe *et al.* (1991)],
- biomedical engineering [Haynor & Woods (1989), Banga & Ghorbel (1993)]
- control [Lai & Chen (1995)],

Introduction (Cont'd)

- atmospheric environmental research [Hanna (1989)],
- vibration analysis [Zoubir & Böhme (1991,1995)].
- power systems [Herman (1996)],
- computer vision [Lange *et al.* (1998), Kanatani & Ohta (1998)]
- image analysis [Archer & Chan (1996)],
- nuclear technology [Yacout *et al.* (1996)],
- metrology [Ciarlini (1997), Cox *et al.* (1997)],
- financial engineering [Bhar & Chiarella (1996), Ankenbrand & Tomassini (1996)].

Introduction (Cont'd)

- Bootstrap methods are potentially superior to large sample techniques for small sample sizes [Hall (1992)].
- A danger exists when applying bootstrap techniques in some circumstances where standard approaches are judged inappropriate and in such circumstances the bootstrap may also fail [Freedman (1981)].
- **Special care** is therefore required when applying the bootstrap in real-life situations [Young (1994)].

Problem Statement

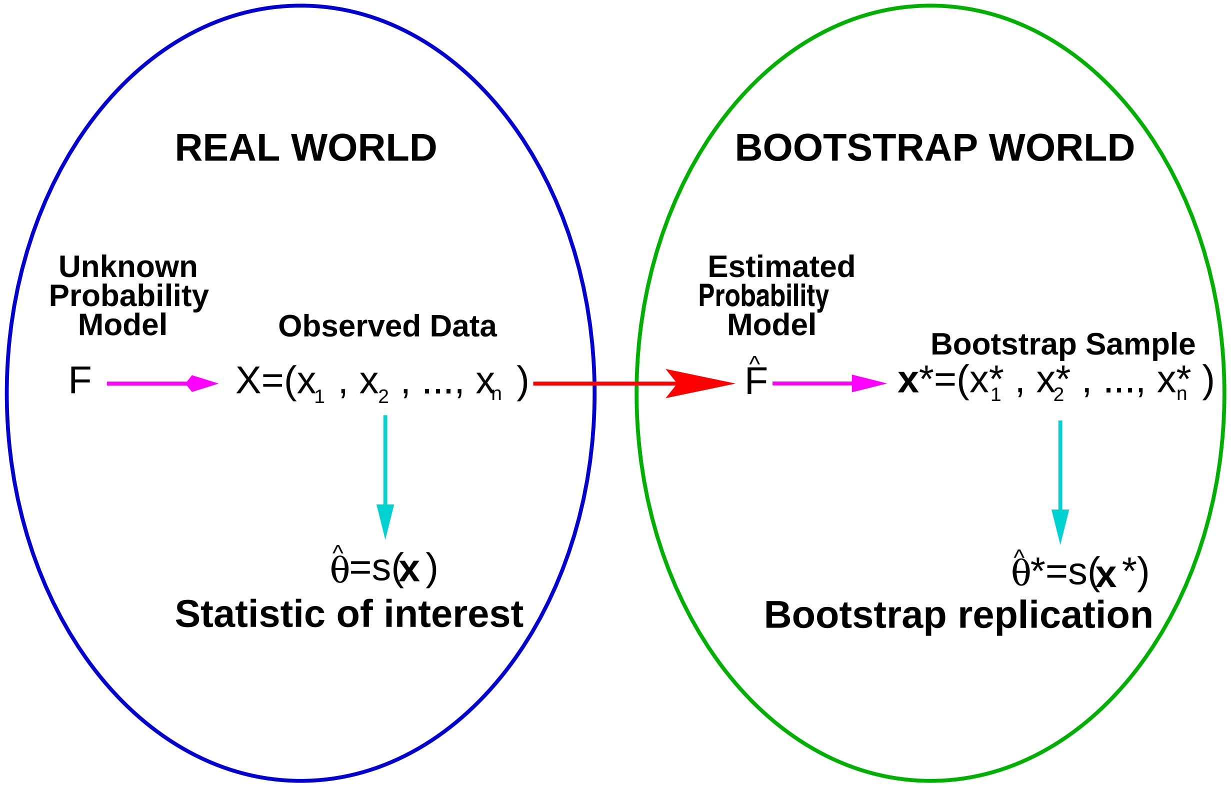
Let $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ be a random sample from *a completely unspecified distribution* F. Let θ denote an unknown characteristic of F, such as its mean or variance.

Find the distribution of $\hat{\theta}$, an estimator of θ, derived from the sample $\mathcal{X}$.

Possible solution: repeat the experiment a sufficient number of times and approximate the distribution of $\hat{\theta}$ by the so obtained empirical distribution.

Problem: may be inapplicable for cost reasons or because the experimental conditions are not reproducible.

Principle of the Bootstrap



From [Efron & Tibshirani (1993)].

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

The Non-Parametric Bootstrap

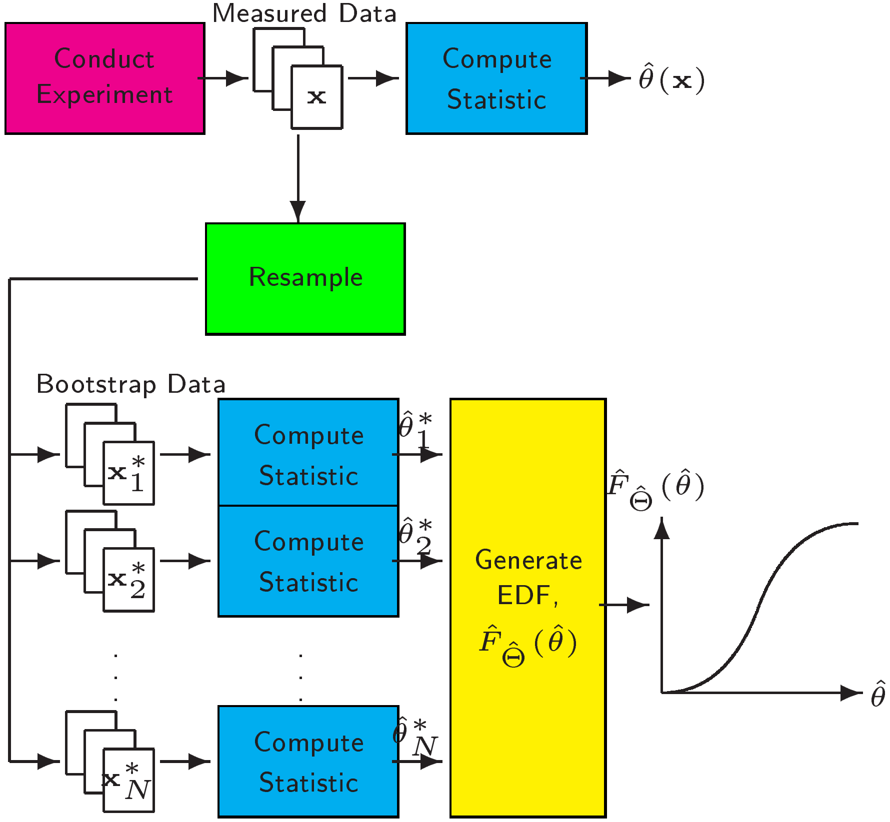
1. Conduct the experiment to obtain the random sample $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ and calculate the estimate $\hat{\theta}$ from $\mathcal{X}$.

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

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

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

The Bootstrap Procedure



Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

Example: Bias Estimation

Consider the problem of estimating the variance of an unknown distribution $F_{\mu,\sigma}$, based on the random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$. Two different estimators can be used:

$$\hat{\sigma}_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 , \quad \hat{\sigma}_b^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 .$$

It can be easily shown that

$$\mathsf{E}\hat{\sigma}_u^2 = \sigma^2 \quad \text{and} \quad \mathsf{E}\hat{\sigma}_b^2 = (1 - \frac{1}{n})\sigma^2$$

With the bootstrap we estimate the bias $b(\hat{\sigma}^2) = \mathsf{E}\hat{\sigma}^2 - \sigma^2$ by $\mathsf{E}_*\hat{\sigma}^{*2} - \hat{\sigma}^2$, where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2, i.e., $\hat{\sigma}_b^2$ and E_* is expectation w.r.t. bootstrap sampling.

Example (Cont'd)

Step 0. *Experiment.* Collect the data into $\mathcal{X} = \{X_1, \ldots, X_n\}$. Compute the estimates $\hat{\sigma}_u^2$ and $\hat{\sigma}_b^2$.

Step 1. *Resampling.* Draw a random sample of size n, with replacement, from $\mathcal{X}$.

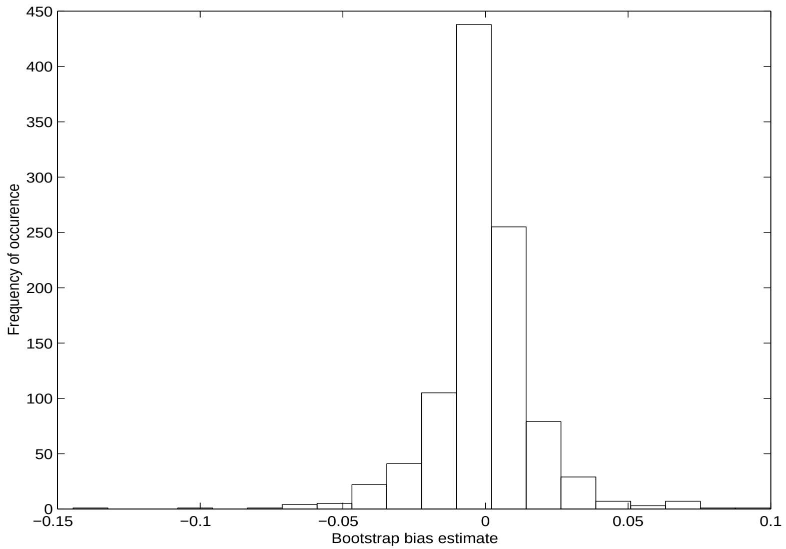
Step 2. *Calculation of the bootstrap estimate.* Calculate the bootstrap estimates $\hat{\sigma}_u^{*2}$ and $\hat{\sigma}_b^{*2}$ from $\mathcal{X}^*$ in the same way $\hat{\sigma}_u^2$ and $\hat{\sigma}_b^2$ were computed but with the resample $\mathcal{X}^*$.

Step 3. *Repetition.* Repeat Steps 1 and 2 to obtain a total of N bootstrap estimates $\hat{\sigma}_{u,1}^{*2}, \ldots, \hat{\sigma}_{u,N}^{*2}$ and $\hat{\sigma}_{b,1}^{*2}, \ldots, \hat{\sigma}_{b,N}^{*2}$.

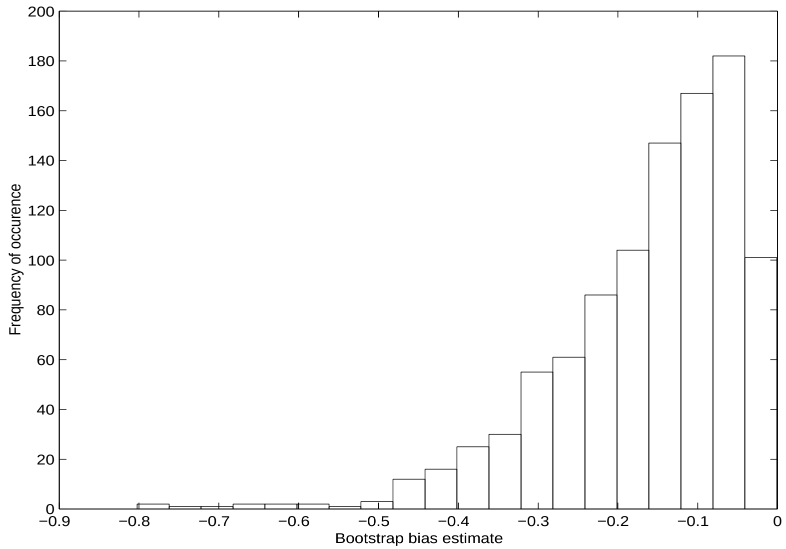
Step 4. *Bias Estimation.* Estimate $b(\hat{\sigma}_u^2)$ by $b_*(\hat{\sigma}_u^{*2}) = 1/N \sum_{i=1}^{N} \hat{\sigma}_{u,i}^{*2} - \hat{\sigma}_b^2$ and $b(\hat{\sigma}_b^2)$ by $b_*(\hat{\sigma}_b^{*2}) = 1/N \sum_{i=1}^{N} \hat{\sigma}_{b,i}^{*2} - \hat{\sigma}_b^2$.

Example (Cont'd)

We considered a sample of size $n = 5$ from the standard normal distribution. With $N = 999$ and 1000 Monte Carlo simulations, we obtained the following histograms for $b_*(\hat{\sigma}_u^{*2})$ and $b_*(\hat{\sigma}_b^{*2})$.



Bootstrap estimation of $b(\hat{\sigma}_u^2)$,
sample mean: -4.9400e-04 (=0 !).



Bootstrap estimation of $b(\hat{\sigma}_b^2)$,
sample mean: -0.1573 (=-0.2 !).

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

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 $\sigma^2_{\hat{\theta}}$.
- Alternatively, one may use asymptotic arguments to compute an estimate $\hat{\sigma}^2_{\hat{\theta}}$ for $\sigma^2_{\hat{\theta}}$.

Problem: In many situations the conditions for the above are not fulfilled.

Solution: The bootstrap provides a simple and accurate alternative to approximate $\sigma^2_{\hat{\theta}}$ by $\hat{\sigma}^{*2}_{\hat{\theta}}$.

Example (Cont'd)

Step 0. *Experiment.* Conduct the experiment and collect the random data into the sample $\mathcal{X} = \{X_1, \dots, X_n\}$.

Step 1. *Resampling.* Draw a random sample of size n, with replacement, from $\mathcal{X}$.

Step 2. *Calculation of the bootstrap estimate.* Evaluate the bootstrap estimate $\hat{\theta}^*$ from $\mathcal{X}^*$ calculated in the same manner as $\hat{\theta}$ but with the resample $\mathcal{X}^*$ replacing $\mathcal{X}$.

Step 3. *Repetition.* Repeat Steps 1 and 2 many times to obtain a total of B bootstrap estimates $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$. Typical values for B are between 25 and 200 [Efron & Tibshirani (1993)].

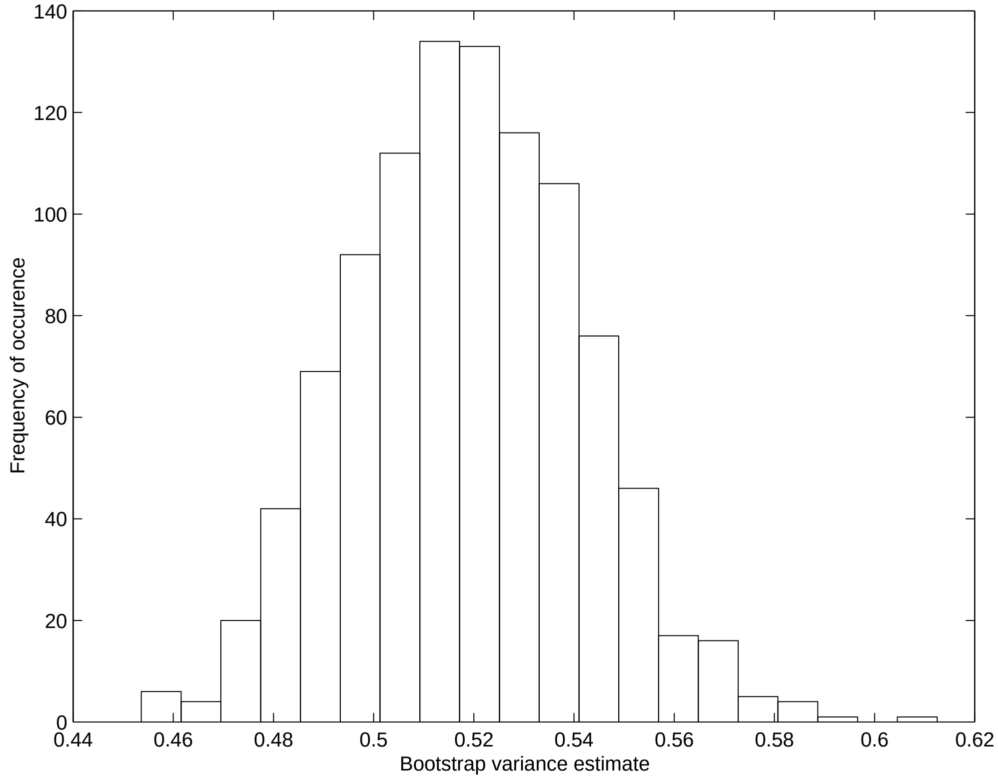
Example (Cont'd)

Step 4. *Estimation of the variance of* $\hat{\theta}$. Estimate the variance $\sigma^2_{\hat{\theta}}$ of $\hat{\theta}$ by

$$\hat{\sigma}^2_{\mathsf{BOOT}} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^*_b - \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*_b \right)^2 .$$

- Suppose $F_{\mu,\sigma}$ is $\mathcal{N}(10, 25)$ and we wish to estimate $\sigma_{\hat{\mu}}$ based on a random sample $\mathcal{X}$ of size $n = 50$.
- Following the above procedure with $B = 25$, a bootstrap estimate of the variance of $\hat{\mu}$ is found to be $\hat{\sigma}^2_{\mathsf{BOOT}} = 0.49$ as compared to the true $\sigma^2_{\hat{\mu}} = 0.5$.

Example (Cont'd)



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

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. **Confidence Interval Estimation for the Mean**
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

Example: Confidence Interval for the Mean

Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be from some unknown distribution $F_{\mu,\sigma}$. We wish to find an estimator and a $100(1-\alpha)\%$ interval for μ. Let

$$\hat{\mu} = \frac{X_1 + \ldots + X_n}{n}.$$

A confidence interval for μ is found by determining the distribution of $\hat{\mu}$, and finding $\hat{\mu}_L, \hat{\mu}_U$ such that

$$\mathsf{Pr}(\hat{\mu}_L \leq \mu \leq \hat{\mu}_U) = 1 - \alpha.$$

The distribution of $\hat{\mu}$ depends on the distribution of the X_i's, which is *unknown*. If n is large, the distribution of $\hat{\mu}$ could be approximated by the normal distribution as per the central limit theorem. *What if n is small?*

Example (Cont'd)

Step 0. *Experiment.* Conduct the experiment and collect $X_1, \ldots, X_n$ into $\mathcal{X}$. Suppose $F_{\mu,\sigma}$ is $\mathcal{N}(10, 25)$ and $\mathcal{X} = \{-2.41, 4.86, 6.06, 9.11, 10.20, 12.81, 13.17, 14.10, 15.77, 15.79\}$ is of size $n = 10$. The mean of all values in $\mathcal{X}$ is $\hat{\mu} = 9.95$.

Step 1. *Resampling.* Draw a sample of 10 values, with replacement, from $\mathcal{X}$. One might obtain the *bootstrap resample* $\mathcal{X}^* = \{9.11, 9.11, 6.06, 13.17, 10.20, -2.41, 4.86, 12.81, -2.41, 4.86\}$. Note that some values from the original sample appear more than once while others do not appear at all.

Step 2. *Calculation of the bootstrap estimate.* Calculate the mean of $\mathcal{X}^*$. The mean of all 10 values in $\mathcal{X}^*$ is $\hat{\mu}_1^* = 6.54$.

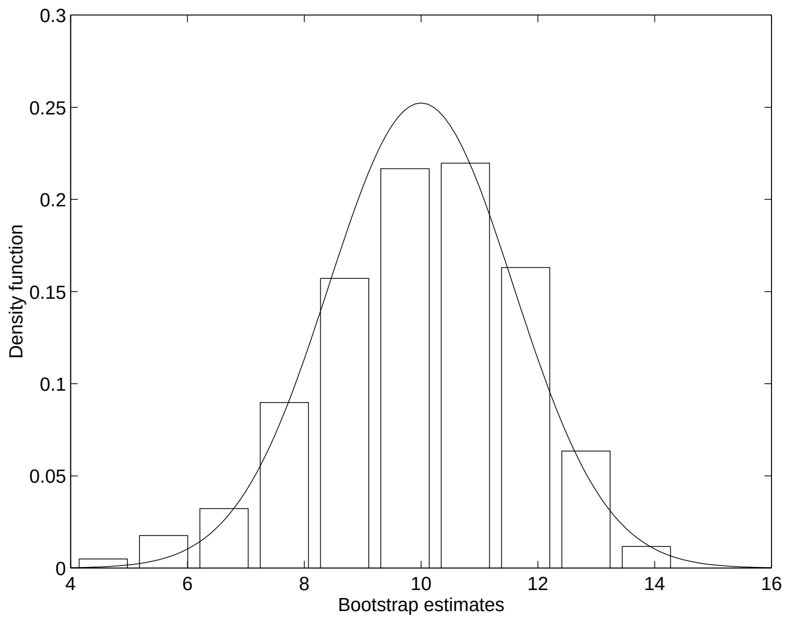
Example (Cont'd)

Step 3. *Repetition.* Repeat Steps 1 and 2 to obtain N bootstrap estimates $\hat{\mu}_1^*, \ldots, \hat{\mu}_N^*$. Let $N = 1000$.

Step 4. *Approximation of the Distribution of $\hat{\mu}$.* Sort the bootstrap estimates to obtain $\hat{\mu}_{(1)}^* \leq \hat{\mu}_{(2)}^* \leq \ldots \leq \hat{\mu}_{(N)}^*$. We might get 3.48, 3.39, 4.46, … , 8.86, 8.88, 8.89, … , 10.07, 10.08, … , 14.46, 14.53, 14.66.

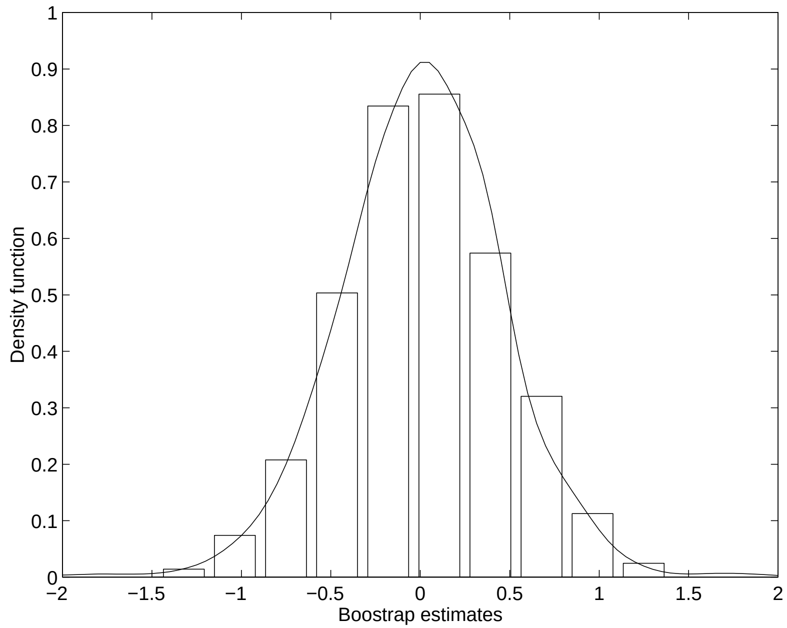
Step 5. *Confidence Interval.* The $100(1-\alpha)\%$ bootstrap confidence interval is $(\hat{\mu}_{(q_1)}^*, \hat{\mu}_{(q_2)}^*)$, where $q_1 = \lfloor N\alpha/2 \rfloor$ and $q_2 = N - q_1 + 1$. For $\alpha = 0.05$ and $N = 1000$, $q_1 = 25$ and $q_2 = 976$, and the 95% confidence interval is found to be $(\mathbf{6.27}, \mathbf{13.19})$ as compared to the theoretical $(\mathbf{6.85}, \mathbf{13.05})$.

Example (Cont'd)



Histogram of $\hat{\mu}_1^*, \hat{\mu}_2^*, \ldots, \hat{\mu}_{1000}^*$, based on the random sample $\mathcal{X} = \{-2.41,$ $4.86,$ $6.06,$ $9.11,$ $10.20,$ $12.81,$ $13.17,$ $14.10,$ $15.77,$ $15.79\}$, together with the density function of a Gaussian variable with mean 10 and variance 2.5.

Example (Cont'd)



Histogram of 1000 bootstrap estimates of the mean of the t_4-distribution and the kernel probability density function obtained from 1000 Monte Carlo simulations. The 95 % confidence interval is (**-0.896,0.902**) and (**-0.886,0.887**) based on the bootstrap and Monte Carlo, respectively.

Example (Cont'd)

- The procedure described above can be substantially improved because the interval calculated is, in fact, an interval with coverage less than the nominal value [Hall (1988)].
- Later, we shall discuss another way that will lead to a more accurate confidence interval for the mean.
- The computational expense to calculate the confidence interval for μ is approximately N times greater than the one needed to compute $\hat{\mu}$.
- This is acceptable given the ever-increasing capabilities of today's computers.

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - **The Parametric Bootstrap**
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

The Parametric Bootstrap

- Bootstrap sampling can be carried out *parametrically*.
- If one has partial knowledge of F, one may use $\hat{F}_{\hat{\theta}}$ instead of $\hat{F}$.
- Draw N samples of size n from the parametric estimate of F

$$\hat{F}_{\hat{\theta}} \longrightarrow (x_1^*, x_2^*, \dots, x_n^*)$$

  and proceed as before.
- When used in a parametric way, the bootstrap provides more accurate answers, provided the model is correct.

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

Example: Confidence Interval for the Mean

Step 0. *Experiment.* Conduct the experiment and collect $X_1, \ldots, X_n$ into $\mathcal{X}$. Suppose $F_{\mu,\sigma}$ is $\mathcal{N}(10, 25)$ and $\mathcal{X} = \{-2.41, 4.86, 6.06, 9.11, 10.20, 12.81, 13.17, 14.10, 15.77, 15.79\}$ is of size $n = 10$. The mean of all values in $\mathcal{X}$ is $\hat{\mu} = 9.95$ and the sample variance is $\hat{\sigma}^2 = 33.15$.

Step 1. *Resampling.* Draw a sample of 10 values, with replacement, from $F_{\hat{\mu},\hat{\sigma}}$. We might obtain $\mathcal{X}^* = \{7.45, 0.36, 10.67, 11.60, 3.34, 16.80, 16.79, 9.73, 11.83, 10.95\}$.

Step 2. *Calculation of the bootstrap estimate.* Calculate the mean of $\mathcal{X}^*$. The mean of all 10 values in $\mathcal{X}^*$ is $\hat{\mu}_1^* = 9.95$.

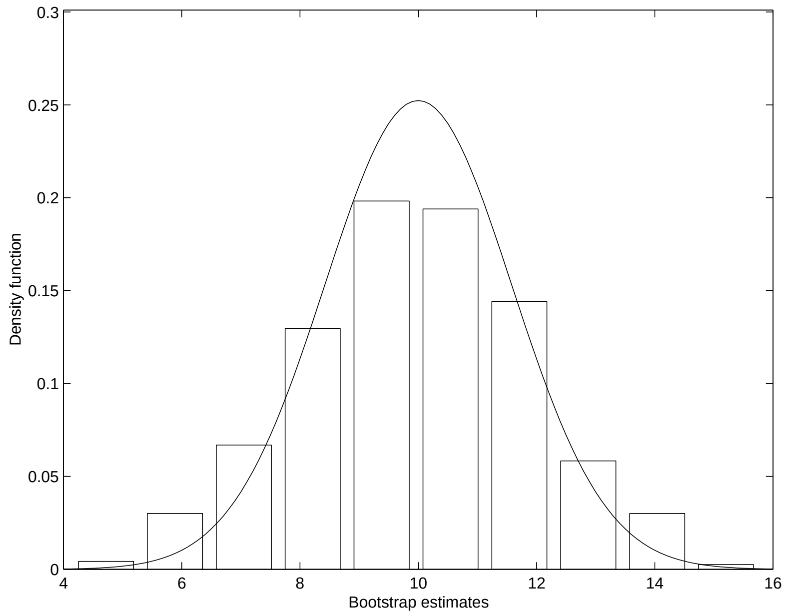
Example (Cont'd)

Step 3. *Repetition.* Repeat Steps 1 and 2 to obtain N bootstrap estimates $\hat{\mu}_1^*, \ldots, \hat{\mu}_N^*$. Let $N = 1000$.

Step 4. *Approximation of the Distribution of $\hat{\mu}$.* Sort the bootstrap estimates to obtain $\hat{\mu}_{(1)}^* \le \hat{\mu}_{(2)}^* \le \ldots \le \hat{\mu}_{(N)}^*$.

Step 5. *Confidence Interval.* The $100(1-\alpha)\%$ bootstrap confidence interval is $(\hat{\mu}_{(q_1)}^*, \hat{\mu}_{(q_2)}^*)$, where $q_1 = \lfloor N\alpha/2 \rfloor$ and $q_2 = N - q_1 + 1$. For $\alpha = 0.05$ and $N = 1000$, $q_1 = 25$ and $q_2 = 976$, and the 95% confidence interval is found to be $(\mathbf{6.01}, \mathbf{13.87})$ as compared to the theoretical $(\mathbf{6.85}, \mathbf{13.05})$.

Example (Cont'd)



Histogram of $\hat{\mu}_1^*, \hat{\mu}_2^*, \ldots, \hat{\mu}_{1000}^*$, based on the random sample $\mathcal{X} = \{-2.41, 4.86, 6.06, 9.11, 10.20, 12.81, 13.17, 14.10, 15.77, 15.79\}$, together with the density function of a Gaussian variable with mean 10 and variance 2.5.

Outline

- Introduction
- Principle of the Bootstrap
  - The Non-Parametric Bootstrap
  - Examples
    1. Bias Estimation
    2. Variance Estimation
    3. Confidence Interval Estimation for the Mean
  - The Parametric Bootstrap
  - An Example: Confidence Interval Estimation for the Mean
  - An Example of Bootstrap Failure

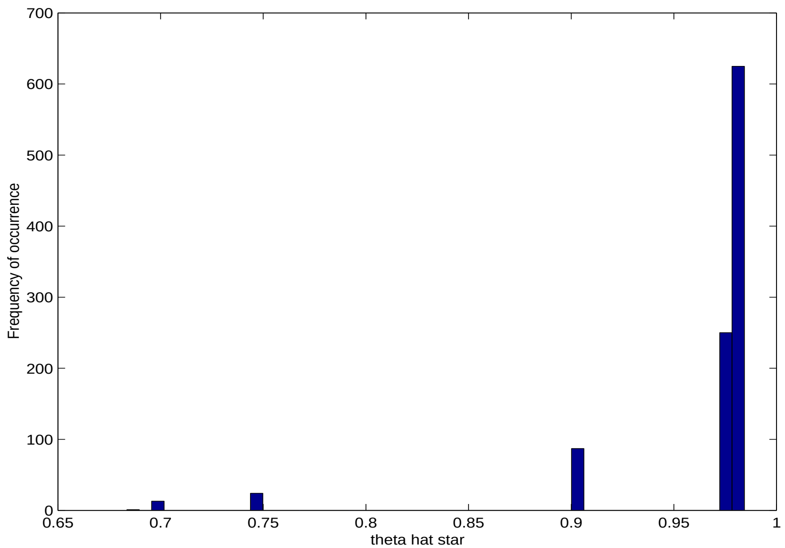
An Example of Bootstrap Failure

Let $X \sim \mathcal{U}(0, \theta)$ and $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$. We wish to estimate θ by $\hat{\theta}$ and its distribution $\hat{F}_{\hat{\Theta}}(\hat{\theta})$. The Maximum Likelihood (ML) estimator of θ is given by $\hat{\theta} = X_{(n)}$.

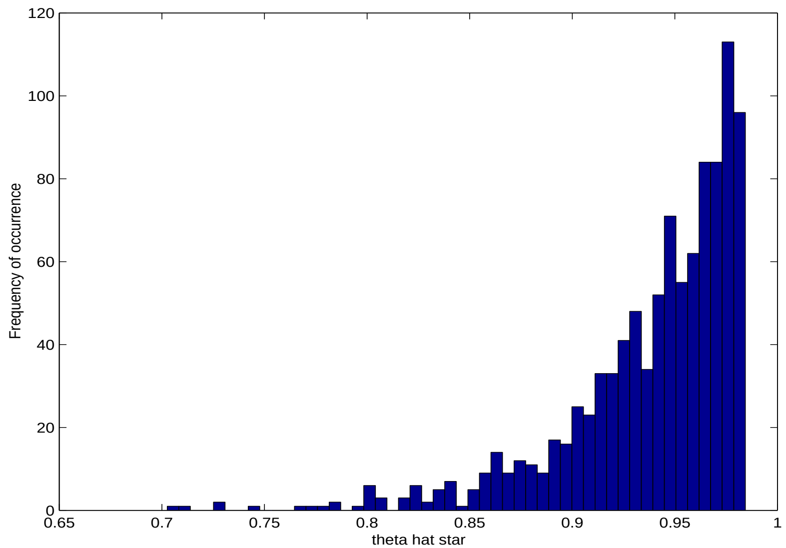
- To obtain an estimate of the density function of $\hat{\theta}$ we sample with replacement from the data and each time estimate $\hat{\theta}^*$ from $\mathcal{X}^*$.
- Alternatively, we could sample from $\mathcal{U}(0, \hat{\theta})$ and estimate $\hat{\theta}^*$ from $\mathcal{X}^*$ (parametric bootstrap).
- We ran an example with $\theta = 1$, $n = 50$ and $N = 1000$. The ML estimate of θ was found to be $\hat{\theta} = 0.9843$.

An Example of Bootstrap Failure (Cont'd)

The non-parametric bootstrap shows that approximately 62% of the values of $\hat{\theta}^*$ equal $\hat{\theta}$. In fact,
$\mathsf{Pr}(\hat{\theta}^* = \hat{\theta}) = 1 - (1 - 1/n)^n \longrightarrow 1 - e^{-1} \approx 0.632$ as $n \longrightarrow \infty$.



Non-parametric Bootstrap



Parametric Bootstrap

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimates
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

The Dependent Data Bootstrap

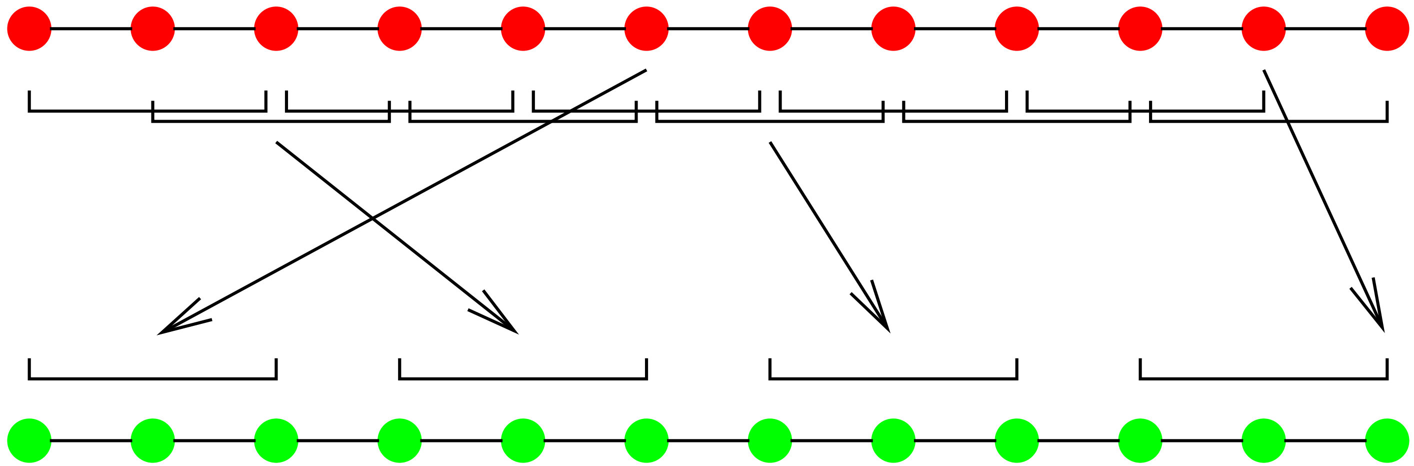
- The assumption of i.i.d. data can break down in practice either because the data is not independent or because it is not identically distributed, or both.
- We can still invoke the bootstrap principle if we **knew the model** that generated the data [Efron & Tibshirani (1993), Bose (1988), Kreiss & Franke (1992), Paparoditis (1996), Zoubir (1993)].
- For example, a way to relax the i.i.d. assumption is to assume that the data is identically distributed but not independent such as in autoregressive (AR) models.

The Dependent Data Bootstrap (Cont'd)

- If no plausible model such as AR is available for the probability mechanism generating stationary observations, we could make the assumption of weak dependence.
- Strong mixing processes[a], for example, satisfy the weak dependence condition.
- The *moving blocks* bootstrap [Künsch (1989), Liu & Singh (1992), Politis & Romano (1992,1994)] has been proposed for bootstrapping weakly dependent data.

[a] Loosely speaking a process is strong mixing if observations far apart (in time) are almost independent [Rosenblatt (1985)].

The Moving Blocks Bootstrap



Schematic diagram of the moving blocks bootstrap for a stationary signal. The red circles are the original signal. A bootstrap realisation of the signal (green circles) is generated by choosing a block length ("3" in the diagram) and sampling with replacement from all possible contiguous blocks of this length.

Other Block Bootstrap Methods

- The circular blocks bootstrap [Politis & Romano (1992), Shao & Yu (1993)] allows blocks which start at the end of the data and wrap around to the start.
- The blocks of blocks bootstrap [Politis *et al.* (1992)] uses two levels of blocking to estimate confidence bands for spectra and cross-spectra.
- The stationary bootstrap [Politis & Romano (1994)] allows blocks to be of random lengths instead of a fixed length.

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

An Example: Variance Estimation in AR Models

We generate n observations $x_t, t = 0, \ldots, n-1$, from

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

where Z_t is white Gaussian noise with $\mathsf{E} Z_t = 0$, $c_{ZZ}(u) = \sigma_Z^2 \delta(u)$, and a such that $|a| < 1$.

After de-trending the data, we fit the AR(1) model to the observation x_t. With $\hat{c}_{xx}(u) = 1/n \sum_{t=0}^{n-|u|-1} x_t x_{t+|u|}$ for $0 \le |u| \le n-1$, we calculate the Maximum Likelihood Estimate (MLE) of a, $\hat{a} = -\hat{c}_{xx}(1)/\hat{c}_{xx}(0)$, which has approximate variance‡ $\hat{\sigma}_{\hat{a}}^2 = (1-a^2)/n$.

‡under some regularity conditions an asymptotic formula for $\hat{\sigma}_{\hat{a}}^2$ can be found in the non-Gaussian case and is a function of a and the variance and kurtosis of Z_t [Porat & Friedlander (1989)].

Example (Cont'd)

Step 0. *Experiment.* Conduct the experiment and collect n observations x_t, $t = 0, \dots, n-1$, from an auto-regressive process of order one, X_t.

Step 1. *Calculation of the residuals.* With the Maximum Likelihood Estimate $\hat{a}$ of a, define the residuals $\hat{z}_t = x_t + \hat{a} \cdot x_{t-1}$ for $t = 1, 2, \dots, n-1$.

Step 2. *Resampling.* Create a bootstrap sample $x_0^*, x_1^*, \dots, x_{n-1}^*$ by sampling $\hat{z}_1^*, \hat{z}_2^*, \dots, \hat{z}_{n-1}^*$, with replacement, from the residuals $\hat{z}_1, \hat{z}_2, \dots, \hat{z}_{n-1}$, then letting $x_0^* = x_0$, and $x_t^* = -\hat{a}x_{t-1}^* + \hat{z}_t^*,\ t = 1, 2, \dots, n-1$.

Example (Cont'd)

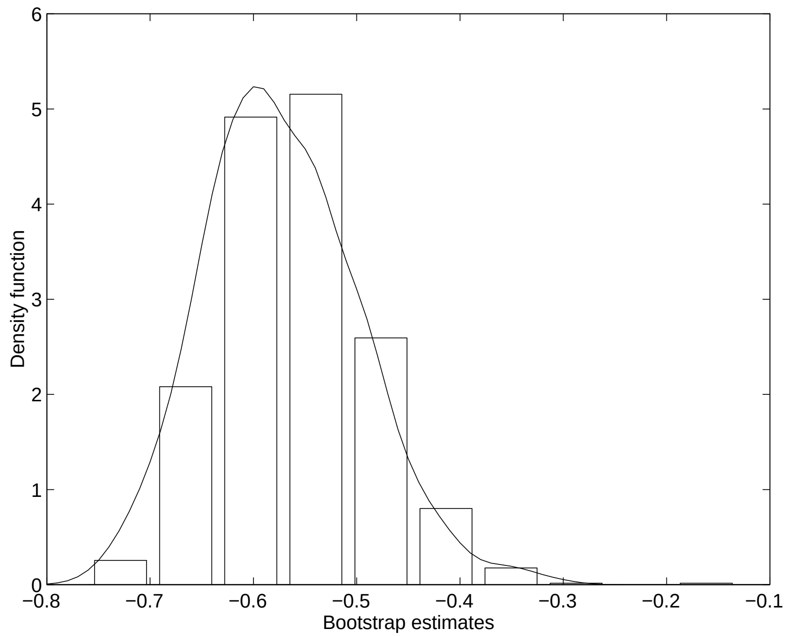
Step 3. *Calculation of the bootstrap estimate.* After centring the time series $x_0^*, x_1^*, \ldots, x_{n-1}^*$, obtain $\hat{a}^*$, using the above formulae but based on $x_0^*, x_1^*, \ldots, x_{n-1}^*$.

Step 4. *Repetition.* Repeat steps 2–3 a large number of times, $N = 1000$, say, to obtain $\hat{a}_1^*, \hat{a}_2^*, \ldots, \hat{a}_N^*$.

Step 5. *Variance estimation.* From $\hat{a}_1^*, \hat{a}_2^*, \ldots, \hat{a}_N^*$, approximate the variance of $\hat{a}$ by

$$\hat{\sigma}_{\hat{a}}^{*2} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{a}_i^* - \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i^*)^2 .$$

Example (Cont'd)



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

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

The Principle of Pivoting

- A statistic $T(X, \theta)$ is called pivotal if it possesses a fixed probability distribution independent of θ [Cramér (1967), Lehmann (1986)].
- Bootstrap confidence intervals or tests have excellent properties even for relatively low fixed resample number [Hall (1992)].
- For example, one can show that the coverage error in confidence interval estimation with the bootstrap is $O_p(n^{-1})$ as compared to $O_p(n^{-1/2})$ when using the normal approximation.
- The accuracy claimed holds whenever the statistic is asymptotically pivotal [Hall & Titterington (1989), Hall (1992)].

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

An Example: Confidence Interval Estimation

We consider the construction of a confidence interval for the mean. Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a random sample from some unknown F_{μ_X, σ_X}. We wish to find an estimator of μ_X with a $100(1-\alpha)\%$ confidence interval.

Let $\hat{\mu}_X$ and $\hat{\sigma}_X^2$ be the sample mean and the sample variance of $\mathcal{X}$, respectively. Alternatively to the previous example, we will base our method for finding a confidence interval for μ_X on the statistic

$$\hat{\mu}_Y = \frac{\hat{\mu}_X - \mu_X}{\hat{\sigma}},$$

where $\hat{\sigma}$ is the standard deviation of $\hat{\mu}_X$. The statistic has asymptotically for large n a distribution free of unknown parameters.

Example (Cont'd)

Step 0. *Experiment.* Conduct the experiment and collect the random data into the sample $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$.

Step 1. *Parameter estimation.* Based on $\mathcal{X}$, calculate $\hat{\mu}_X$ and its standard deviation $\hat{\sigma}$, using a nested bootstrap.

Step 2. *Resampling.* Draw a random sample, $\mathcal{X}^*$ of n values, with replacement, from $\mathcal{X}$.

Step 3. *Calculation of the pivotal statistic.* Calculate the mean of all values in $\mathcal{X}^*$ and using a nested bootstrap, calculate $\hat{\sigma}^*$. Then, form

$$\hat{\mu}_Y^* = \frac{\hat{\mu}_X^* - \hat{\mu}_X}{\hat{\sigma}^*}$$

Example (Cont'd)

Step 4. *Repetition.* Repeat Steps 2-3 many times to obtain a total of N bootstrap estimates $\hat{\mu}^*_{Y,1}, \ldots, \hat{\mu}^*_{Y,N}$.

Step 5. *Ranking.* Sort the bootstrap estimates to obtain $\hat{\mu}^*_{Y,(1)} \le \hat{\mu}^*_{Y,(2)} \le \ldots \le \hat{\mu}^*_{Y,(1000)}$.

Step 6. *Confidence Interval.* If $(\hat{\mu}^*_{Y,(q_1)}, \hat{\mu}^*_{Y,(q_2)})$ is an interval containing $(1-\alpha)N$ of the means $\hat{\mu}^*_Y$, where $q_1 = \lfloor N\alpha/2 \rfloor$ and $q_2 = N - q_1 + 1$, then

$$(\hat{\mu}_X - \hat{\sigma}\hat{\mu}^*_{Y,(q_2)}, \hat{\mu}_X - \hat{\sigma}\hat{\mu}^*_{Y,(q_1)})$$

is a $100(1-\alpha)\%$ confidence interval for μ_X.

Such an interval is known as a *percentile-t confidence interval* [Efron (1987), Hall (1988)].

Example (Cont'd)

For the same random sample $\mathcal{X}$ as before, we obtained the confidence interval $(\mathbf{3.54}, \mathbf{13.94})$ as compared to $(\mathbf{6.01}, \mathbf{13.87})$.

- This interval is larger than the one obtained earlier and enforces the statement that the interval obtained there has coverage less than the nominal 95%.
- It also yields better results than an interval derived using the assumption that $\hat{\mu}_Y$ is $\mathcal{N}(0,1)$- or the (better) approximation that $\hat{\mu}_Y$ is t_{n-1}-distributed.
- The interval obtained here accounts for skewness in the underlying population or other errors [Hall (1988,1992), Efron & Tibshirani (1993), Zoubir (1993)].

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

Variance Stabilisation

To ensure pivoting, usually the statistic is "studentised", i.e., we form

$$\hat{T} = \frac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}}$$

The percentile-t method is particularly applicable to location statistics, such as the sample mean, sample median, etc. [Efron & Tibshirani (1993)]. However, for more general statistics, it may not be accurate.

Problem: Studentising results in confidence intervals with erratically varying lengths and end points.

Solution: Pivoting often does not hold unless an appropriate variance stabilising transformation is applied first. *How do we "automatically" get a variance stabilising transformation?*

Variance Stabilisation (Cont'd)

Step 1. *Estimation of the variance stabilising transformation.*

(a) Generate B_1 bootstrap samples $\mathcal{X}_i^*$ from $\mathcal{X}$ and for each calculate the value of the statistic $\hat{\theta}_i^*$, $i = 1, \ldots, B_1$. For example $B_1 = 100$.

(b) Generate B_2 bootstrap samples from $\mathcal{X}_i^*$, $i = 1, \ldots, B_1$, and calculate $\hat{\sigma}_i^{*2}$, a bootstrap estimate for the variance of $\hat{\theta}_i^*$, $i = 1, \ldots, B_1$. For example $B_2 = 25$.

(c) Estimate the variance function $\zeta(\theta)$ by smoothing the values of $\hat{\sigma}_i^{*2}$ against $\hat{\theta}_i^*$, using, for example, a fixed-span 50% "running lines" smoother [Hastie & Tibshirani (1990)].

Variance Stabilisation (Cont'd)

Step 1. *Estimation of the variance stabilising transformation.*

(d) Estimate the variance stabilising transformation $h(\hat{\theta})$ from

$$h(\theta) = \int^{\theta} \{\zeta(s)\}^{-1/2} ds.$$

Step 2. *Bootstrap quantile estimation.* Generate B_3 bootstrap samples and compute $\hat{\theta}_i^*$ and $h(\hat{\theta}_i^*)$ for each sample i. Approximate the distribution of $h(\hat{\theta}) - h(\theta)$ by that of $h(\hat{\theta}^*) - h(\hat{\theta})$.

Outline

- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Confidence Interval for the Mean
    2. Coherence Gain for Engine Knock Data

Example: Correlation coefficient

Let $\theta = \varrho$ be the correlation coefficient of two unknown populations, and let $\hat{\varrho}$ and $\hat{\sigma}^2$ be estimates of ϱ and the variance of $\hat{\varrho}$, respectively, based on $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$.

Let $\mathcal{X}^*$ and $\mathcal{Y}^*$ be resamples, drawn with replacement from $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let $\hat{\varrho}^*$ and $\hat{\sigma}^{*2}$ be bootstrap versions of $\hat{\varrho}$ and $\hat{\sigma}^2$.

By repeated resampling from $\mathcal{X}$ and $\mathcal{Y}$ we compute $\hat{s}_\alpha$ and $\hat{t}_\alpha$, such that with $0 < \alpha < 1$

$$\mathsf{Pr}\left((\hat{\varrho}^* - \hat{\varrho})/\hat{\sigma}^* \leq \hat{s}_\alpha \,\middle|\, \mathcal{X}, \mathcal{Y} \right) = \frac{\alpha}{2} = \mathsf{Pr}\left((\hat{\varrho}^* - \hat{\varrho})/\hat{\sigma}^* \geq \hat{t}_\alpha \,\middle|\, \mathcal{X}, \mathcal{Y} \right) .$$

The confidence interval (percentile-t) for ϱ is given by

$$\mathcal{I}(\mathcal{X}, \mathcal{Y}) = \left(\hat{\varrho} - \hat{\sigma}\hat{t}_\alpha \, , \, \hat{\varrho} - \hat{\sigma}\hat{s}_\alpha \right) .$$

Example: Correlation coefficient (Cont'd)

There exists a transformation called Fisher's z-transform [Fisher (1921), Anderson (1984)], which is stabilising and normalising:

$$\breve{\varrho} = \tanh^{-1}\hat{\varrho} = \frac{1}{2}\log\frac{1+\hat{\varrho}}{1-\hat{\varrho}}\,.$$

We could first find a confidence interval for $\xi = \tanh^{-1}\varrho$ and then transform the endpoints back with the inverse transformation $\varrho = \tanh\xi$ to obtain a confidence interval for ϱ.

For X and Y bivariate normal ($\breve{\varrho} \sim \mathcal{N}(\xi, 1/(n-3))$), a 95%, for example, confidence interval for ϱ is obtained from

$$\left(\tanh(-1.96/\sqrt{n-3}+\breve{\varrho})\,,\,\tanh(1.96/\sqrt{n-3}+\breve{\varrho})\right)\,.$$

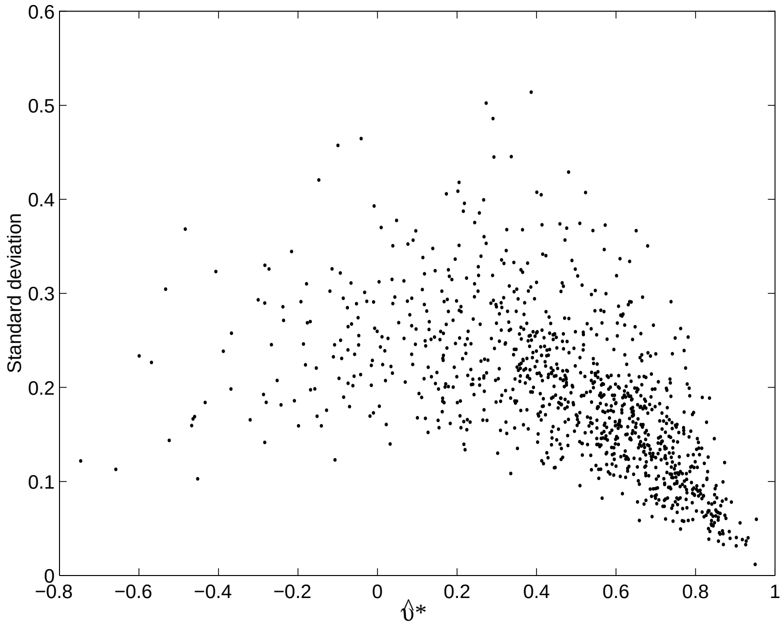
Example: Correlation coefficient (Cont'd)

Let $X = Z_1 + W$ and $Y = Z_2 + W$, where Z_1, Z_2 and W are pairwise i.i.d. Then, $\varrho_{XY} = 0.5$. We drew $n = 15$ realisations $z_{1,i}, z_{2,i}$ and w_i, from the normal distribution and calculated x_i, y_i, $i = 1, \ldots, 15$.

- Assuming a normal distribution we found $\hat{\varrho}_{XY} = 0.36$ and the 95% confidence interval $(\mathbf{-0.18}, \mathbf{0.74})$ for $\varrho_{X,Y}$.
- using the bootstrap percentile-t method, we found with $N = 1000$ the 95% confidence interval $(\mathbf{-0.05}, \mathbf{1.44})$.
- Using Fisher's z-transform and the bootstrap (without assuming bivariate normality), a confidence interval was found to be $(\mathbf{-0.28}, \mathbf{0.93})$.

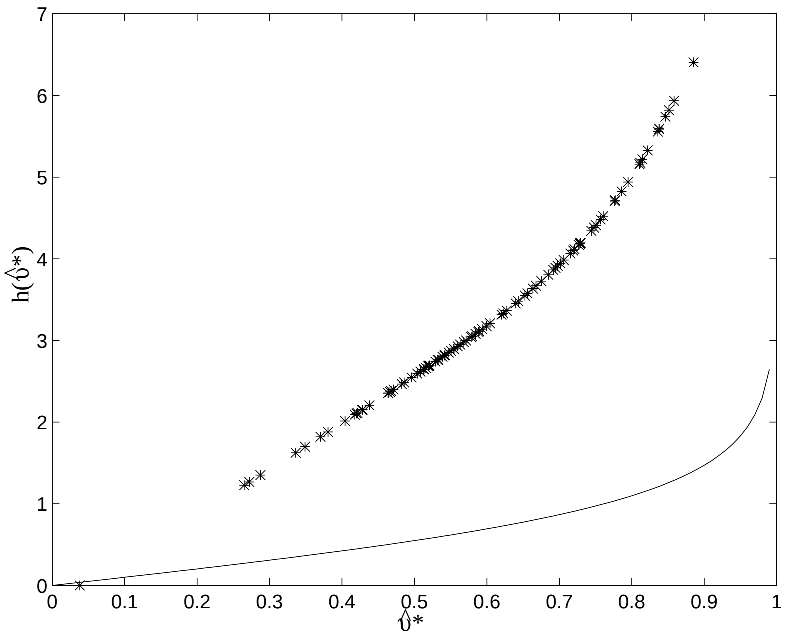
The interval found using the percentile-t method is over-covering

Example: Correlation coefficient (Cont'd)



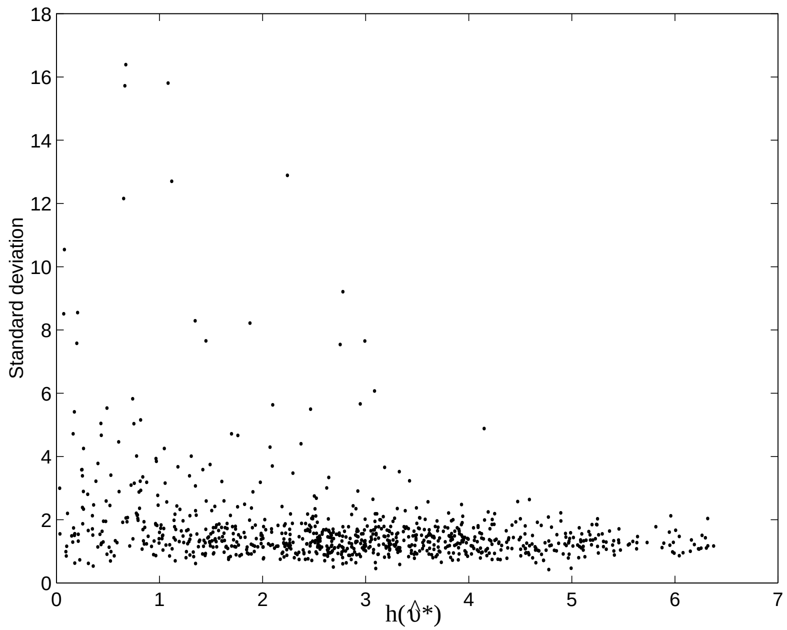
Bootstrap estimates of the standard deviation of $B_3 = 1000$ (bootstrap) estimates of the correlation coefficient before variance stabilisation.

Example: Correlation coefficient (Cont'd)



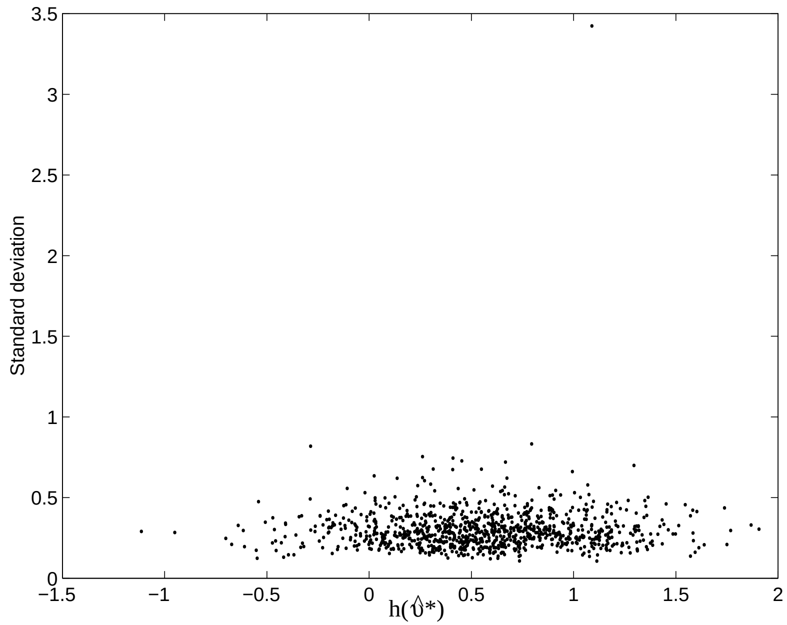
Variance stabilising transformation for the correlation coefficient estimated using $B_1 = 100$ and $B_2 = 25$. The solid line is a plot of Fisher's z-transform.

Example: Correlation coefficient (Cont'd)



Bootstrap estimates of the standard deviation of $B_3 = 1000$ (new bootstrap) estimates of the correlation coefficient after variance stabilisation, obtained through bootstrap. The confidence interval found was $(\mathbf{0.06}, \mathbf{0.97})$.

Example: Correlation coefficient (Cont'd)

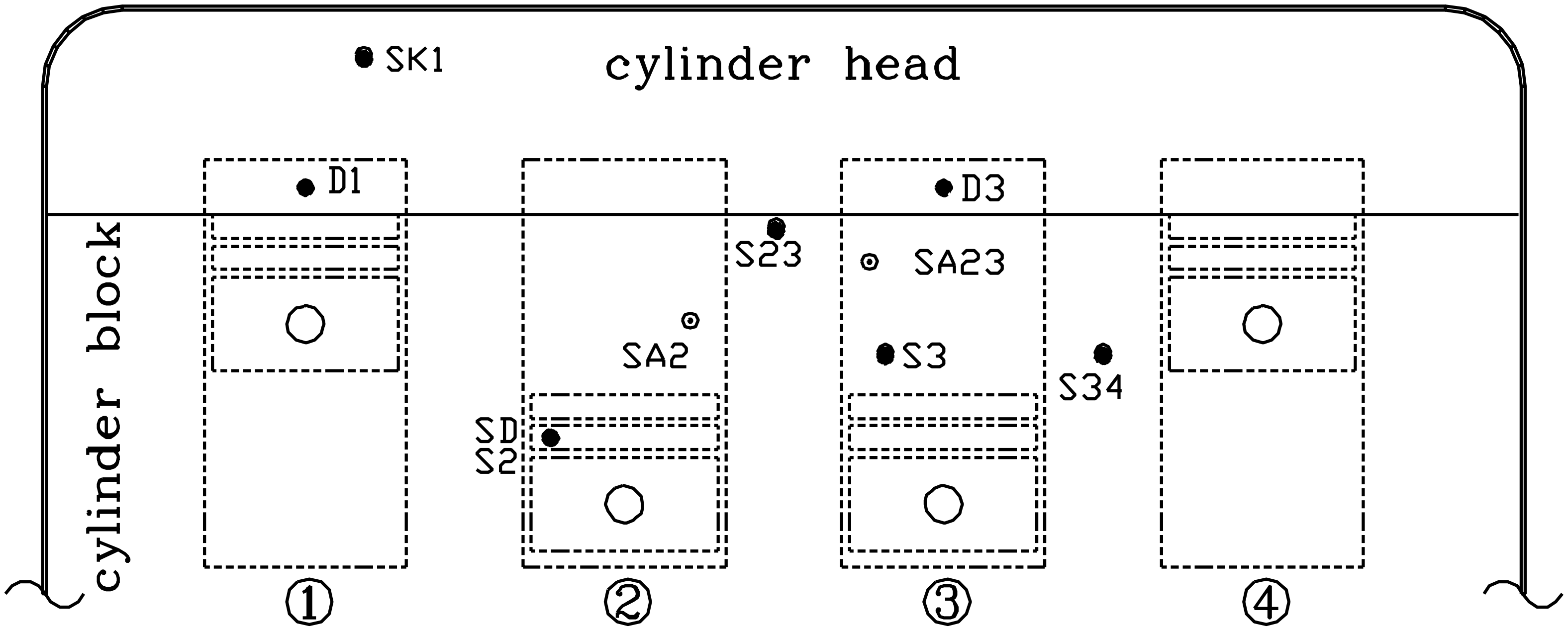


Bootstrap estimates of the standard deviation of $B_3 = 1000$ (bootstrap) estimates of the correlation coefficient after applying Fisher's variance stabilising transformation $\tanh^{-1}$.

Outline

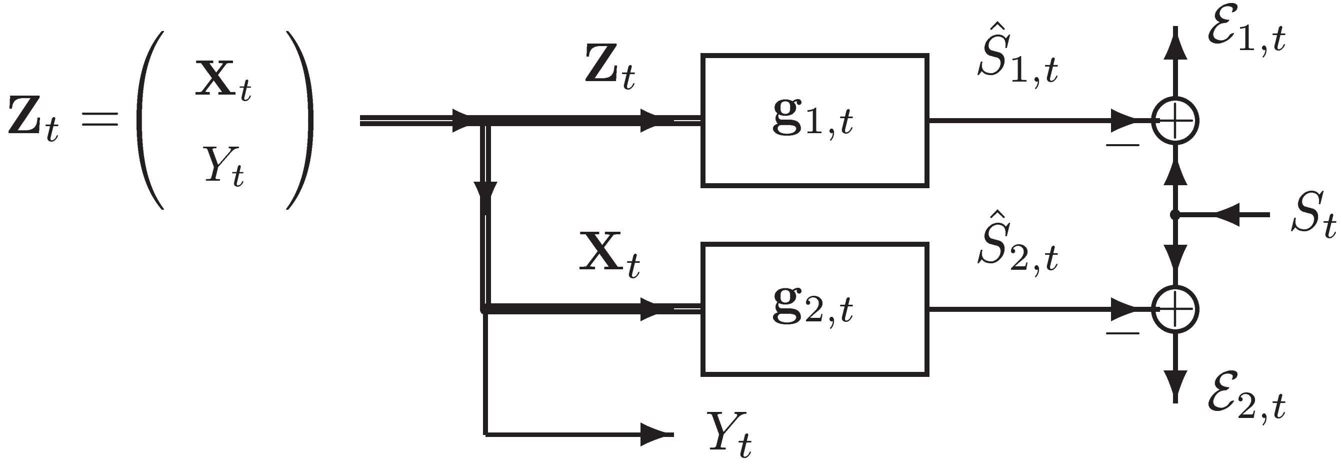
- Principle of the Bootstrap
  - The Dependent Data Bootstrap
  - An Example: Variance Estimation for Autoregressive Parameter Estimators
- The Principle of Pivoting
  - An Example: Confidence Interval for the Mean
- Variance Stabilisation
  - Examples
    1. Correlation Coefficient
    2. Coherence Gain for Engine Knock Data

Example: Coherence Gain for Knock Data



vibration sensors distributed on the block of a Volkswagen Passat four-cylinder engine with 1.8l, 79 kW and 10:1 compression ratio.

Coherence Gain for Knock Data (Cont'd)



$\mathbf{Z}_t$: vector of vibration signals, observed;

S_t: cylinder pressure signal, observed;

$\mathbf{g}_{1,t}$: prediction filter impulse response, unknown;

$\mathbf{g}_{2,t}$: prediction filter impulse response, unknown;

$\mathcal{E}_{i,t}$: prediction error, $\mathcal{E}_{i,t} = S_t - \hat{S}_{i,t}$, $i = 1, 2$, unknown.

Coherence Gain for Knock Data (Cont'd)

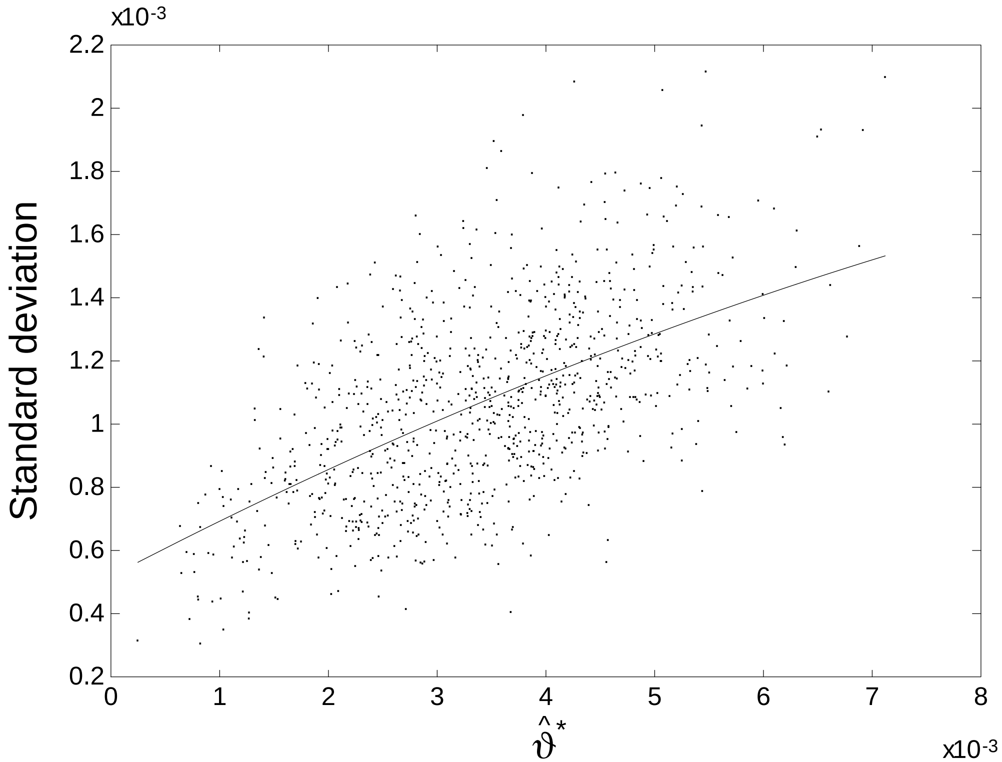
The closeness to zero of $\theta(\omega) = R^2_{SZ}(\omega) - R^2_{SX}(\omega)$, the coherence gain explained by the sensor with output signal Y_t, is an irrelevancy measure of this sensor among the remaining sensors.

This suggests to

test	$\mathsf{H} : R^2_{SZ}(\omega) - R^2_{SX}(\omega) \leq \theta_0(\omega)$	
against	$\mathsf{K} : R^2_{SZ}(\omega) - R^2_{SX}(\omega) > \theta_0(\omega)\,,$	$0 < \theta_0(\omega) < 1\,.$

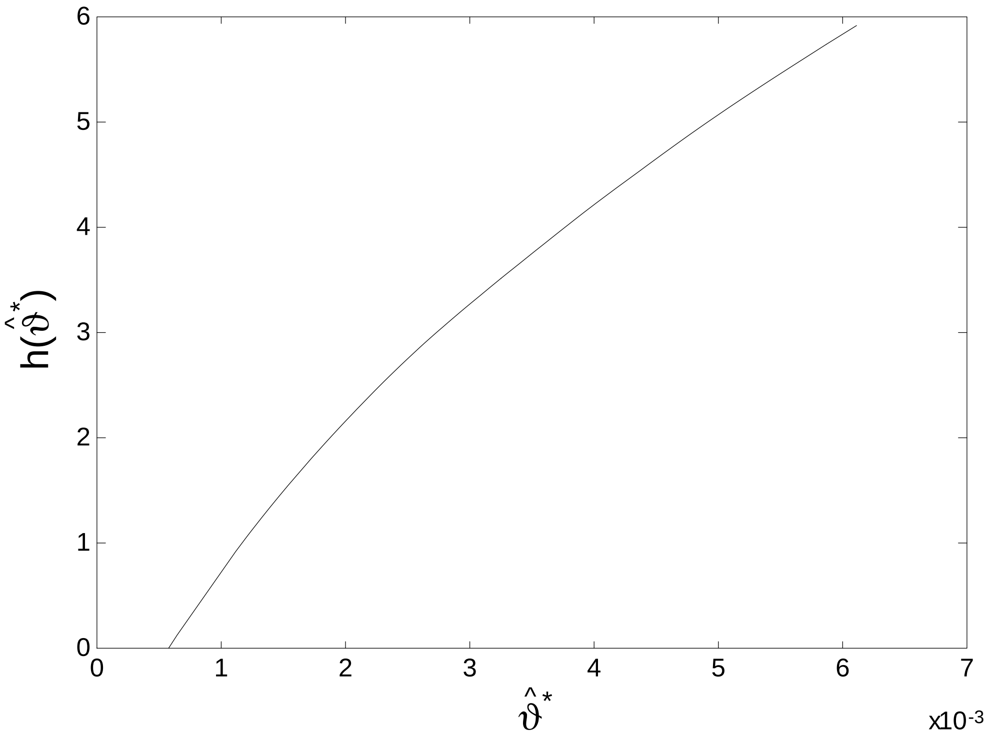
Problem: Distribution and variance stabilisation for the test statistic $\hat{T}(\omega) = (\hat{\theta}(\omega) - \theta_0(\omega))/\hat{\sigma}(\omega)$, with $\hat{\theta}(\omega) = \hat{R}^2_{SZ}(\omega) - \hat{R}^2_{SX}(\omega)$, are unknown.

Example (Cont'd)



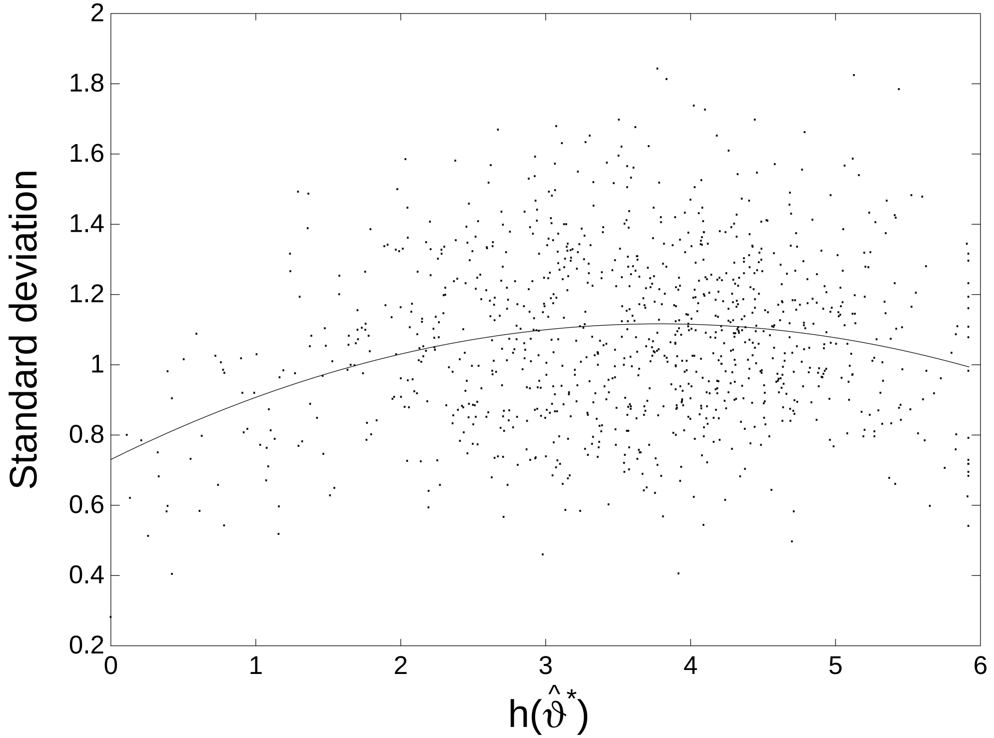
Standard deviation of bootstrap estimates at one mode frequency without variance stabilisation. Herein, $\hat{\theta}(\omega) = \hat{R}^2_{SZ}(\omega) - \hat{R}^2_{SX}(\omega)$.

Example (Cont'd)



Estimated variance stabilising transformation. The transformation was found using $B_1 = 200, B_2 = 25$ and a fixed-span running lines smoother with span of 50%.

Example (Cont'd)



Standard deviation of (new) bootstrap estimates $\hat{\theta}^*(\omega)$ at one mode frequency after variance stabilisation.

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Hypothesis Testing with the Bootstrap

Consider a random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ observed from an unspecified probability distribution F. Let θ be an unknown parameter of F.

We wish to test the hypothesis

$$\mathsf{H} : \theta \leq \theta_0 \qquad \text{against} \qquad \mathsf{K} : \theta > \theta_0 \,,$$

where θ_0 is some known constant. Let $\hat{\theta}$ be an estimator of θ and $\hat{\sigma}^2$ an estimator of the variance σ^2 of $\hat{\theta}$.

Define the statistic

$$\hat{T} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}} \,.$$

Hypothesis Testing (Cont'd)

Step 1. Draw $\mathcal{X}^*$, with replacement, from $\mathcal{X}$.

Step 2. Calculate

$$\hat{T}^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*},$$

where $\hat{\theta}^*$ and $\hat{\sigma}^*$ are versions of $\hat{\theta}$ and $\hat{\sigma}$ computed from $\mathcal{X}^*$.

Step 3. Repeat Steps 1 and 2 to obtain $\hat{T}_1^*, \ldots, \hat{T}_N^*$.

Step 4. Rank $\hat{T}_1^*, \ldots, \hat{T}_N^*$ into $\hat{T}_{(1)}^* \leq \cdots \leq \hat{T}_{(N)}^*$. Reject H if $\hat{T} \geq \hat{T}_{(q)}^*$, where $q = \lfloor (N+1)(1-\alpha) \rfloor$.

Hypothesis Testing (Cont'd)

For a double-sided alternative we would test

$$\mathsf{H} : \theta = \theta_0 \qquad \text{against} \qquad \mathsf{K} : \theta \neq \theta_0 \, ,$$

where θ_0 is some known constant. The statistic used is given by

$$\hat{T}_d = \frac{|\hat{\theta} - \theta_0|}{\hat{\sigma}} \, .$$

We would proceed as before and rank the bootstrap statistics $\hat{T}^*_{d,1}, \ldots, \hat{T}^*_{d,N}$ into $\hat{T}^*_{d,(1)} \leq \cdots \leq \hat{T}^*_{d,(N)}$. Then, we would reject H at level α if $\hat{T}_d \geq \hat{T}^*_{d,(q)}$, where $q = \lfloor (N+1)(1-\alpha) \rfloor$.

Hypothesis Testing: Variance Estimation

If an estimator $\hat{\sigma}^2$ is unavailable, use the bootstrap: Given $\hat{\theta}_1^*, \ldots, \hat{\theta}_{B_1}^*$, we estimate the variance $\hat{\sigma}^2$ of $\hat{\theta}$ by

$$\hat{\sigma}^2_{\mathsf{BOOT}} = \frac{1}{B_1 - 1} \sum_{b=1}^{B_1} \left(\hat{\theta}_b^* - \frac{1}{B_1} \sum_{b=1}^{B_1} \hat{\theta}_b^* \right)^2 .$$

In the case of $\hat{\sigma}^{*2}$, the procedure involves two nested levels of resampling. For each resample $\mathcal{X}_b^*$, $b = 1, \ldots, B_1$, we draw resamples $\mathcal{X}_b^{**}$, $b = 1, \ldots, B_2$, evaluate $\hat{\theta}_b^{**}$ from each resample to obtain B_2 replications, and calculate

$$\hat{\sigma}^{*2}_{\mathsf{BOOT}} = \frac{1}{B_2 - 1} \sum_{b=1}^{B_2} \left(\hat{\theta}_b^{**} - \frac{1}{B_2} \sum_{b=1}^{B_2} \hat{\theta}_b^{**} \right)^2 .$$

Variance Estimation (Cont'd)

The jackknife can be thought of as drawing n samples of size $n-1$ each *without* replacement from the original sample of size n [Miller (1974)].

The jackknife is based on the sample *delete-one observation at a time*, $\mathcal{X}^{(i)} = \{X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$, $i = 1, 2, \ldots, n$, called jackknife sample. For each ith jackknife sample, we calculate the ith jackknife estimate $\hat{\theta}^{(i)}$ of θ, $i = 1, \ldots, n$ and compute

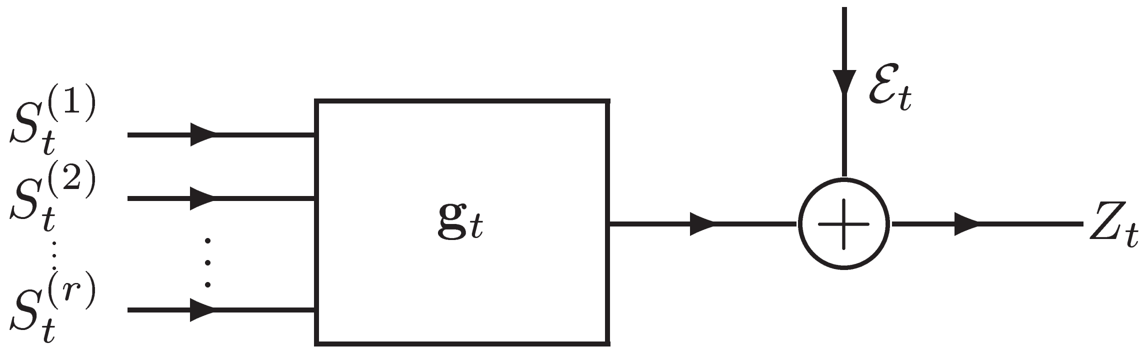
$$\hat{\sigma}^2_{\mathsf{JACK}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}^{(i)} - \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}^{(i)} \right)^2 ,$$

which is less expensive than the bootstrap if n is less than the number of replicates used by the bootstrap for standard deviation.

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

An Example: Testing the Frequency Response



$\mathbf{S}_t$: r vector-valued stationary signal, observed;

Z_t: output signal, observed;

$\mathbf{g}_t$: filter impulse response, unknown;

$\mathcal{E}_t$: noise, unknown, $\mathcal{E}_t$ and $\mathbf{S}_t$ independent for $t = 0, \pm 1, \pm 2, \ldots$.

$$Z_t = \sum_{u=-\infty}^{\infty} \mathbf{g}_u' \mathbf{S}_{t-u} + \mathcal{E}_t \ ,$$

Problem: which element $G_l(\omega)$, $1 \leq l \leq r$, is zero at a given ω?

Example (Cont'd)

Let $\mathbf{G}(\omega) = (\mathbf{G}^{(l)}(\omega)', G_l(\omega))'$. Given $\mathbf{S}_t$ and Z_t for n independent observations of length T each, we wish to test

$$\mathsf{H} : G_l(\omega) = 0 \ (\mathbf{G}^{(l)}(\omega) \text{ unspecified}) \quad \text{against} \quad \mathsf{K} : G_l(\omega) \neq 0 \,.$$

We compute the frequency data $\mathbf{d}_{\mathbf{s}}(\omega) = (\mathbf{d}_{S_1}(\omega), \ldots, \mathbf{d}_{S_r}(\omega))$,
$\mathbf{d}_{S_l}(\omega) = (d_{S_l}(\omega, 1), \ldots, d_{S_l}(\omega, n))'$, $l = 1, \ldots, r$,
$\mathbf{d}_Z(\omega) = (d_Z(\omega, 1), \ldots, d_Z(\omega, n))'$, with
$d_Z(\omega, i) = \sum_{t=0}^{T-1} w(t/T) \cdot Z_{t,i}\, e^{-j\omega t}, i = 1, \ldots, n$, and consider the complex regression

$$\mathbf{d}_Z(\omega) = \mathbf{d}_{\mathbf{s}}(\omega)\mathbf{G}(\omega) + \mathbf{d}_{\mathcal{E}}(\omega) \,.$$

Let the Least-Squares Estimate (LSE) of $\mathbf{G}(\omega)$ be

$$\hat{\mathbf{G}}(\omega) = (\mathbf{d}_{\mathbf{s}}(\omega)^H \mathbf{d}_{\mathbf{s}}(\omega))^{-1} (\mathbf{d}_{\mathbf{s}}(\omega)^H \mathbf{d}_Z(\omega)) \,.$$

Example (Cont'd)

Conventional techniques assume T large so that $\mathbf{d}_{\mathcal{E}}(\omega)$ becomes complex Gaussian [Brillinger (1981)]. Under this condition and H, the statistic

$$\hat{T}(\omega) = (n-r)\frac{\| \mathbf{d}_Z(\omega) - \mathbf{d}_{\mathbf{s}^{(l)}}(\omega)\hat{\mathbf{G}}^{(l)}(\omega) \|^2 - \| \mathbf{d}_Z(\omega) - \mathbf{d}_{\mathbf{s}}(\omega)\hat{\mathbf{G}}(\omega) \|^2}{\| \mathbf{d}_Z(\omega) - \mathbf{d}_{\mathbf{s}}(\omega)\hat{\mathbf{G}}(\omega) \|^2}$$

is $F_{2,2(n-r)}$-distributed, where

$$\mathbf{d}_{\mathbf{s}^{(l)}}(\omega) = (\mathbf{d}_{S_1}(\omega), \ldots, \mathbf{d}_{S_{l-1}}(\omega), \mathbf{d}_{S_{l+1}}(\omega), \ldots, \mathbf{d}_{S_r}(\omega))'$$

is obtained from $\mathbf{d}_{\mathbf{s}}(\omega) = (\mathbf{d}_{\mathbf{s}^{(l)}}(\omega), \mathbf{d}_{S_l}(\omega))$ by deleting the lth vector $\mathbf{d}_{S_l}(\omega)$, and $\hat{\mathbf{G}}^{(l)}(\omega) = (\mathbf{d}_{\mathbf{s}^{(l)}}(\omega)^H \mathbf{d}_{\mathbf{s}^{(l)}}(\omega))^{-1}(\mathbf{d}_{\mathbf{s}^{(l)}}(\omega)^H \mathbf{d}_Z(\omega))$ [Shumway (1983)].

Example (Cont'd)

Step 0. *Experiment.* Conduct the experiment and calculate $\mathbf{d}_S(\omega,1),\ldots,\mathbf{d}_S(\omega,n)$, and $d_Z(\omega,1),\ldots,d_Z(\omega,n)$.

Step 1. *Resampling.* Conduct two totally independent resampling operations in which $\{\mathbf{d}^*_S(\omega,1),\ldots,\ \mathbf{d}^*_S(\omega,n)\}$ is drawn, with replacement, from $\{\mathbf{d}_S(\omega,1),\ldots,\mathbf{d}_S(\omega,n)\}$, where $\mathbf{d}_S(\omega,i)=(d_{S_1}(\omega,i),\ldots,d_{S_r}(\omega,i))$, $i=1,\ldots,n$, and a resample $\{d^*_{\hat{\mathcal{E}}}(\omega,1),\ldots,\ d^*_{\hat{\mathcal{E}}}(\omega,n)\}$ is drawn, with replacement, from $\{d_{\hat{\mathcal{E}}}(\omega,1),\ldots,d_{\hat{\mathcal{E}}}(\omega,n)\}$, collected into the vector $\mathbf{d}_{\hat{\mathcal{E}}}(\omega)=(d_{\hat{\mathcal{E}}}(\omega,1),\ldots,d_{\hat{\mathcal{E}}}(\omega,n))'$, so that

$$\mathbf{d}_{\hat{\mathcal{E}}}(\omega)=\mathbf{d}_Z(\omega)-\mathbf{d}_{\mathbf{s}}(\omega)\hat{\mathbf{G}}(\omega)\,.$$

Example (Cont'd)

Step 2. *Generation of bootstrap data.* Centre the frequency data resamples and compute

$$\mathbf{d}_Z^*(\omega) = \mathbf{d}_\mathbf{s}^*(\omega)\hat{\mathbf{G}}(\omega) + \mathbf{d}_{\hat{\mathcal{E}}}(\omega)\,.$$

The joint distribution of $\{(\mathbf{d}_S^*(\omega, i), d_Z^*(\omega, i)), 1 \leq i \leq n\}$, conditional on $\mathcal{X}(\omega) = \{(\mathbf{d}_S(\omega, 1), d_Z(\omega, 1)), \ldots,$ $(\mathbf{d}_S(\omega, n), d_Z(\omega, n))\}$ is the bootstrap estimate of the unconditional joint distribution of $\mathcal{X}(\omega)$.

Step 3. *Calculation of bootstrap estimates.* With the new $\mathbf{d}_Z^*(\omega)$ and $\mathbf{d}_\mathbf{s}^*(\omega)$, calculate the LSE $\hat{\mathbf{G}}^*(\omega)$, using the resamples $\mathbf{d}_Z^*(\omega)$ and $\mathbf{d}_\mathbf{s}^*(\omega)$, replacing $\mathbf{d}_Z(\omega)$ and $\mathbf{d}_\mathbf{s}(\omega)$, respectively.

Example (Cont'd)

Step 4. *Calculation of the bootstrap statistic.* Calculate the statistic, replacing $\mathbf{d_s}(\omega)$, $\mathbf{d}_{\mathbf{s}^{(l)}}(\omega)$, $\mathbf{d}_Z(\omega)$, $\hat{\mathbf{G}}(\omega)$, and $\hat{\mathbf{G}}^{(l)}(\omega)$ by their bootstrap counterparts to yield $\hat{T}^*(\omega)$.

Step 5. *Repetition.* Repeat Steps 1–4 a large number of times, say N, to obtain $\hat{T}_1^*(\omega), \ldots, \hat{T}_N^*(\omega)$.

Step 6. *Distribution estimation.* Approximate the distribution of $\hat{T}(\omega)$ by the distribution of $\hat{T}^*(\omega)$ obtained.

An alternative bootstrap approach to the one described here can be obtained by expressing $\hat{T}(\omega)$ using multiple coherences [Zoubir (1993,1994), Zoubir & Boashash (1998)].

Example (Cont'd)

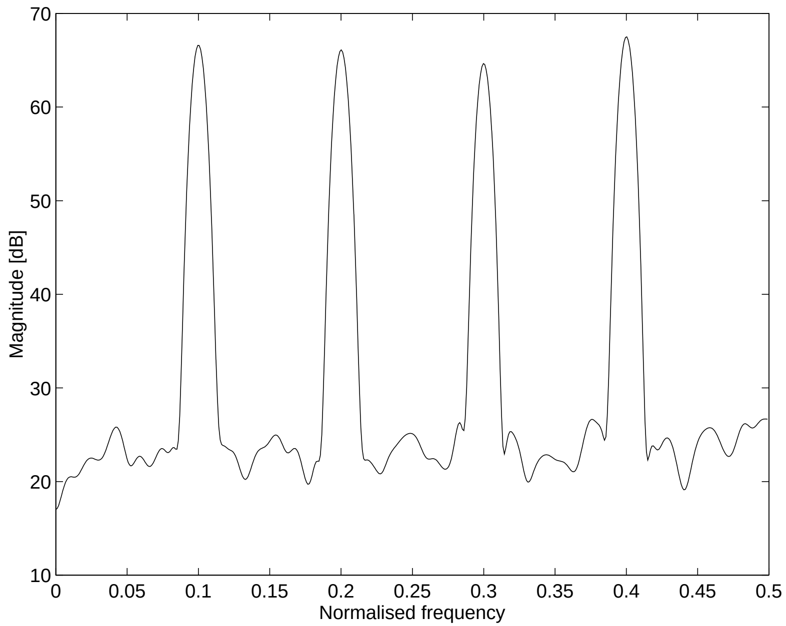
Let $n = 20$ be independent records of $\mathbf{S}_t$ with $r = 5$ and let

$$S_{l,t} = \sum_{k=1}^{K} A_{k,l} \cos(\omega_k t + \Phi_{k,l}) + U_{l,t}\,, \quad l = 1, \ldots, 5\,,$$

where $A_{k,l}$ and $\Phi_{k,l}$ are mutually independent random amplitudes and phases, respectively, ω_k are arbitrary resonance frequencies for $k = 1, \ldots, K$ and $U_{l,t}$ is a white noise process, $l = 1, \ldots, r$.

With $K = 4$ and $T = 128$, we generated data for Φ and A from a uniform distribution on the interval $[0, 2\pi)$ and $[0, 1)$, respectively. We selected $f_1 = 0.1$, $f_2 = 0.2$, $f_3 = 0.3$ and $f_4 = 0.4$, where $f_k = \omega_k/2\pi$, $k = 1, \ldots, 4$. The added noise was uniformly distributed.

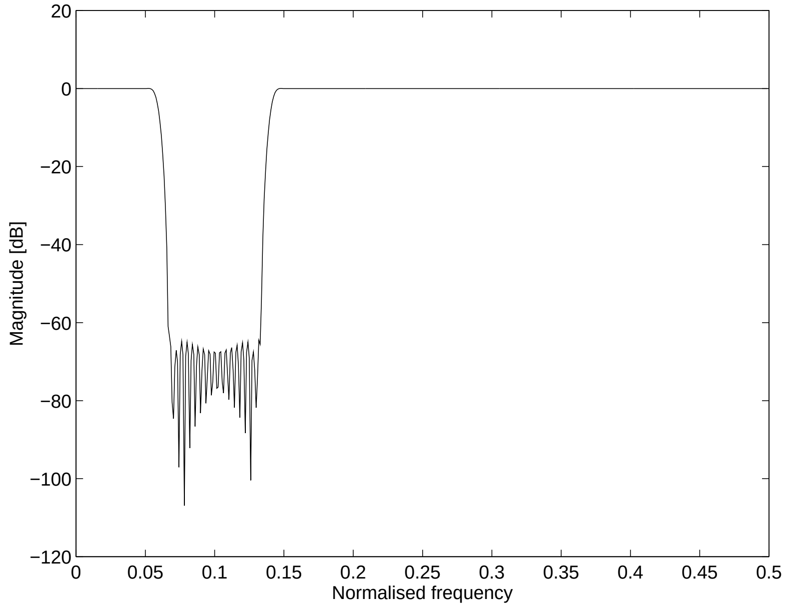
Example (Cont'd)



Spectral estimate of $S_{l,t}$, $l = 1, \ldots, r$, $\mathbf{d}_{S_l}(\omega)^H \mathbf{d}_{S_l}(\omega)/n$, obtained by averaging $n = 20$ periodograms.

Example (Cont'd)

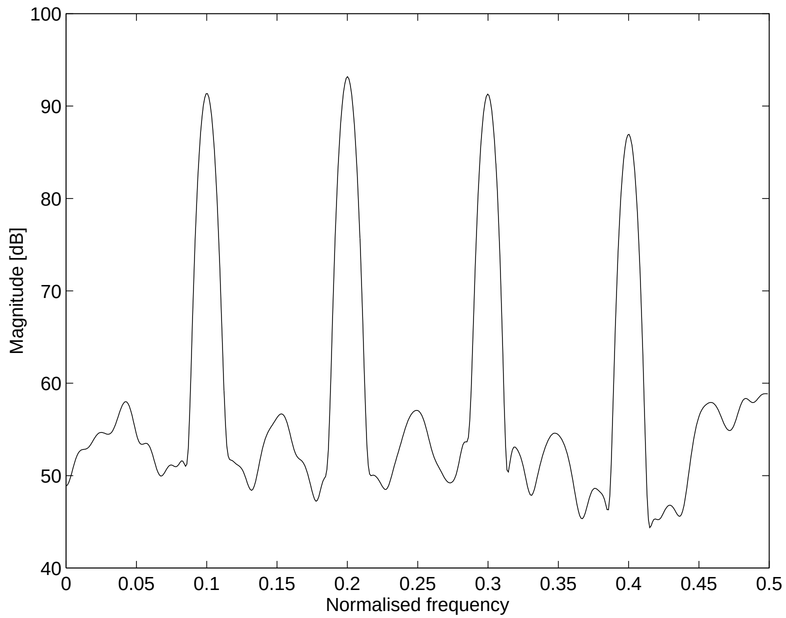
We generated bandstop filters (FIR filters with 256 coefficients) with bands centred about the four resonance frequencies f_1, f_2, f_3 and f_4.



Frequency response of the first channel, $G_1(\omega)$, obtained using an FIR filter with 256 coefficients.

Example (Cont'd)

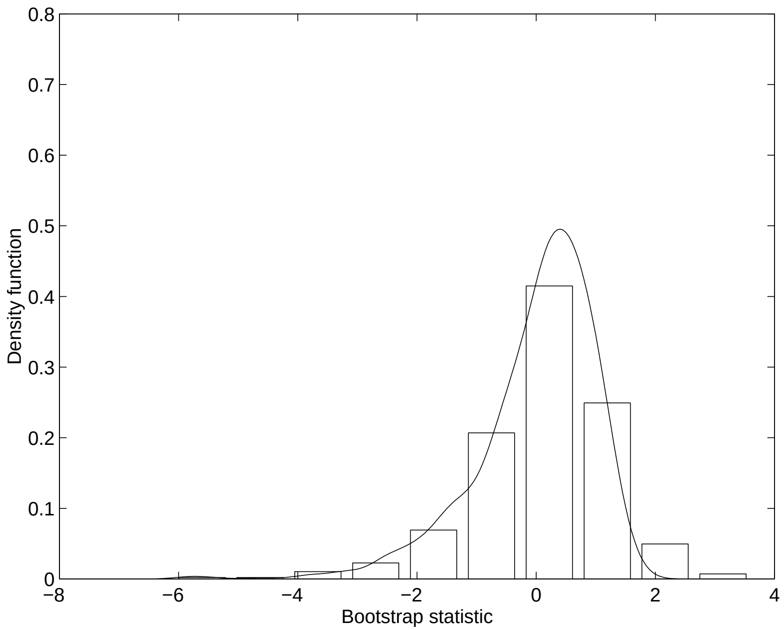
We filtered $\mathbf{S}_t$ and added independent uniformly distributed noise $\mathcal{E}_t$ to generate Z_t (SNR $= 5$ dB with respect to the component $S_{l,t}$, $l = 1, \ldots, r$, $t = 0, \ldots, T-1$, with highest power).



Spectral estimate of Z_t, $\mathbf{d}_Z(\omega)^H \mathbf{d}_Z(\omega)/n$, with $n = 20$.

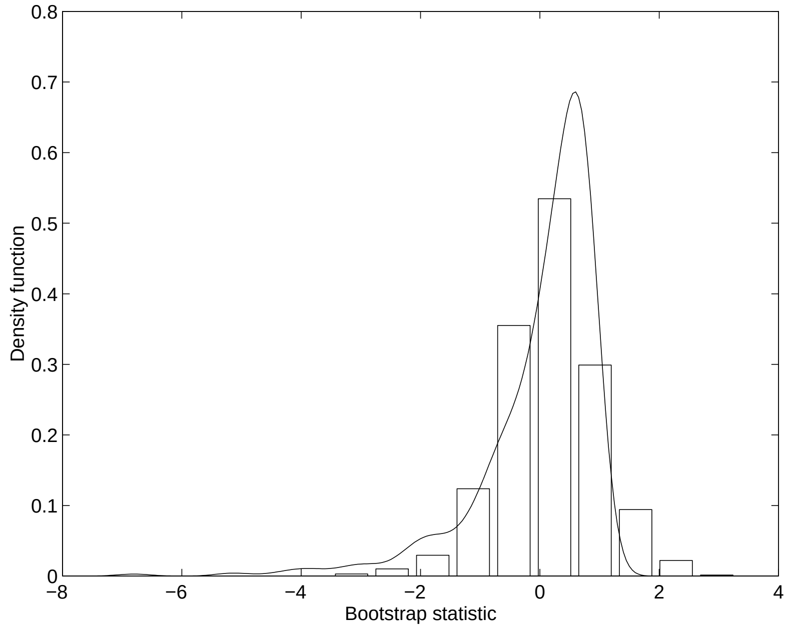
Example (Cont'd)

We selected arbitrarily $G_l(\omega)$, $l = 1, \ldots, r$, and tested H.



Histogram of 1000 bootstrap values of the statistic $\hat{T}^*(\omega) = (\hat{\theta}^*(\omega) - \hat{\theta}(\omega))/\hat{\sigma}^*(\omega)$ at a frequency bin where $\mathsf{H}: \ G_2(\omega) = 0$ retained. The solid line represents a kernel density estimate using 1000 Monte Carlo simulations.

Regression Analysis (Cont'd)



Histogram of 1000 bootstrap values of the statistic $\hat{T}^*(\omega) = (\hat{\theta}^*(\omega) - \hat{\theta}(\omega))/\hat{\sigma}^*(\omega)$ at a frequency bin where the hypothesis H : $G_4(\omega) = 0$ was retained. The solid line represents a kernel density estimate using 1000 Monte Carlo simulations.

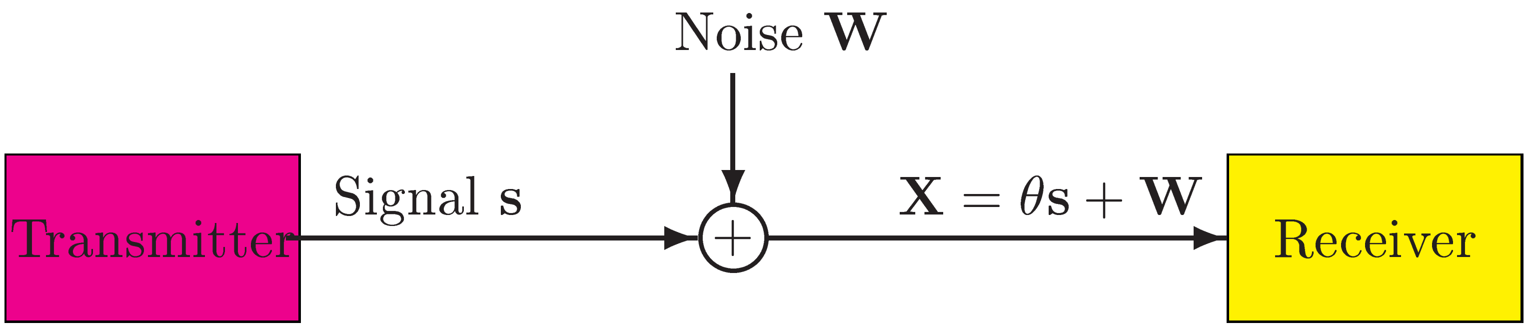
Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Signal Detection

- Detection of signals in interference is a key area in signal processing applications such as radar, sonar, and telecommunications.
- Detection theory is well established when the interference is Gaussian.
- In many applications such as high-resolution radar and radar at low grazing angles interference such as clutter is non-Gaussian.
- Existing methods for the detection of signals in non-Gaussian interference are often cumbersome and/or non-optimal

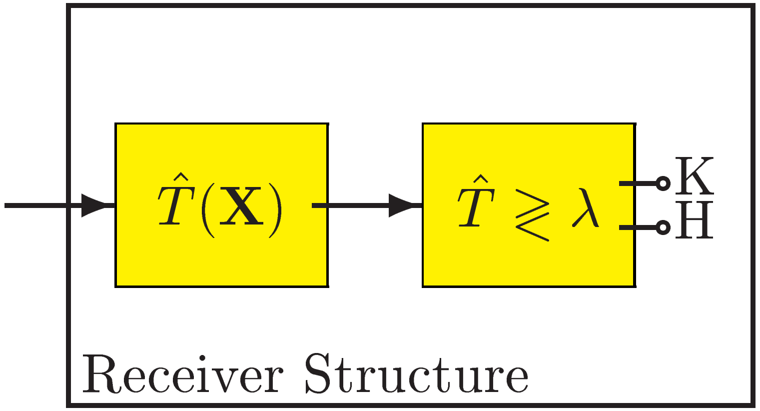
The Signal Detection Problem



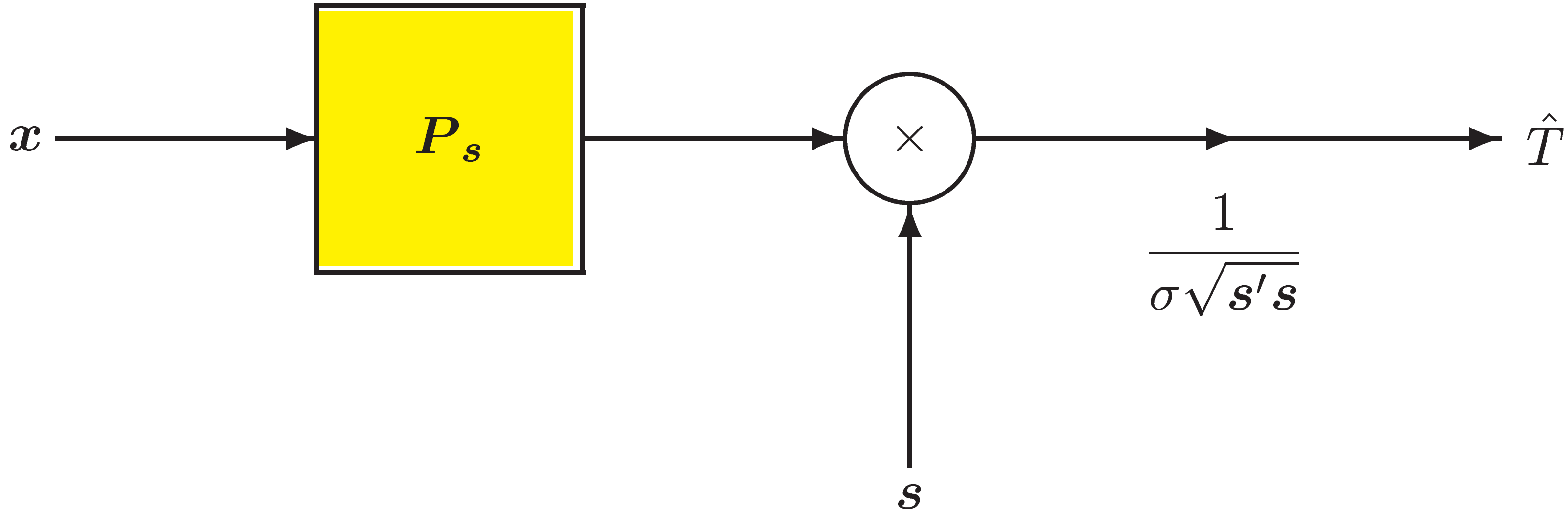
Test the hypotheses:

$$H : \theta = 0$$

$$K : \theta > 0$$

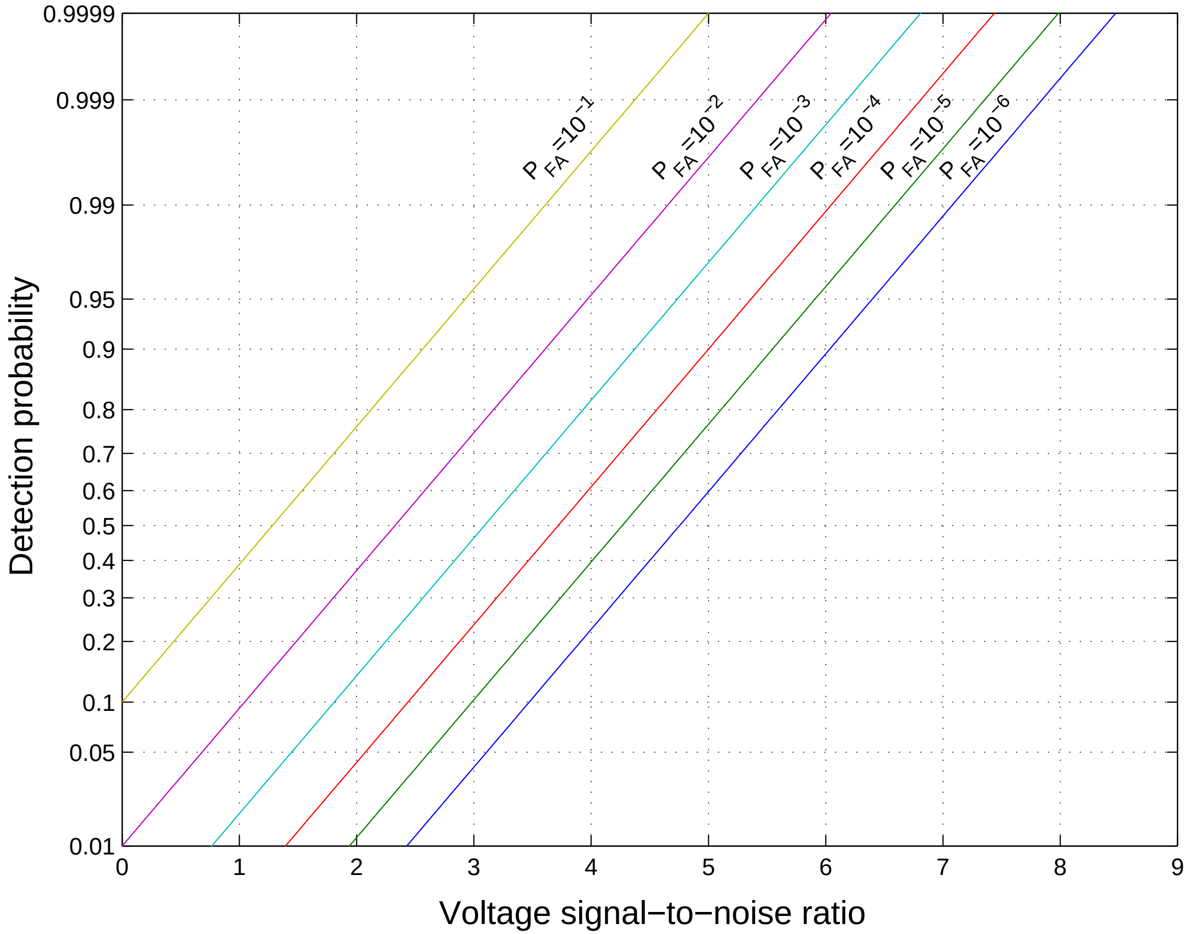


Signal Detection with the Matched Filter

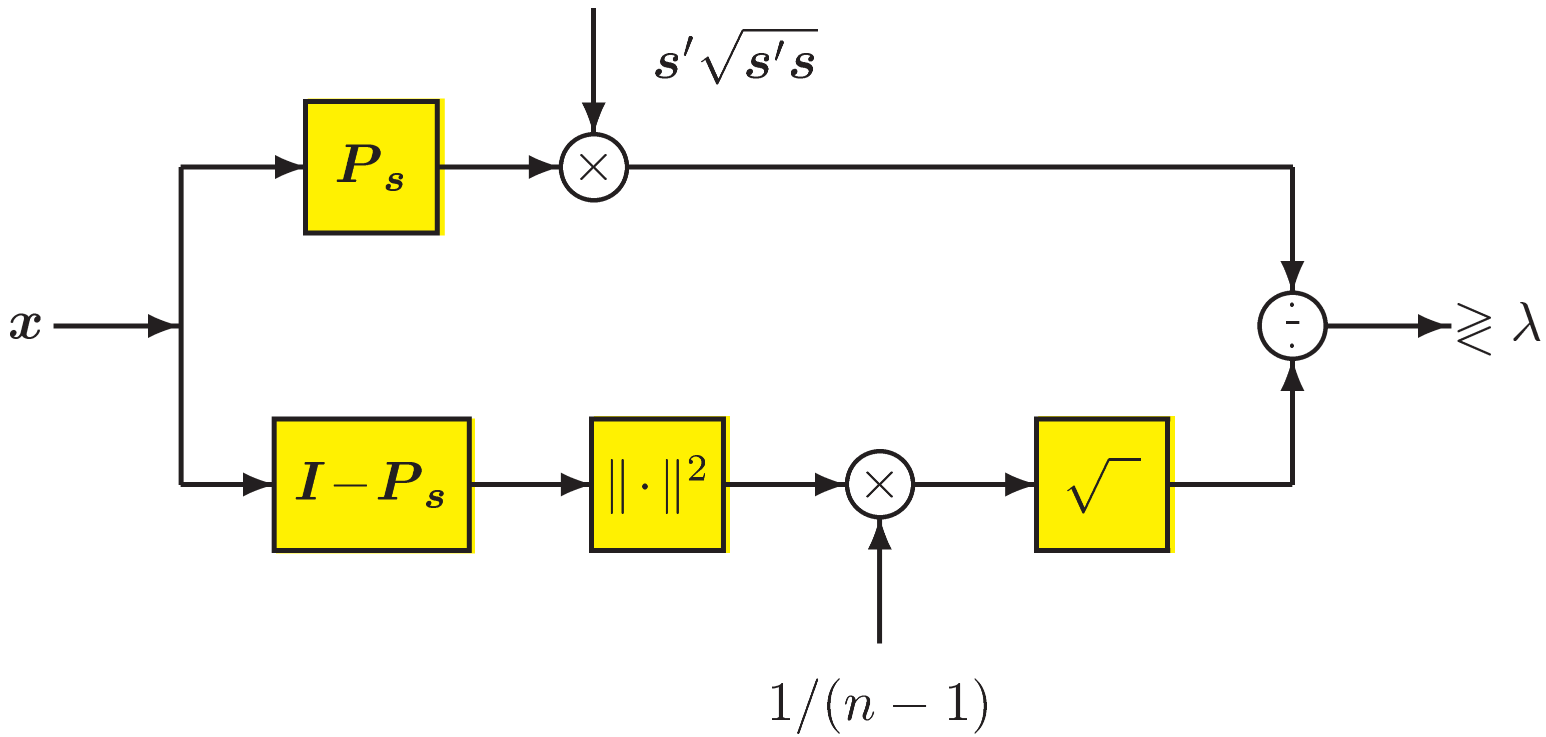


$$\hat{T} = \frac{\mathbf{s}'\mathbf{x}}{\sigma\sqrt{\mathbf{s}'\mathbf{s}}} = \frac{\mathbf{s}'\mathbf{P}_s\mathbf{x}}{\sigma\sqrt{\mathbf{s}'\mathbf{s}}} \sim \mathcal{N}\left(\frac{\theta\sqrt{\mathbf{s}'\mathbf{s}}}{\sigma}, 1\right), \quad \mathbf{P}_s = \mathbf{s}\mathbf{s}'/\mathbf{s}'\mathbf{s}\,.$$

Performance of the Matched Filter

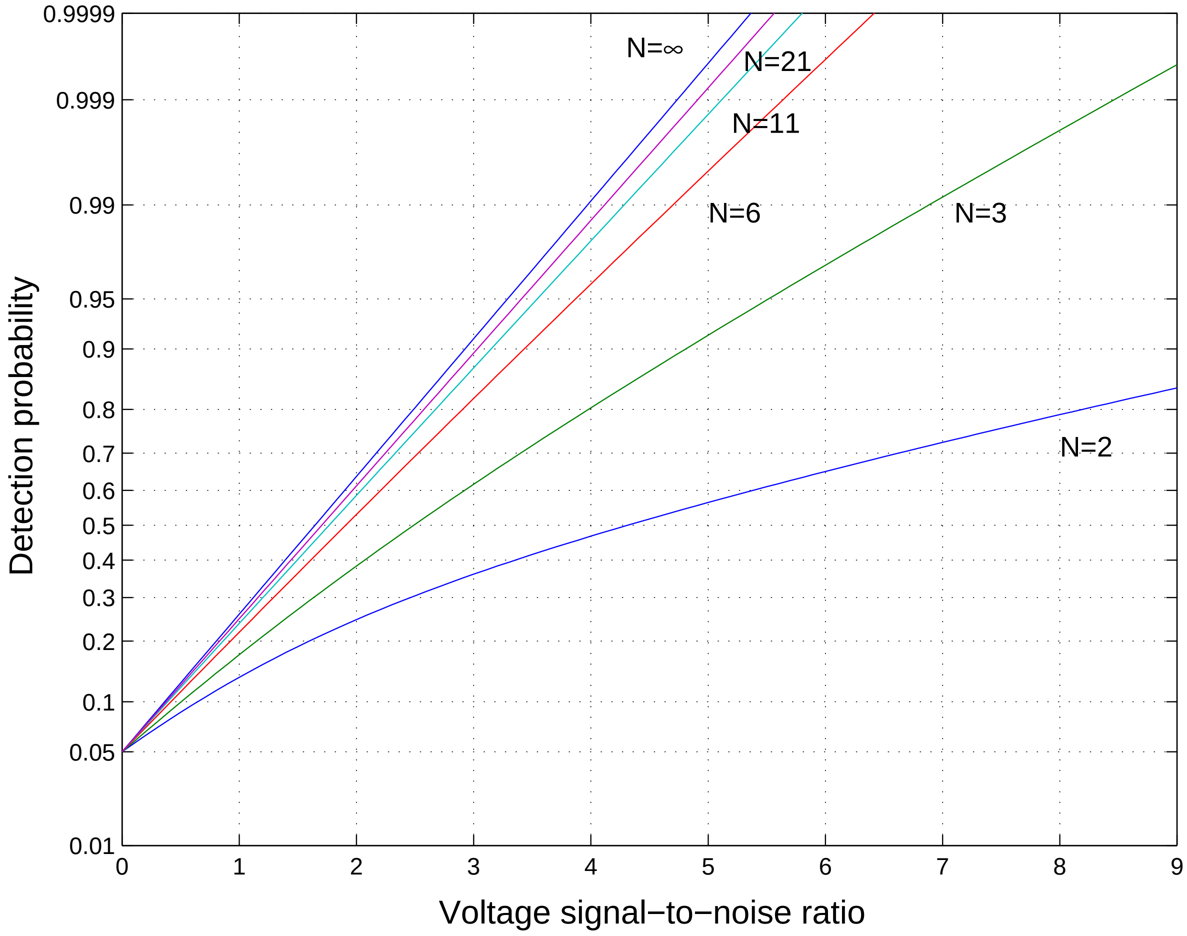


Signal Detection with the CFAR Matched Filter



$$T = \frac{\mathbf{s}'\mathbf{P}_s\mathbf{x}\sqrt{\sigma^2\mathbf{s}'\mathbf{s}}}{\sqrt{\mathbf{x}'(\mathbf{I}-\mathbf{P}_s)\mathbf{x}/(n-1)\sigma^2}} \sim t_{n-1}\left(\frac{\theta\sqrt{\mathbf{s}'\mathbf{s}}}{\sigma}\right)$$

Performance of the CFAR Matched Filter



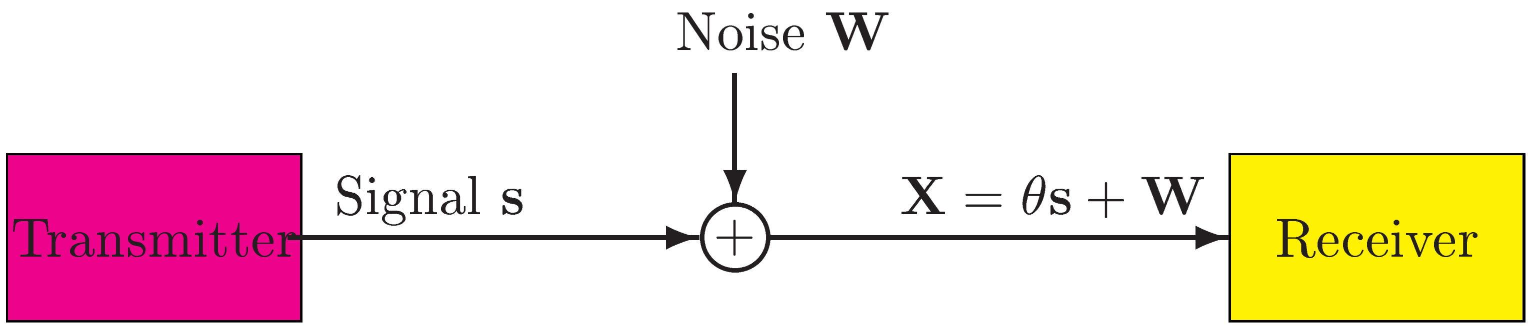
Limitations of the Matched Filter

- The matched filter and the CFAR matched filter are designed (and optimal) for Gaussian interference.
- Although they show high probability of detection in the non-Gaussian case, they are *unable to maintain the preset level of significance* for small sample sizes.
- The matched filter fails in the case where the interference/noise is non-Gaussian and the data size is small.
- The goal is to develop techniques which require little in the way of modelling and assumptions.

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

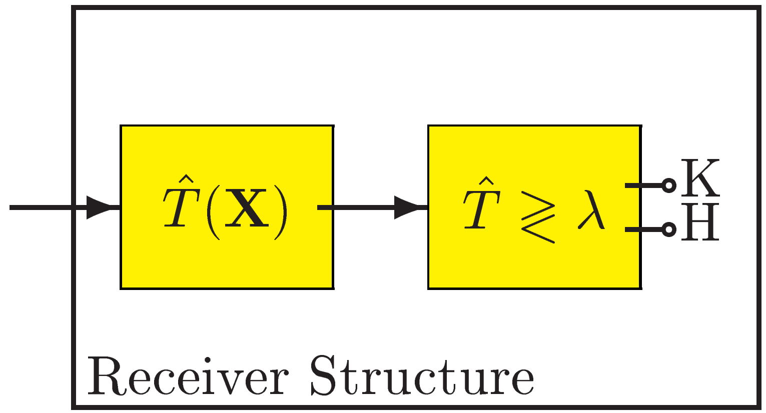
The Signal Detection Problem



Test the hypotheses:

$$\mathrm{H} : \theta = 0$$

$$\mathrm{K} : \theta > 0$$



The Bootstrap Matched Filter

Step 0. *Experiment.* Run the experiment and collect the data x_t, $t = 0, \dots, n-1$.

Step 1. *Estimation.* Compute the LSE $\hat{\theta}$ of θ, $\hat{\sigma}_{\hat{\theta}}$, and

$$\hat{T} = \left. \frac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}} \right|_{\theta=0} .$$

Step 2. *Resampling.* Compute the residuals

$$\hat{w}_t = x_t - \hat{\theta} s_t, \qquad t = 0, \dots, n-1,$$

and after centering, resample the residuals, assumed to be i.i.d., to obtain $\hat{w}_t^*$, $t = 0, \dots, n-1$.

The Bootstrap Matched Filter (Cont'd)

Step 3. *Bootstrap Test Statistic.* Compute new measurements

$$x_t^* = \hat{\theta} s_t + \hat{w}_t^* , \qquad t = 0, \ldots , n-1 ,$$

the LSE $\hat{\theta}^*$ based on the resamples x_t^*, $t = 0, \ldots , n-1$, and

$$\hat{T}^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*_{\hat{\theta}^*}} .$$

Step 4. *Repetition.* Repeat Steps 2 and 3 a large number of times to obtain $\hat{T}_1^*, \ldots , \hat{T}_N^*$.

Step 5. *Bootstrap Test.* Sort $\hat{T}_1^*, \ldots , \hat{T}_N^*$ to obtain $\hat{T}^*_{(1)} \leq \cdots \leq \hat{T}^*_{(N)}$. Reject H if $\hat{T} \geq \hat{T}^*_{(q)}$, where $q = \lfloor (1-\alpha)(N+1) \rfloor$.

Simulation Results (Gaussian Case)

Let $s_t = \sin(2\pi t/6)$, $n = 10$, $N = 999$ (25 for variance estimation), $\alpha = 5\%$, and the number of independent runs be 5,000.

$\mathcal{N}(0,1)$, SNR=7 dB

Detector	$\hat{P}_f$ [%]	$\hat{P}_d$ [%]
Matched Filter (MF)	5	99
CFAR MF	5	98
Boot. (σ known)	4	99
Boot. (σ unknown)	5	98

Simulation Results (Non-Gaussian Case)

Let $s_t = \sin(2\pi t/6)$, $n = 10$, $N = 999$ (25 for variance estimation), $\alpha = 5\%$, and the number of independent runs be 5,000.

$$W_t \sim \sum_{i=1}^{4} a_i \mathcal{N}(\mu_i, \sigma_i^2) \text{ with } \mathbf{a} = (0.5, 0.1, 0.2, 0.2),$$
$$\boldsymbol{\mu} = (-0.1, 0.2, 0.5, 1),\ \boldsymbol{\sigma} = (0.25, 0.4, 0.9, 1.6), \text{ and SNR = -6 dB}$$

Detector	$\hat{P}_f$ [%]	$\hat{P}_d$ [%]
Matched Filter (MF)	**8**	83
CFAR MF	**7**	77
Boot. (σ known)	5	77
Boot. (σ unknown)	**7**	79

Interpretation of the Results

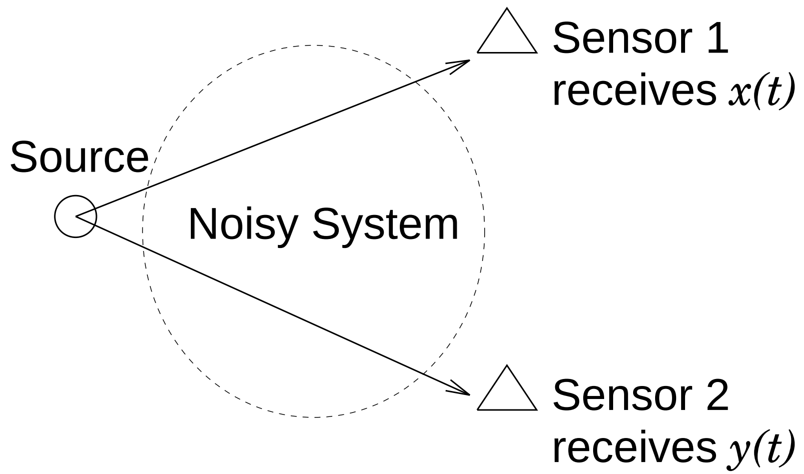
- The bootstrap is able to maintain a constant false alarm while achieving a reasonably high detection power.
- These results can be improved in several ways, and the methods extended to the correlated data case.
- The bootstrap is not proposed as an alternative to existing non-parametric/parametric signal detection schemes in the non-Gaussian interference case.
- The examples suffice to show the power of the bootstrap in signal detection when little is known about the distribution of the interference.

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Detection of a Non-Gaussian Signal Common to Two Sensors

Problem: Detection of a non-Gaussian signal common to two sensors embedded in interference which is either mutually independent at each sensor or have a vanishing cross bispectrum [Tugnait (1993), Ong *et al.* (1997)].



Cross-Bispectral Detection Scheme

- Let x_t and y_t be two jointly-stationary, zero-mean, discrete-time random processes modelling the sensor output signals.
- A detection scheme for the problem is based on testing the cross bispectrum of the sensor output signals for zero, i.e.

$$\begin{array}{ll} \mathsf{H}: & C(j,k) = 0, \\ \mathsf{K}: & C(j,k) \neq 0, \end{array} \qquad \forall (j,k) \in \mathcal{D}',$$

where $C(j,k) = C_{xxy}(2\pi j/n, 2\pi k/n)$, is the cross bispectrum of x_t and y_t at discrete frequencies $\omega_j = 2\pi j/n, \omega_k = 2\pi k/n$, and $\mathcal{D}' = \{(j,k) : |j| < k < n/2\}$.

Limitations with Existing Methods

- Current detection methods [Tugnait (1993)] based on the cross bispectrum assume enough data are available for asymptotic results to apply.
- In some applications these assumptions are not valid leading to degradation in the detector performance.
- Specifically, for small sample sizes, the probability of false alarm is not maintained at the nominal level.

Problem: Seek a solution in the case where the data size is small and asymptotic tests are inapplicable.

Principle of the Bootstrap Procedure

We use the following *approximate* regression

$$I_{xxy}(j,k) = C_{xxy}(j,k) + \varepsilon(j,k)V(j,k),$$

where

$$I_{xxy}(j,k) = \frac{1}{n} d_x(j) d_x(k) \bar{d}_y(j+k)$$

and

$$V(j,k)^2 = n[1+\delta(j-k)]C_{xx}(j)C_{xx}(k)C_{yy}(j+k).$$

This regression is consistent with asymptotic results.

Bootstrap Procedure (Cont'd)

Step 1. Calculate $I^{(i)}(j,k)$, $i = 1,\ldots,P$ and $\hat{C}(j,k) = \frac{1}{P}\sum_{i=1}^{P} I^{(i)}(j,k)$, $(j,k) \in \mathcal{D}'$.

Step 2. Form residuals of the regression

$$\hat{\varepsilon}^{(i)}(j,k) = \frac{I^{(i)}(j,k) - \hat{C}(j,k)}{\hat{V}^{(i)}(j,k)}$$

Step 3. Repeat N times (after mean-subtracting $\hat{\varepsilon}^{(i)}(j,k)$):

$$\begin{aligned} I_b^{(i)*}(j,k) &= \hat{C}(j,k) + \hat{\varepsilon}_b^{(i)*}(j,k)\hat{V}^{(i)}(j,k) \\ \hat{C}_b^{*}(j,k) &= \frac{1}{P}\sum_{i=1}^{P} I_b^{(i)*}(j,k), \qquad b = 1,\ldots,N. \end{aligned}$$

Bootstrap Procedure (Cont'd)

Step 4. Calculate test statistic

$$\hat{T} = \sum_{(j,k)\in\mathcal{D}'} \left| \frac{\hat{C}(j,k) - C_0(j,k)}{\hat{\sigma}(j,k)} \right| .$$

Step 5. Calculate the bootstrap statistics

$$\hat{T}_b^* = \sum_{(j,k)\in\mathcal{D}'} \left| \frac{\hat{C}_b^*(j,k) - \hat{C}(j,k)}{\hat{\sigma}_b^*(j,k)} \right| , \quad b = 1, \dots, N .$$

Step 6. Rank $\hat{T}_1^*, \dots, \hat{T}_N^*$ to obtain $\hat{T}_{(1)}^* \leq \dots \leq \hat{T}_{(N)}^*$.

Step 7. Reject the null hypothesis if $\hat{T} > \hat{T}^*_{(\lfloor (N+1)(1-\alpha)\rfloor)}$.

GLRT Simulation Results

Interference		% False Alarms	% Detected
Common Gaussian interference	i.i.d.	5	99
	AR(1)	6	99
	AR(5)	6	99
Independent exponential interference	i.i.d.	**31**	99
	AR(1)	**30**	99
	AR(5)	**28**	99

Detection results for a common MA(10) exponential signal in Gaussian and non-Gaussian interference for $n = 512$ and $\alpha = 5\%$.

Bootstrap Simulation Results (Cont'd)

Interference		% False Alarms	% Detected
Common	i.i.d.	4	93
Gaussian	AR(1)	1	94
interference	AR(5)	1	93
Independent	i.i.d.	**8**	94
exponential	AR(1)	**1**	93
interference	AR(5)	**1**	95

Detection results for a common MA(10) exponential signal in Gaussian and non-Gaussian interference for $n = 512$ and $\alpha = 5\%$.

Detection of a Non-Gaussian Signal Common to Multiple Sensors

Problem: How can the detection that uses measurements from two sensors be extended when measurements are available from multiple sensors so that a higher probability of detection is achieved?

Relevance: This problem is important in array processing applications such as in sonar, radar, and communications.

Solution: We propose the use of the Bonferroni test of multiple hypotheses coupled with a bootstrap method [Ong & Zoubir (1997)].

Multiple Tests

- Let $x_{1,t}, \ldots, x_{L+1,t}$, $t = 0, \ldots, n-1$, be the discrete-time measurements from $L+1$ sensors.
- They are assumed to be zero-mean, jointly stationary, random sequences.
- A non-Gaussian signal common to any two adjacent sensors can be detected by testing the multiple hypotheses,

$$\begin{aligned} \mathsf{H}_l: &\quad C_{x_l x_l x_{l+1}}(j,k) \equiv 0, \\ \mathsf{K}_l: &\quad C_{x_l x_l x_{l+1}}(j,k) \not\equiv 0, \end{aligned} \qquad l = 1, \ldots, L,$$

where $C_{x_l x_l x_{l+1}}(j,k)$ is the cross bispectrum of $x_{l,t}$ and $x_{l+1,t}$ evaluated at discrete bifrequencies $(2\pi j/n, 2\pi k/n)$.

Multiple Tests (Cont'd)

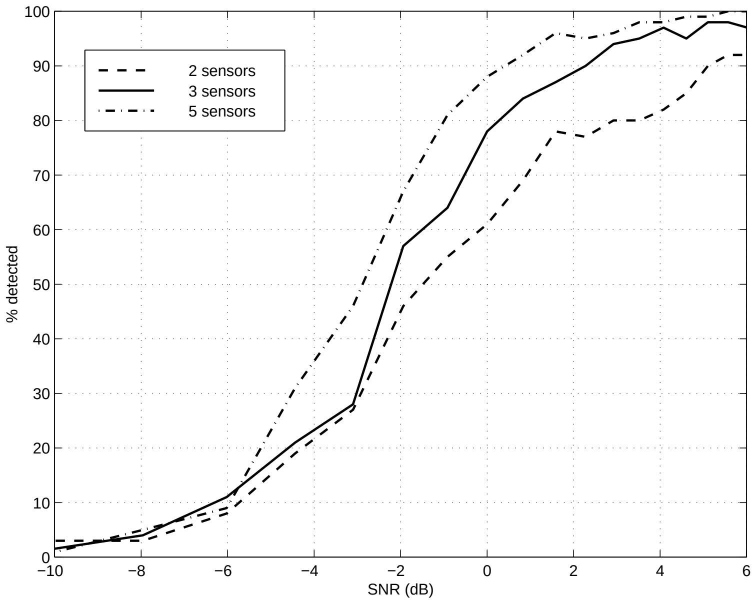
1. Apply a detection method for two sensors on $x_{l,t}$ and $x_{l+1,t}$ for $l = 1, \ldots, L$.
2. Form a set of L p-values, $P_1, \ldots, P_L$.
3. Compare the minimum p-value, $P_{(1)}$, to α/L, where α is the nominal test level.
4. If $P_{(1)} < \alpha/L$, conclude that a non-Gaussian signal is present in at least two adjacent sensors.

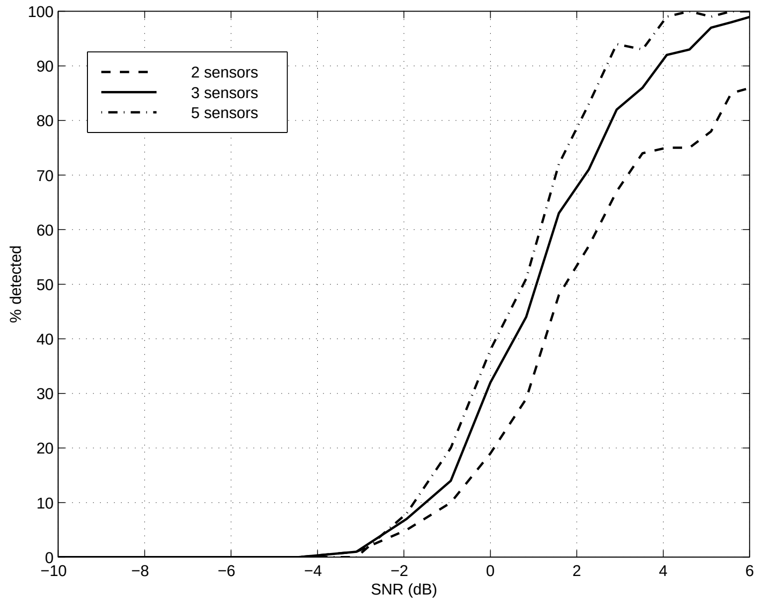
Using the Bonferroni level, α/L, limits the global level to α.

This procedure is equivalent to running two sensor tests with α/L instead of α and rejecting if any test rejects.

Simulation Results

Detection of an MA(10) exponential signal in AR(1) (left plot) and AR(5) (right plot) Gaussian interference (512 data points):

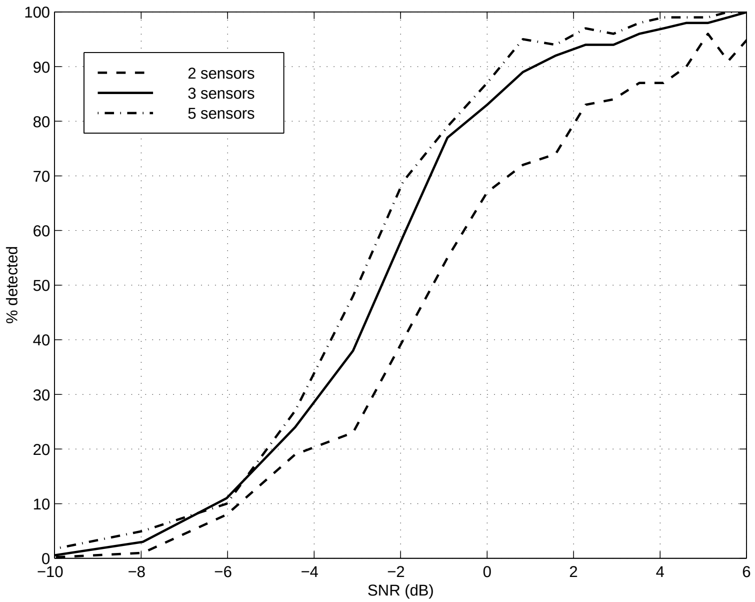


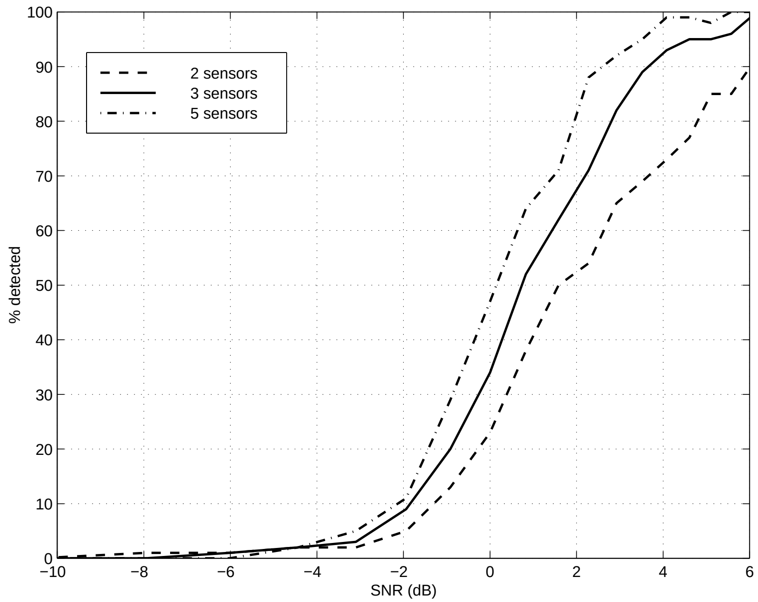


In all cases, % false alarms = 0 (100 simulation runs)

Simulation Results (Cont'd)

Detection of an MA(10) exponential signal in AR(1) (left plot) and AR(5) (right plot) **exponential** interference (512 data points):





In all cases, % false alarms < 3 (100 simulation runs)

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Model Selection

- Several model selection procedures exist. They include Akaike's information criterion, Rissanen's minimum description length criterion, and Hannan and Quinn's criterion.
- Bootstrap methods for model selection are simple and computationally efficient.
- If one uses a bootstrap approach for the model selection and for the subsequent inference, then the bootstrap observations generated for model selection can also be used in the inference procedure.
- Thus, the model selection procedure can be done at no extra computational cost.

Outline

- Hypothesis Testing with the Bootstrap
  - An Example: Testing the Frequency Response for Zero
- Signal Detection
  - Examples
    1. Bootstrap Matched-Filter
    2. Detection of a Non-Gaussian Signal at Multiple Sensors
- Model Selection
  - Linear Models

Model Selection in Linear Models

Consider the linear model

$$Y_t = \boldsymbol{x}_t'\boldsymbol{b} + Z_t, \qquad t = 0, \ldots, n-1,$$

where Z_t is a set of i.i.d. random variables of unknown distribution with $\mu_Z = 0$ and $\sigma_Z^2 > 0$, and $\boldsymbol{b}$ is an unknown p-vector parameter. Here, $\boldsymbol{x}_t$ is the t-th value of the p vector of explanatory variables, assumed to be known.

With $\boldsymbol{Y} = (Y_0, \ldots, Y_{n-1})'$, $\boldsymbol{x} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1})'$, $\boldsymbol{b} = (b_1, \ldots, b_p)'$ and $\boldsymbol{Z} = (Z_0, \ldots, Z_{n-1})'$, we have

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{b} + \boldsymbol{Z}\,.$$

Model Selection in Linear Models (Cont'd)

With β being as subset of $\{1, \ldots, p\}$, a model corresponding to β is given by

$$\boldsymbol{Y} = \boldsymbol{x}_\beta \boldsymbol{b}_\beta + \boldsymbol{Z}\,,$$

where $\boldsymbol{b}_\beta$ is a sub-vector of $\boldsymbol{b}$ containing the components of $\boldsymbol{b}$ indexed by integers in β and $\boldsymbol{x}_\beta$ is a matrix containing the columns of $\boldsymbol{x}$ indexed by integers in β.

Problem: Estimate β_0 based on $y_0, \ldots, y_{n-1}$, where β_0 is such that $\boldsymbol{b}_{\beta_0}$ contains all non-zero components of $\boldsymbol{b}$ only.

Model Selection in Linear Models (Cont'd)

We consider an estimator of the mean-squared error given by

$$\Gamma_n(\beta) = \frac{1}{n}\sum_{t=0}^{n-1}\left(Y_t - \boldsymbol{x}'_{\beta t}\hat{\boldsymbol{b}}_\beta\right)^2 = \frac{\|\boldsymbol{Y} - \boldsymbol{x}_\beta\hat{\boldsymbol{b}}_\beta\|^2}{n}$$

with $\boldsymbol{x}'_{\beta t}$ being the t-th row of $\boldsymbol{x}_\beta$ and $\hat{\boldsymbol{b}}_\beta$ the LSE for $\boldsymbol{b}_\beta$. Then,

$$\mathsf{E}[\Gamma_n(\beta)] = \sigma_Z^2 - \frac{\sigma_Z^2 p_\beta}{n} + \Delta_n(\beta),$$

where $\Delta_n(\beta) = n^{-1}\boldsymbol{\mu}'(\boldsymbol{I} - \boldsymbol{x}_\beta(\boldsymbol{x}'_\beta\boldsymbol{x}_\beta)^{-1}\boldsymbol{x}'_\beta)\boldsymbol{\mu}$, with $\boldsymbol{\mu} = \mathsf{E}[\boldsymbol{Y}]$, $\boldsymbol{h}_\beta = \boldsymbol{x}_\beta(\boldsymbol{x}'_\beta\boldsymbol{x}_\beta)^{-1}\boldsymbol{x}'_\beta$ being the $p \times p$ projection matrix and p_β is the size of $\boldsymbol{b}_\beta$. If β is correct, then $\Delta_n(\beta) = 0$.

Model Selection: Minimise $\mathsf{E}[\Gamma_n(\beta)]$ over β.

Model Selection in Linear Models (Cont'd)

- A bootstrap model selection approach would minimise over β

$$\tilde{\Gamma}_n(\beta) = \frac{1}{n}\sum_{t=0}^{n-1} \mathsf{E}_* \left(y_t - \boldsymbol{x}'_{t\beta}\hat{\boldsymbol{b}}^*_\beta \right)^2 = \mathsf{E}_* \left[\frac{\|\boldsymbol{y} - \boldsymbol{x}_\beta \hat{\boldsymbol{b}}^*_\beta\|^2}{n} \right],$$

where $\hat{\boldsymbol{b}}^*_\beta$ is the LSE based on $(y^*_t, \boldsymbol{x}_{\beta t})$.

- The estimator $\tilde{\Gamma}_n(\beta)$ is biased. A better estimator is given by

$$\hat{\Gamma}^*_{n,m}(\beta) = \mathsf{E}_* \left[\frac{\|\boldsymbol{y} - \boldsymbol{x}_\beta \hat{\boldsymbol{b}}^*_{\beta,m}\|^2}{n} \right],$$

where $\hat{\boldsymbol{b}}^*_{\beta,m}$ is obtained from $y^*_t = \boldsymbol{x}'_{\beta t}\hat{\boldsymbol{b}}_\beta + \hat{z}^*_t$, $t = 0, \ldots, n-1$, with $\hat{z}^*_t$ being a resample from $\sqrt{n/m}(\hat{z}_t - \hat{z}_\bullet)/\sqrt{1-p/n}$ and $\hat{z}_\bullet = n^{-1}\sum_{t=0}^{n-1}\hat{z}_t$.

Model Selection in Linear Models (Cont'd)

Step 1. Based on $y_0, \ldots, y_{n-1}$, compute the LSE $\hat{\boldsymbol{b}}$ and $\hat{z}_t = y_t - \boldsymbol{x}'_{\alpha t}\hat{\boldsymbol{b}}_\alpha$, $t = 0, \ldots, n-1$, where $\alpha = \{1, \ldots, p\}$.

Step 2. Resample with replacement from $\sqrt{n/m}(\hat{z}_t - \hat{z}_\bullet)/\sqrt{1-p/n}$ to obtain $\hat{z}_t^*$, where $\frac{m}{n} \to 0$ and $\frac{n}{m}\max_{t\leq n} h_{\beta t} \to 0$ for all β in the class of models to be selected.

Step 3. Compute $y_t^* = \boldsymbol{x}'_{\beta t}\hat{\boldsymbol{b}}_\beta + \hat{z}_t^*$, $t = 0, \ldots, n-1$ and the LSE $\hat{\boldsymbol{b}}^*_{\beta,m}$ from $(y_t^*, \boldsymbol{x}_{\beta t})$.

Step 4. Repeat Steps 2-3 to obtain $\hat{\boldsymbol{b}}^{*(i)}_{\beta,m}$ and $\hat{\Gamma}^{*(i)}_{n,m}(\beta)$ for $i = 1, \ldots, N$.

Step 5. Average $\hat{\Gamma}^{*(i)}_{n,m}(\beta)$ over $i = 1, \ldots, N$ and minimise over β to obtain $\hat{\beta}_0$.

Example: Trend Estimation

Let $\boldsymbol{x}_t = (1, t, \ldots, t^p)$, $t = 0, \ldots, 63$ and $\boldsymbol{b} = (0, 0, 0.035, -0.0005)'$. We simulate $Y_t = \boldsymbol{x}_t'\boldsymbol{b} + \boldsymbol{Z}_t$ by adding standard normal and t_3-distributed noise. With $N = 100$ and $m = 2$ we obtain the following results (based on 1,000 simulations).

	$\mathcal{N}(0,1)$			t_3		
Model β	$\hat{\Gamma}^*$	AIC	MDL	$\hat{\Gamma}^*$	AIC	MDL
$(0, 0, b_2, b_3)$	**100**	**91**	**98**	**99**	**89**	**98**
$(0, b_1, b_2, b_3)$	0	5	1	1	5	1
$(b_0, 0, b_2, b_3)$	0	3	1	0	3	1
(b_0, b_1, b_2, b_3)	0	2	0	0	3	0

Empirical probabilities (%), excluding models not selected by any method.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Model Selection in Non-Linear Models

We define a nonlinear model by

$$Y_t = g(\boldsymbol{x}_t, \boldsymbol{b}) + Z_t, \qquad t = 0, \ldots, n-1,$$

where Z_t is a noise sequence of i.i.d. random variables of unknown distribution with $\mu_Z = 0$ and $\sigma_Z^2 > 0$, and g is a known function.

Define the collection of subsets of $\alpha = \{1, \ldots, p\}$ by $\mathcal{B}$ and $g_{\beta t}(\boldsymbol{b}_\beta) = g_\beta(\boldsymbol{x}_{\beta t}, \boldsymbol{b}_\beta)$, where $\beta \in \mathcal{B}$ and g_β is the restriction of the function g to the admissible set of $(\boldsymbol{x}_{\beta t}, \boldsymbol{b}_\beta)$. Let $\tilde{\mathcal{B}}$ be the admissible set for $\boldsymbol{b}$.

With $\dot{\boldsymbol{g}}(\gamma) = \frac{\partial g(\gamma)}{\partial \gamma}$ and $\boldsymbol{m}_\beta(\gamma) = \sum_{t=0}^{n-1} \dot{\boldsymbol{g}}_{\beta t}(\gamma) \dot{\boldsymbol{g}}_{\beta t}(\gamma)'$, a consistent bootstrap procedure for selecting β is as follows.

Model Selection in Non-Linear Models (Cont'd)

Step 1. With y_t, $t = 0, \ldots, n-1$, find $\hat{\boldsymbol{b}}_\alpha$, the solution of $\sum_{t=0}^{n-1} (y_t - g_{\alpha t}(\boldsymbol{\gamma}))\, \dot{\boldsymbol{g}}_{\alpha t}(\boldsymbol{\gamma}) = 0$, for all $\boldsymbol{\gamma} \in \tilde{\mathcal{B}}$ and the residuals $\hat{z}_t = y_t - g_{\alpha t}(\hat{\boldsymbol{b}}_\alpha)$, $t = 0, \ldots, n-1$.

Step 2. Get $\hat{z}_t^*$ by resampling $\sqrt{n/m}(\hat{z}_t - \hat{z}_\bullet)/\sqrt{1-p/n}$.

Step 3. Compute $\hat{\boldsymbol{b}}^*_{\beta,m} = \hat{\boldsymbol{b}}_\beta + \boldsymbol{m}_\beta(\hat{\boldsymbol{b}}_\beta)^{-1} \sum_{t=0}^{n-1} \hat{z}_t^* \dot{\boldsymbol{g}}_{\beta t}(\hat{\boldsymbol{b}}_\beta)$.

Step 4. Repeat Steps 2-3 to obtain $\hat{\boldsymbol{b}}^{*(i)}_{\beta,m}$, $i = 1, \ldots, N$.

Step 5. To find $\hat{\beta}_0$, minimise over β

$$\hat{\Gamma}^*_{n,m}(\beta) = N^{-1} \sum_{i=1}^{N} \sum_{t=0}^{n-1} \frac{\left(y_t - g_{\beta t}(\hat{\boldsymbol{b}}^{*(i)}_{\beta,m})\right)^2}{n}.$$

Example: Oscillations in Noise

Let $Y_t = \cos\omega_1 t(1+\cos\omega_2 t) + Z_t$, $t = 0, \ldots, 39$. In this case $\mathcal{B} = \{\beta_k, k = 1,2,3\}$ so that $g_{\beta_1 t}(\boldsymbol{b}_{\beta_1}) = 2\cos\omega_1 t$ $(\omega_2 = 0)$, $g_{\beta_2 t}(\boldsymbol{b}_{\beta_2}) = 1 + \cos\omega_2 t$ $(\omega_1 = 0)$, and $g_{\beta_3 t}(\boldsymbol{b}_{\beta_3}) = \cos\omega_1 t(1+\cos\omega_2 t)$ $(\omega_1, \omega_2 \neq 0)$.

We chose $\omega_1 = 0.2\pi$ and $\omega_2 = 0.1\pi$ and run simulations at -1.2 dB SNR with $m = 35$.

Method	β_1	β_2	β_3
Bootstrap	3	0	**97**
AIC	0	3	**97**
MDL	0	5	**95**

Empirical probabilities (%) based on 100 simulations.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Order Selection in Autoregressive Models

Consider

$$Y_t = b_1 Y_{t-1} + b_2 Y_{t-2} + \cdots b_p Y_{t-p} + Z_t, \quad t \in \mathbb{Z},$$

where p is the order, b_k, $k = 1, \ldots, p$, are unknown parameters and Z_t are i.i.d. random variables with $\mu_Z = 0$ and $\sigma_Z^2 > 0$. Let $(y_{-p}, \ldots, y_{-1}, y_0, \ldots, y_{n-1})$ be observations and $\hat{\boldsymbol{b}}$ the LSE of $\boldsymbol{b} = (b_1, \ldots, b_p)'$.

Select a model β from $\mathcal{B} = \{1, \ldots, p\}$ where *each* β *corresponds to the autoregressive model of order* β, i.e., $Y_t = b_1 Y_{t-1} + \cdots b_\beta Y_{t-\beta} + Z_t$.

Objective: Find the optimal order, i.e., $\beta_0 = \max\{k : 1 \le k \le p, b_k \ne 0\}$, where p is the largest order.

Order Selection in Autoregressive Models (Cont'd)

Step 1. Resample the residuals $(\hat{z}_t - \hat{z}_\bullet)$ to obtain $\hat{z}_t^*$.

Step 2. Find $\hat{\boldsymbol{b}}^*_{\beta,m}$ the LSE of $\boldsymbol{b}_\beta$ under β from $y_t^* = \sum_{k=1}^{\beta} \hat{b}_k y^*_{t-k} + \hat{z}_t^*$ for $t = -p, \ldots, m-1$, with m replacing n and $\left\{y^*_{-p}, \ldots, y^*_0\right\}$ replacing $\left\{y^*_{-2p}, \ldots, y^*_{-p-1}\right\}$.

Step 3. Repeat Steps 1-2 to obtain $\hat{\boldsymbol{b}}^{*(1)}_{\beta,m}, \ldots, \hat{\boldsymbol{b}}^{*(N)}_{\beta,m}$ and

$$\hat{\Gamma}^*_{n,m}(\beta) = N^{-1} \sum_{i=1}^{N} \sum_{t=0}^{n-1} \frac{\left(y_t - \sum_{k=1}^{\beta} y_{t-k+1} \hat{b}^{*(i)}_{k,m}\right)^2}{n}$$

Step 5. Minimise $\hat{\Gamma}^*_{n,m}(\beta)$ over β to find $\hat{\beta}_0$.

The procedure is consistent for $m \to \infty$ and $m/n \to 0$ as $n \to \infty$.

Example: Order Selection in an AR Model

We consider determining the order of the process described by

$$Y_t = -0.4Y_{t-1} + 0.2Y_{t-2} + Z_t, \quad t \in \mathbb{Z},$$

where Z_t is a standard Gaussian variable. With $n = 128$ we obtained:

Method	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$
Bootstrap	28.0	**65.0**	5.0	2.0
AIC	17.8	**62.4**	12.6	7.2
MDL	43.2	**54.6**	2.1	0.1

Empirical Probabilities (%) of selecting the true AR model, $\beta_0 = 2$, $n = 128$ and $m = 40$, based on 1,000 simulations.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

More Applications in Signal Processing

The relatively simple examples showed that the bootstrap is powerful. This suggests its use in real-life applications and more complex problems. We show now how the bootstrap can be applied for:

1. Confidence Intervals for Spectra [Franke & Härdle (1992), Politis *et al.* (1992), Zoubir & Iskander (1996)]
2. Bispectrum Based Gaussianity Tests [Zoubir & Iskander (1996,1999)]
3. Noise Floor Estimation in Radar [Zoubir & Boashash (1996)]
4. Confidence Intervals for Flight Parameters in Passive Sonar [Reid *et al.* (1996), Zoubir & Boashash (1998)].
5. Model Selection of Polynomial Phase Signals [Zoubir & Iskander (1998,1999)].

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Confidence Intervals for Spectra

- Several methods for spectrum estimation exist [Brillinger (1981), Priestley (1981), Marple (1987), Kay (1989)].
- It is often desirable to provide a confidence interval for the spectrum based on the estimate as an accuracy measure.
- We present three different methods based on the bootstrap to estimate confidence bands for the power spectrum.
- These methods are compared with the chi-squared approximation for both Gaussian and non-Gaussian weakly dependent time series.

Confidence Intervals for Spectra (Cont'd)

Let $X_0, \ldots, X_{n-1}$ be observations from a strictly stationary real-valued time series X_t with $\mu_X = 0$, $\sigma_X^2 > 0$ and spectral density

$$C_{XX}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \mathsf{E}\, X_0\, X_{|\tau|}\, e^{-j\,\tau\,\omega}.$$

Denote the periodogram [Brillinger (1981), Marple(1987)] by

$$I_{XX}(\omega) = \frac{1}{2\pi n} \left| \sum_{k=0}^{n-1} X_k\, e^{j\,k\,\omega} \right|^2, \qquad -\pi \leq \omega \leq \pi\,.$$

Confidence Intervals for Spectra (Cont'd)

We will consider a kernel estimate of $C_{XX}(\omega)$

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

- $K(\theta)$ is symmetric and nonnegative (here Bartlett-Priestley window [Priestley (1981)]),
- h is its bandwidth,
- M denotes the largest integer $\leq n/2$
- $\omega_k = 2\pi k/n$, $-M \leq k \leq M$.

Confidence Intervals for Spectra (Cont'd)

Asymptotically for large n, $\hat{C}_{XX}(\omega_1), \ldots, \hat{C}_{XX}(\omega_M)$ are independent variates with distribution $C_{XX}(\omega_k)\, \chi^2_{4\,m+2}/(4\,m+2)$, $k = 1, \ldots, M$ with $m = \lfloor (h\,n - 1)/2 \rfloor$.

A $100\alpha\%$ confidence interval [Brillinger (1981)] is given by

$$\frac{(4\,m+2)\,\hat{C}_{XX}(\omega;\, h)}{\chi^2_{4\,m+2}\left(\frac{1+\alpha}{2}\right)} < C_{XX}(\omega) < \frac{(4\,m+2)\,\hat{C}_{XX}(\omega;\, h)}{\chi^2_{4\,m+2}\left(\frac{1-\alpha}{2}\right)},$$

where $\chi^2_\nu(\alpha)$ is such that $\mathsf{Pr}\left[\chi^2_\nu < \chi^2_\nu(\alpha)\right] = \alpha$.

An alternative method proposed by Franke & Härdle (1992) is presented below. It exploits the approximate regression $\varepsilon_k = I_{XX}(\omega_k)/C_{XX}(\omega_k)$, $k = 1, \ldots, M$.

Confidence Intervals for Spectra (Cont'd)

Step 1. *Compute Residuals.* Choose an $h_i > 0$ which does not depend on ω and compute

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

Step 2. *Rescaling.* Rescale the empirical residuals to

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

Step 3. *Resampling.* Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_M^*$ from the empirical distribution of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_M$.

Confidence Intervals for Spectra (Cont'd)

Step 4. *Bootstrap estimates.* With a bandwidth g, find

$$\begin{aligned} I^*_{XX}(\omega_k) &= I^*_{XX}(-\omega_k) = \hat{C}_{XX}(\omega_k; g)\,\tilde{\varepsilon}^*_k, \quad k = 1, \ldots, M, \\ \hat{C}^*_{XX}(\omega; h) &= \frac{1}{n\,h} \sum_{k=-M}^{M} K\left(\frac{\omega - \omega_k}{h}\right) I^*_{XX}(\omega_k). \end{aligned}$$

Step 5. *Confidence bands estimation.* Repeat Steps 3 – 4 and find c^*_U (and proceed similarly for c^*_L) such that

$$\mathsf{Pr}_*\left(\sqrt{n\,h}\,\frac{\hat{C}^*_{XX}(\omega; h) - \hat{C}_{XX}(\omega; g)}{\hat{C}_{XX}(\omega; g)} \le c^*_U\right) = \alpha\,.$$

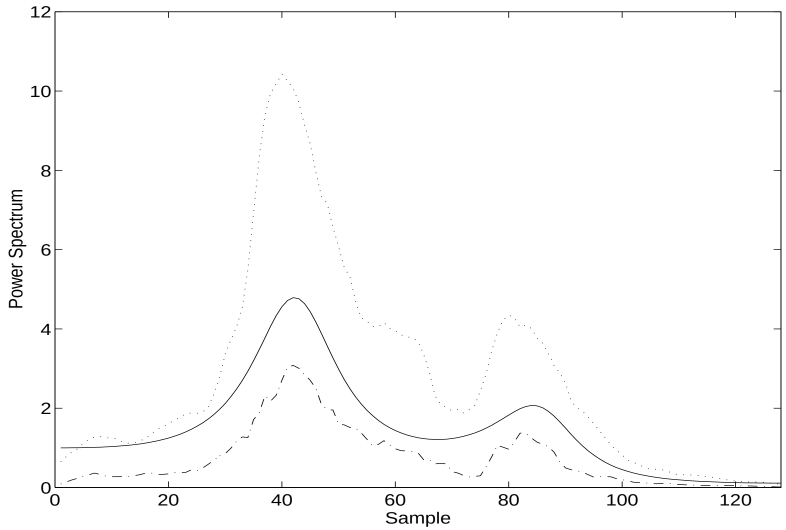
That is $\{1 + c^*_U(n\,h)^{-1/2}\}^{-1}\hat{C}_{XX}(\omega; h)$ is the upper bound of an $(1-2\alpha)\%$-confidence interval for $C_{XX}(\omega)$.

Simulation Results

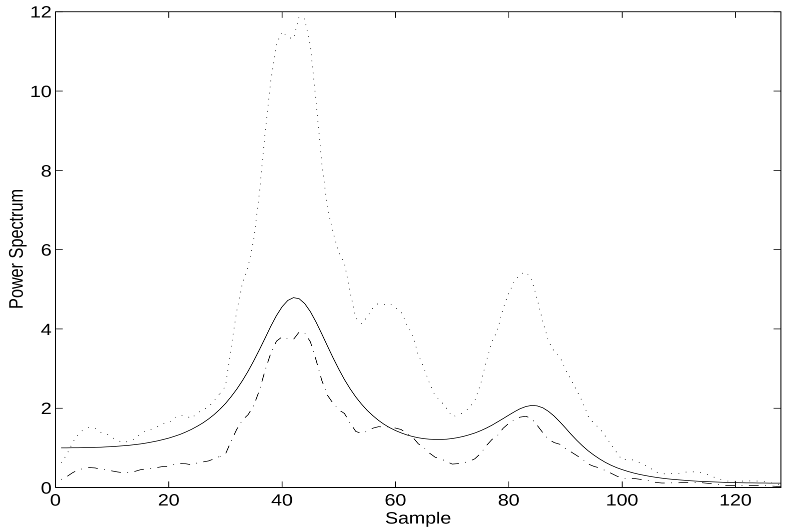
Consider the AR process

$$X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + \varepsilon_t \,,$$

where ε_t is an i.i.d. $\mathcal{N}(0,1)$ process. With $n = 256$, $N = 399$ and $h = 0.1$, we obtain the following 95% confidence interval.



Residuals based method



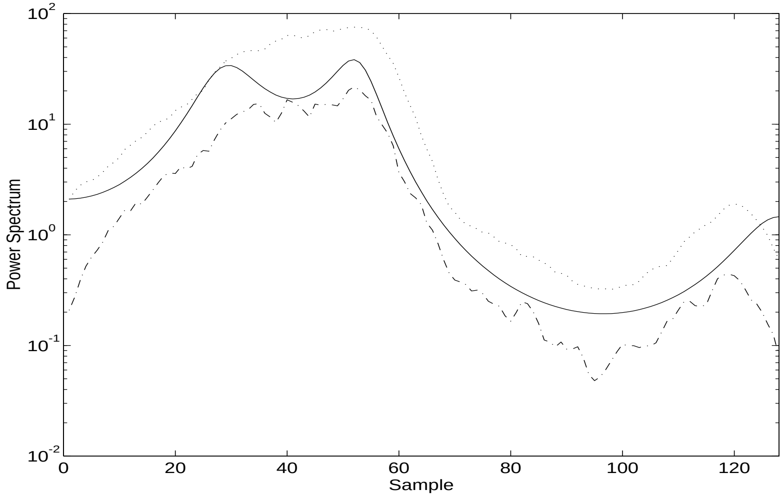
χ^2 approximation

Simulation Results (Cont'd)

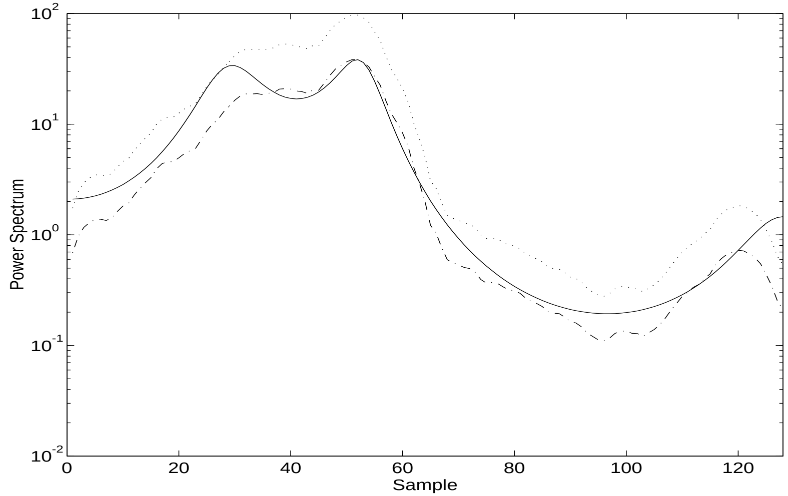
Consider the AR process

$$Y_t = Y_{t-1} - 0.7Y_{t-2} - 0.4Y_{t-3} + 0.6Y_{t-4} - 0.5Y_{t-5} + \zeta_t \, ,$$

where ζ_t is an i.i.d. $\mathcal{U}(-2.5, 2.5)$ process. With $n = 256$, $N = 399$ and $h = 0.1$, we obtain the following 95% confidence interval.



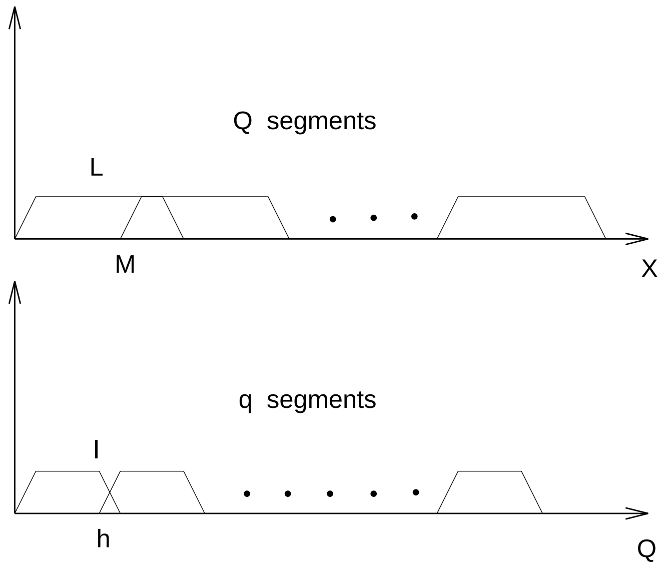
Residuals based method



χ^2 approximation

The Block of Blocks Bootstrap

The *block of blocks* bootstrap initially suggested in [Künsch (1989)] was proposed for setting confidence bands for spectra in [Politis & Romano (1992)]. An application of the block of blocks bootstrap to higher-order cumulants can be found in [Zhang *et al.* (1993)].



Principle of the block of blocks bootstrap

Block of Blocks Bootstrap for Spectra

Step 1. *First block.* Given $X_0, \ldots, X_{n-1}$, obtain Q overlapping $(0 < M < L)$ or non-overlapping $(M = 0)$ segments of L samples and estimate $\hat{C}_{XX}^{(i)}(\omega)$, $i = 1, \ldots, Q$.

Step 2. *Second block.* Divide $\hat{C}_{XX}^{(1)}(\omega), \ldots, \hat{C}_{XX}^{(Q)}(\omega)$ into q overlapping $(0 < h < l)$ or non-overlapping $(h = 0)$ blocks, say $\mathcal{C}_j$, $j = 1, \ldots, q$, each containing l estimates.

Step 3. *Resampling.* Generate k bootstrap samples $y_1^*, \ldots, y_k^*$ of size l each, from $\mathcal{C}_1, \ldots, \mathcal{C}_q$.

Step 4. *Reshaping.* Concatenate $y_1^*, \ldots, y_k^*$ into a vector $\mathbf{Y}^*$ and estimate $\hat{C}_{XX}^*(\omega)$.

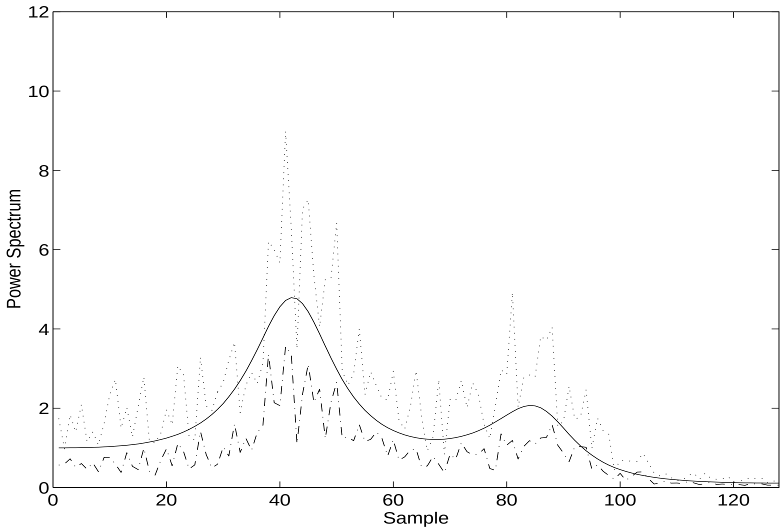
Step 5. *Confidence interval.* Repeat Steps 3-4 and proceed as before to obtain a confidence interval for $C_{XX}(\omega)$.

Simulation Results

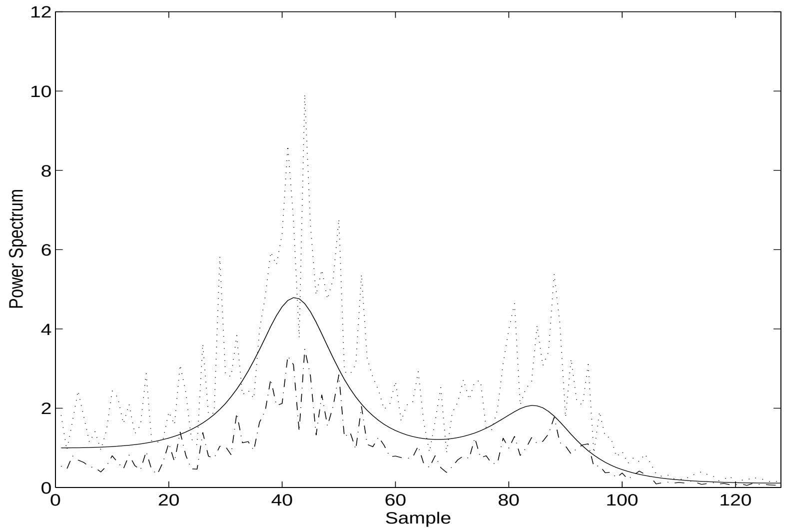
Consider the AR process driven by an i.i.d. $\mathcal{N}(0,1)$ process:

$$X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + \varepsilon_t \,.$$

With $n = 2,000$, $L = 128$, $M = 20$, $l = 6$, $h = 2$ and $N = 100$, we obtain the following 95% confidence interval, based on 100 runs.



Block of blocks bootstrap method.



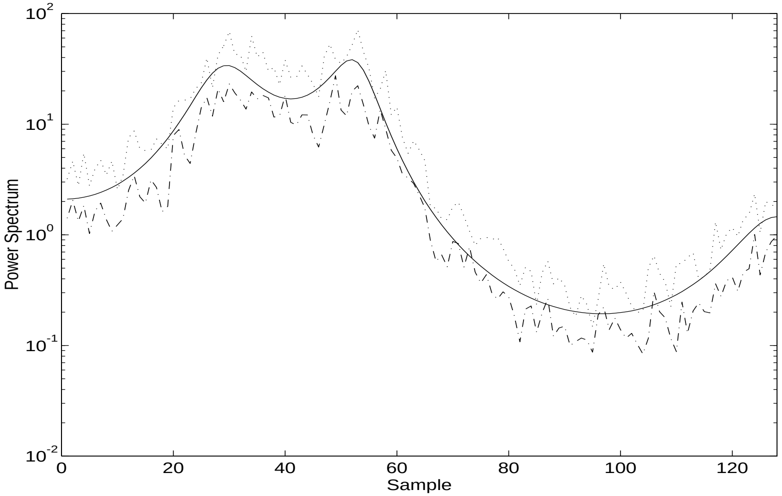
Circular block bootstrap method.

Simulation Results (Cont'd)

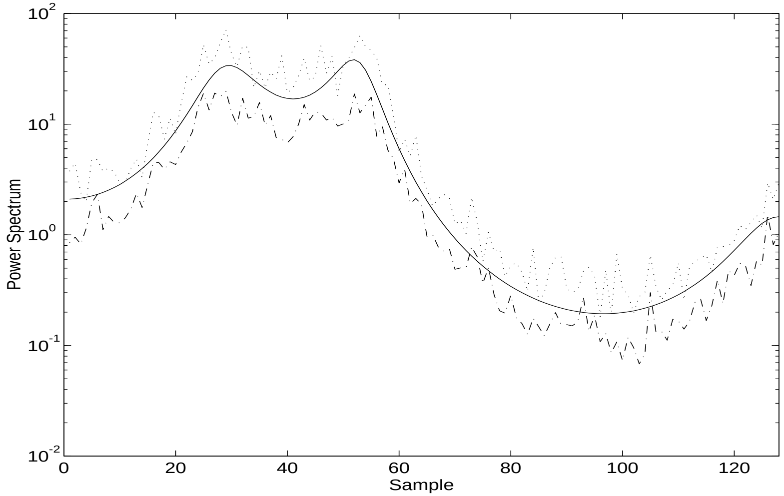
Consider the AR process driven by an i.i.d. $\mathcal{U}(-2.5, 2.5)$ process:

$$Y_t = Y_{t-1} - 0.7Y_{t-2} - 0.4Y_{t-3} + 0.6Y_{t-4} - 0.5Y_{t-5} + \zeta_t\,.$$

With $n = 2,000$, $L = 128$, $M = 20$, $l = 6$, $h = 2$ and $N = 100$, we obtain the following 95% confidence interval, based on 100 runs.



Block of blocks bootstrap method.



Circular block bootstrap method.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Testing for Departure from Gaussianity

- Tests for departure from Gaussianity [Subba Rao & Gabr (1980), Hinich (1982)] have received considerable interest among signal processing practitioners [Swami *et al.* (1995)].
- A limitation of the tests is the large amount of data required for the asymptotic distribution of the test statistics to hold.
- The bootstrap can be used to test for departure from Gaussianity with high power while maintaining the level of significance, even for small sample sizes [Zoubir & Iskander (1999)].
- The bootstrap can also be used to set confidence bands for the bicoherence [Zoubir & Iskander (1996)].

Testing for Departure from Gaussianity (Cont'd)

- Let $X_0, \ldots, X_{n-1}$ be observations from a strictly stationary real-valued time-series X_t, $t \in \mathbb{Z}$, with $\mu_X = 0, \sigma_X^2 > 0$ and $C_{XXX}(\omega_j, \omega_k), -\pi \leq (\omega_j, \omega_k) \leq \pi$.
- For a Gaussian process (as well as any other symmetric process) $C_{XXX}(\omega_j, \omega_k) \equiv 0$. Testing the bispectrum for zero may be seen as testing for departure from Gaussianity.
- Rejection of the hypothesis implies that the process is non-Gaussian; otherwise, it may be non-Gaussian.
- We would test $C_{XXX}(\omega_j, \omega_k)$ for zero when ω_j and ω_k are restricted to $0 \leq \omega_k \leq \omega_j$, $\omega_k + 2\omega_j \leq 2\pi$ due to symmetry.

Testing for Departure from Gaussianity (Cont'd)

Divide the observations X_t, $t = 0, \ldots, T-1$ into P non-overlapping segments of n consecutive measurements and calculate the periodogram $I^{(i)}_{XXX}(\omega_j, \omega_k)$ for each segment $i = 1, \ldots, P$,

$$I^{(i)}_{XXX}(\omega_j, \omega_k) = \frac{1}{n} d^{(i)}_X(\omega_j) d^{(i)}_X(\omega_k) \bar{d_X}^{(i)}(\omega_j + \omega_k), \quad -\pi \leq \omega_j, \omega_k \leq \pi\,,$$

where $d^{(i)}_X(\omega_i)$ is the finite Fourier transform of the i-th segment and $\bar{d_X}$ is its complex conjugate.

An estimate of $C_{XXX}(\omega_j, \omega_k)$ is obtained through

$$\hat{C}_{XXX}(\omega_j, \omega_k) = \frac{1}{P} \sum_{i=1}^{P} I^{(i)}_{XXX}(\omega_j, \omega_k).$$

Testing for Departure from Gaussianity (Cont'd)

Our procedure is based on the approximate regression

$$I^{(i)}_{XXX}(\omega_j, \omega_k) = C_{XXX}(\omega_j, \omega_k) + \varepsilon_{j,k} V(\omega_j, \omega_k), \qquad (j,k) \in \mathcal{D},$$

$$\begin{aligned} V(\omega_j, \omega_k)^2 &= nC_{XX}(\omega_j) C_{XX}(\omega_k) C_{XX}(\omega_j + \omega_k) \\ &\times \left[1 + \delta(j-k) + \delta(n-2j-k) + 4\delta(n-3j)\delta(n-3k)\right]. \end{aligned}$$

Herein, $C_{XX}(\omega)$ is the spectrum of X_t, $\delta(k)$ is Kronecker's delta function, $\omega_j = 2\pi j/n$ and $\omega_k = 2\pi k/n$ are discrete frequencies and $\mathcal{D} = \{0 < k \le j,\ 2j + k \le n\}$.

We shall assume that $\varepsilon_{j,k}$ are independent and identically distributed random variates which holds for a reasonably large n.

Testing for Departure from Gaussianity (Cont'd)

Step 1. Calculate $I_{XX}^{(i)}(\omega_j)$, $I_{XXX}^{(i)}(\omega_j, \omega_k)$, $\hat{C}_{XX}(\omega_j)$, $\hat{C}_{XXX}(\omega_j, \omega_k)$, $\hat{\sigma}(\omega_j, \omega_k)$ (using the bootstrap) and

$$\tilde{C} = \sum_{j,k \in \mathcal{D}} \frac{|\hat{C}_{XXX}(\omega_j, \omega_k)|}{\hat{\sigma}(\omega_j, \omega_k)} \,.$$

Step 2. For each segment, estimate the residuals

$$\hat{\varepsilon}_{j,k}^{(i)} = \frac{I_{XXX}^{(i)}(\omega_j, \omega_k) - \hat{C}_{XXX}(\omega_j, \omega_k)}{\hat{V}(\omega_j, \omega_k)}, \qquad j, k \in \mathcal{D} \,.$$

Step 3. Centre the residuals to obtain $\tilde{\varepsilon}_{j,k}^{(i)} = \hat{\varepsilon}_{j,k}^{(i)} - \bar{\varepsilon}^{(i)}$, $i = 1, \ldots, P$, where $\bar{\varepsilon}^{(i)}$ is an average over all $\hat{\varepsilon}_{j,k}^{(i)}$.

Testing for Departure from Gaussianity (Cont'd)

Step 4. Draw independent bootstrap residuals $\tilde{\varepsilon}_{j,k}^{(i)*}$.
Step 5. Compute the bootstrap biperiodogram ordinates

$$I_{XXX}^{(i)*}(\omega_j, \omega_k) = \hat{C}_{XXX}(\omega_j, \omega_k) + \tilde{\varepsilon}_{j,k}^{(i)*}\hat{V}(\omega_j, \omega_k).$$

Step 6. Obtain the bootstrap bispectral estimate

$$\hat{C}_{XXX}^{*}(\omega_j, \omega_k) = \frac{1}{P}\sum_{i=1}^{P} I_{XXX}^{(i)*}(\omega_j, \omega_k)\,,$$

Step 7. Compute the statistic

$$\tilde{C}^{*} = \sum_{j,k\in\mathcal{D}} \frac{|\hat{C}_{XXX}^{*}(\omega_j, \omega_k) - \hat{C}_{XXX}(\omega_j, \omega_k)|}{\hat{\sigma}^{*}(\omega_j, \omega_k)}.$$

Testing for Departure from Gaussianity (Cont'd)

Step 8. Repeat Steps 4–7 a large number of times, to obtain a total of N bootstrap statistics[a] $\tilde{C}_1^*, \ldots, \tilde{C}_N^*$.

Step 9. Rank the collection $\tilde{C}_1^*, \ldots, \tilde{C}_N^*$ into increasing order to obtain $\tilde{C}_{(1)}^* \leq \cdots \leq \tilde{C}_{(N)}^*$. Reject the hypothesis of Gaussianity at level α if $\tilde{C} > \tilde{C}_{(q)}^*$, where $q = \lfloor (N+1)(1-\alpha) \rfloor$.

---

[a] $\hat{\sigma}^*(\omega_j, \omega_k)^2$ is obtained as follows. For each $I_{XXX}^*(\omega_j, \omega_k)$, repeat Steps 3–6 (nested bootstrap) a small number of times e.g., $B = 25$), replacing $\hat{C}_{XXX}(\omega_j, \omega_k)$ and $I_{XXX}^{(i)*}(\omega_j, \omega_k)$ by $\hat{C}_{XXX}^*(\omega_j, \omega_k)$ and $I_{XXX}^{(i)**}(\omega_j, \omega_k)$, respectively. Then, compute

$$\hat{\sigma}^*(\omega_j, \omega_k)^2 = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{C}_{XXX}^{**(b)}(\omega_j, \omega_k) - \frac{1}{B} \sum_{b'=1}^{B} \hat{C}_{XXX}^{**(b')}(\omega_j, \omega_k) \right)^2 ,$$

where $\hat{C}_{XXX}^{**(b)}(\omega_j, \omega_k)$, $b = 1, \ldots, B_1$, is a bootstrap version of $\hat{C}_{XXX}^*(\omega_j, \omega_k)$.

Simulation Results

We considered three processes:

- a white process
- an auto-regressive process of order five [AR(5)], and
- a moving-average process of order two [MA(2)].

Specifically, we assumed $X_{1,t} = Y_t$,

$$X_{2,t} = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Y_t,$$

$$X_{3,t} = 0.5Y_t + 0.3Y_{t-1} + 0.5Y_{t-2}, \quad t \in \mathbb{Z},$$

where Y_t is an independent random process with distribution F_Y,

Simulation Results

F_Y	i.i.d.	AR(5)	MA(2)
$N(0,1)$	6	4	16
$U(0,1)$	2	7	6
χ_2^2	92	91	88
χ_8^2	39	34	14
Laplace	23	27	20
$K(1,1)$	62	66	44
LogN	100	100	98

Rejection rate ($\alpha = 5\%$) using Subba-Rao and Gabr's test ($T = 256$).

F_Y	i.i.d.	AR(5)	MA(2)
$N(0,1)$	5	2	1
$U(0,1)$	5	2	1
χ_2^2	95	72	63
χ_8^2	69	22	12
Laplace	16	2	1
$K(1,1)$	78	42	20
LogN	100	91	77

Rejection rate ($\alpha = 5\%$) using the bootstrap test ($T = 256$, $n = 22$).

Simulation Results

F_Y	i.i.d.	AR(5)	MA(2)
$N(0,1)$	8	8	7
$U(0,1)$	7	4	10
χ_2^2	100	100	98
χ_8^2	61	48	32
Laplace	39	38	45
$K(1,1)$	85	82	68
$LogN$	100	100	96

Rejection rate ($\alpha = 5\%$) using Subba-Rao and Gabr's test ($T = 512$).

F_Y	i.i.d.	AR(5)	MA(2)
$N(0,1)$	3	2	3
$U(0,1)$	6	1	2
χ_2^2	100	94	89
χ_8^2	89	60	38
Laplace	14	8	3
$K(1,1)$	98	84	57
$LogN$	100	100	93

Rejection rate ($\alpha = 5\%$) using the bootstrap test ($T = 512$, $n = 22$).

Interpretation of the Results

- Subba Rao and Gabr's test is unable to maintain the 5% level of significance for symmetric processes, e.g. for coloured Gaussian or independent/coloured Laplace process.
- The test based on the bootstrap maintains the nominal level of significance at below 5%, except in the case of independent uniform and Laplace processes for $T = 512$.
- A comparison of power is appropriate only if the tests maintain the same nominal level of significance. We found that the bootstrap test achieves power comparable to or better than Subba Rao and Gabr's test.

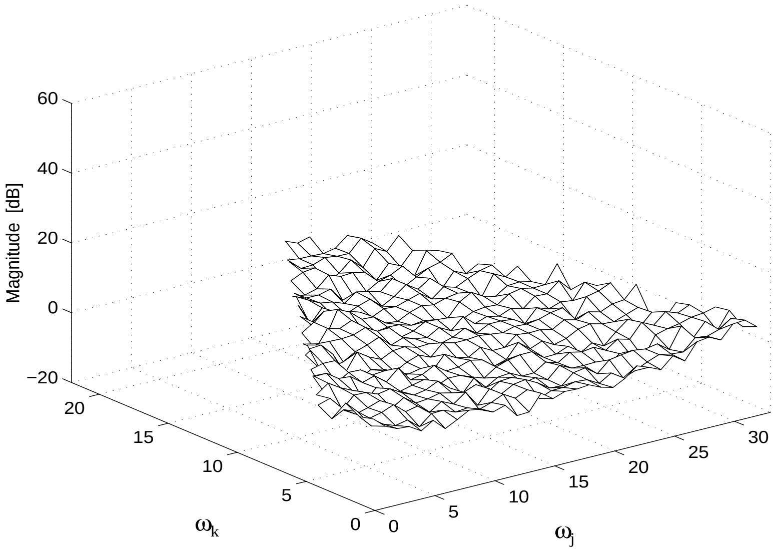
Confidence Bands estimation for the Bicoherence

- We can use the above method to set confidence bands for higher order spectra or cumulants.
- Repeating steps 4–6 a large number of times, one can obtain a total of N bootstrap estimates of the bicoherence, as

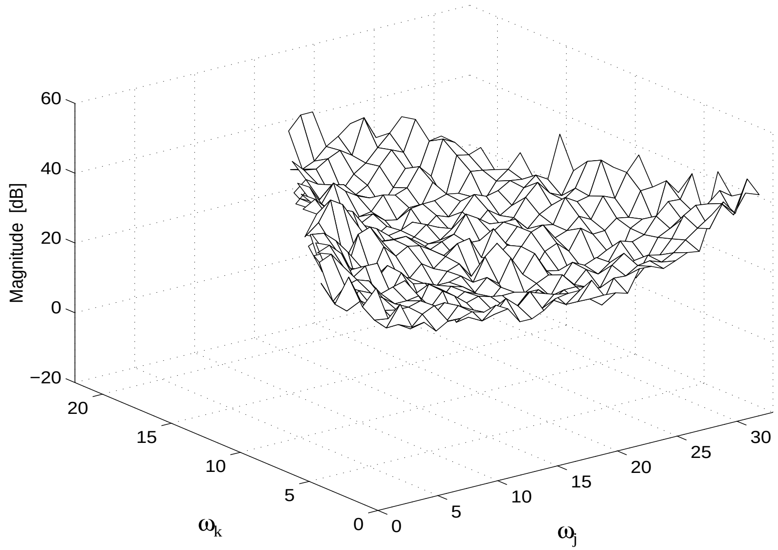
$$|\hat{C}^{*(1)}_{XXX}(\omega_j,\omega_k)|/\hat{\sigma}(\omega_j,\omega_k),\ldots,|\hat{C}^{*(B)}_{XXX}(\omega_j,\omega_k)|/\hat{\sigma}(\omega_j,\omega_k).$$

- After sorting these estimates at each frequency pair (ω_j,ω_k) one is able to determine the confidence bands using estimated percentiles obtained as in the above procedure [Zoubir & Iskander (1996,1999)].

Confidence Bands estimation for the Bicoherence



Lower 95% confidence band for the bicoherence of a white Gaussian process ($T = 4096$ and $n = 64$).



Upper 95% confidence band for the bicoherence of a white Gaussian process ($T = 4096$ and $n = 64$).

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Noise Floor Estimation in Radar

- The noise floor is one of the main radar performance statistics.
- It is computed by taking a trimmed mean of power estimates in the Doppler spectrum, after excluding the low Doppler power associated with any stationary targets or ground clutter.
- The radar operator requires an optimal value for the trim and also supplementary information describing the accuracy of the computed noise floor estimate.
- The limited number of samples available and the non-Gaussian nature of the noise field make it particularly difficult to estimate confidence measures reliably. This suggests the use of the bootstrap.

Optimal Selection of the Trim

- The goal is to use the bootstrap to optimally select the amount of trimming, α, say.
- The problem can be stated as one of estimation of θ from a class of estimates $\{\hat{\theta}(\alpha) : \ \alpha \in \mathcal{A}\}$ indexed by a parameter α, which is selected according to some optimality criterion.
- It is reasonable to use the estimate that leads to a minimum variance. Because the variance of $\hat{\theta}(\alpha)$ may depend on unknown parameters, the optimal parameter α_o is unknown.

Optimal Selection of the Trim (cont'd)

- The objective is to find α by estimating the unknown variance of $\hat{\theta}(\alpha)$ and select α which minimises the variance estimate.
- The problem of trimming has been solved theoretically [Jaeckel (1971)] for the mean of symmetric distributions. The trim is, however, restricted to the interval $[0, 25]\%$.
- In addition to the fact that for a small sample size asymptotic results are invalid, explicit expressions for the asymptotic variance may not be available outside the interval $[0, 25]\%$.

An Example: Selection of the Trim for the Mean

- Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a random sample drawn from a distribution function F.
- Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics.
- For an integer α less then $n/2$, the α-trimmed mean based on the sample $\mathcal{X}$ is given by

$$\hat{\mu}(\alpha) = \frac{1}{n - 2\alpha} \sum_{i=\alpha+1}^{n-\alpha} X_{(i)}$$

- For an asymmetric distribution we use the $\alpha\beta$-trimmed mean:

$$\hat{\mu}(\alpha, \beta) = \frac{1}{n - \alpha - \beta} \sum_{i=\alpha+1}^{n-\beta} X_{(i)}\,, \quad \alpha, \beta < n/2\,.$$

Example

Step 1. *Initial conditions.* Select the initial trim $\alpha = 0$ and $\beta = 0$.

Step 2. *Resampling.* Using a pseudo-random number generator, draw a large number, say N, of independent samples

$$\mathcal{X}_1^* = \{X_{11}^*, \ldots, X_{1n}^*\}, \ldots, \mathcal{X}_N^* = \{X_{N1}^*, \ldots, X_{Nn}^*\}$$

of size n from $\mathcal{X}$. Each sample is taken with replacement.

Step 3. *Calculation of the bootstrap statistic.* For each bootstrap sample $\mathcal{X}_j^*$, $j = 1, \ldots, N$, calculate the trimmed mean

$$\hat{\mu}_j^*(\alpha, \beta) = \frac{1}{n - \alpha - \beta} \sum_{i=\alpha+1}^{n-\beta} X_{j(i)}^*, \qquad j = 1, \ldots, N.$$

Example (cont'd)

Step 4. *Variance estimation.* Using bootstrap based trimmed mean values, calculate the estimate of the variance

$$\hat{\sigma}^{*2} = \frac{1}{N-1} \sum_{j=1}^{N} \left(\hat{\mu}_j^*(\alpha, \beta) - \frac{1}{N} \sum_{j=1}^{N} \hat{\mu}_j^*(\alpha, \beta) \right)^2 .$$

Step 5. *Repetition.* Repeat steps 2–4 using different combinations of the trim α and β.

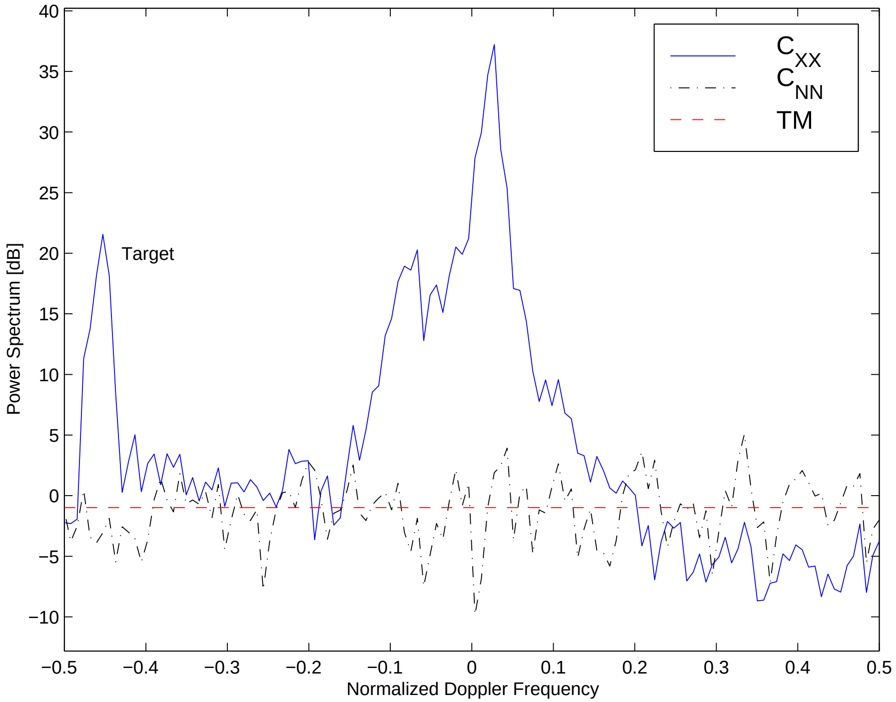
Step 6. *Optimal trim.* Choose the setting of the trim that results in a minimal variance estimate.

Noise Floor Estimation in Radar

- In radar, we estimate residuals for the Doppler spectrum by taking the ratio of the periodogram and a spectral estimate of the raw data. The residuals are used for resampling to generate bootstrap spectral estimates as in a previous example.
- We proceed as in the example for the trimmed mean, replacing X_i by $\hat{C}_{XX}(\omega_i)$, where $\omega_i = 2\pi i/n$, $i = 1, \ldots, n$, are discrete frequencies and n is the number of observations.
- The procedure makes the assumption that the spectral estimates are i.i.d. for distinct discrete frequencies.
- The optimal noise floor is found by minimising the bootstrap variance estimate of the *trimmed spectrum* w.r.t. (α, β).

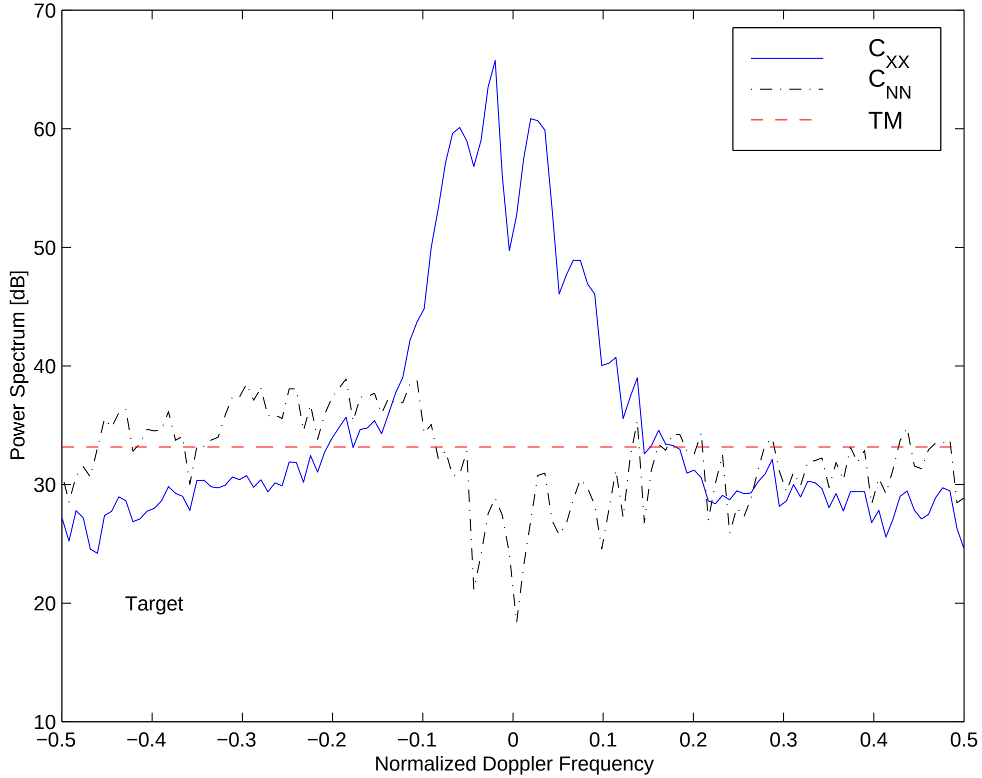
Noise Floor Estimation in Radar (Cont'd)

The radar return is assumed to consist of 128 observations from a complex-valued AR(4) process, driven by a Gaussian process.



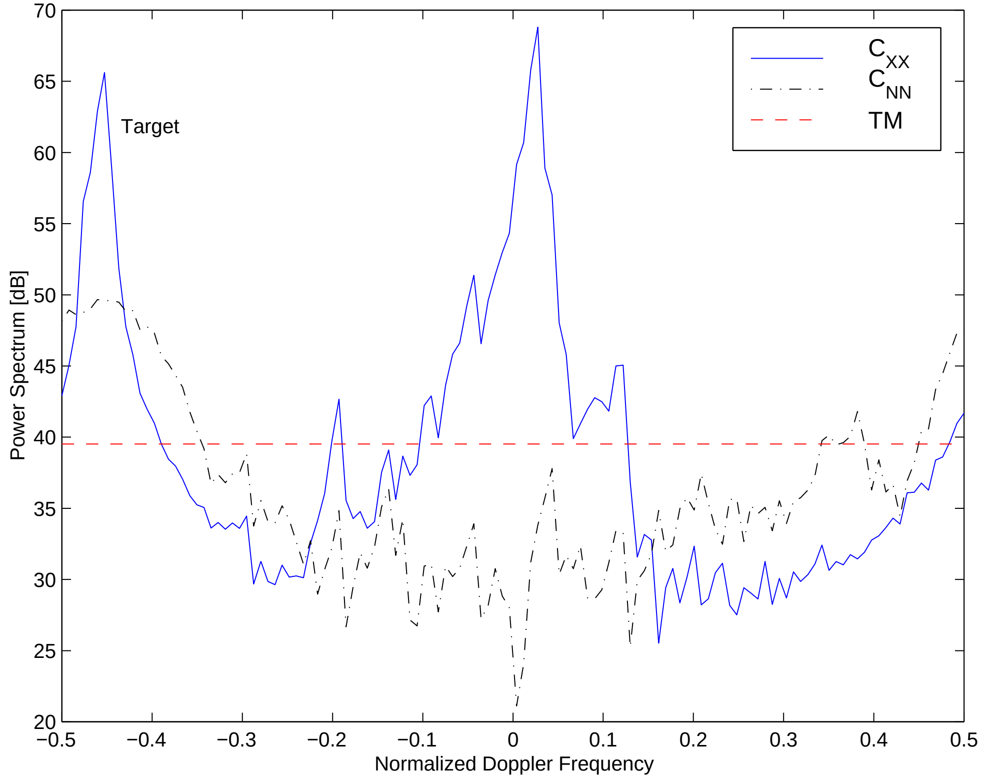
Doppler spectrum and estimated noise floor. With $N = 99$, $\hat{\theta}(\hat{\alpha}_0, \hat{\beta}_0) = 0.80$.

Noise Floor in Over-The-Horizon Radar



Doppler spectrum and estimated noise floor for a real over-the-horizon radar return (no target present). With $N = 99$, $\hat{\theta}(\hat{\alpha}_0, \hat{\beta}_0) = 2.0738 \times 10^3$.

Noise Floor in Over-The Horizon Radar

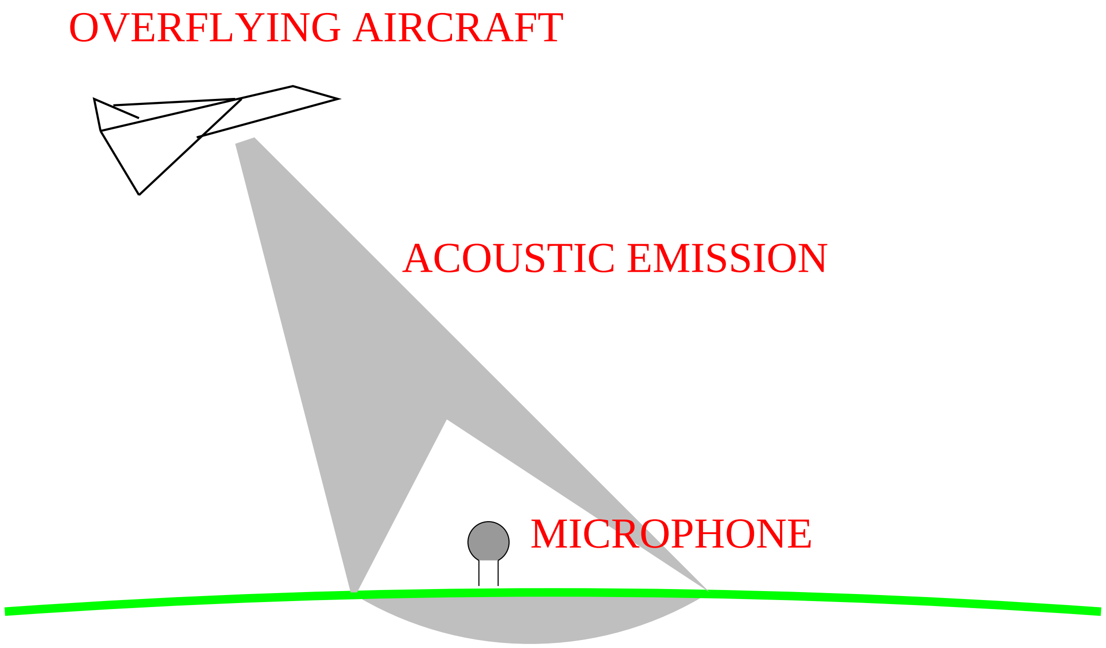


Doppler spectrum and estimated noise floor for a real over-the-horizon radar return (target present). With $N = 99$, $\hat{\theta}(\hat{\alpha}_0, \hat{\beta}_0) = 8.9513 \times 10^3$.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - **Confidence Intervals for Flight Parameters in Passive Sonar**
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Confidence Intervals for an Aircraft's Flight Parameters



Schematic of the passive acoustic scenario.

The Passive Acoustic Approach

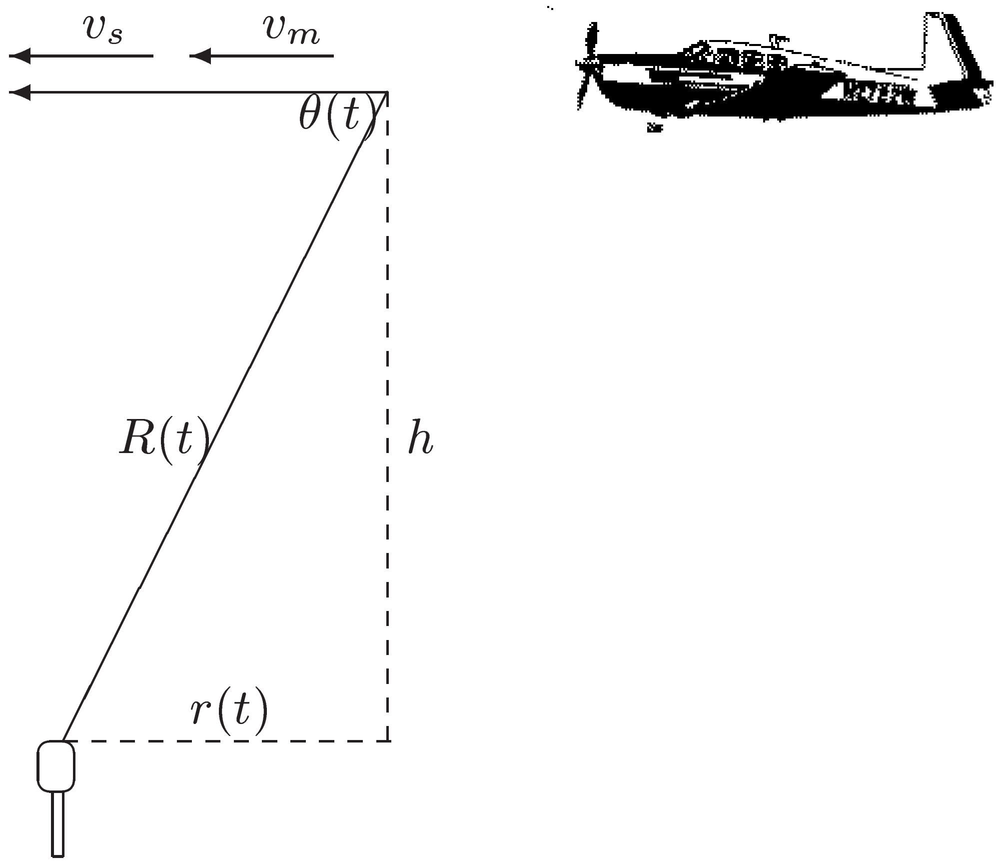
- Information about the flight parameters is contained in the acoustic signal as heard by the stationary observer.
- The frequency of the observed acoustic signal $s(t)$ from the over-flying aircraft undergoes a time varying Doppler shift.
- Thus, the phase of $s(t)$ undergoes a time varying rate of change.

A simple model for the aircraft acoustic signal, as heard by a stationary observer, is given by

$$X(t) = z(t) + U(t) = Ae^{j\phi(t)} + U(t) = Ae^{j2\pi \int_{-\infty}^{t} f(\tau)d\tau} + U(t)\,,$$

where $z(t) = s(t) + j\mathsf{H}[s(t)]$ and $\mathsf{H}[\cdot]$ is the Hilbert transform.

Problem Description



Schematic of the geometric arrangement of the aircraft and observer in terms of the physical parameters of the model.

Problem Description (Cont'd)

Context: Estimation of an aircraft's flight parameters (height, speed, range and acoustic frequency) from the acoustic signal as heard by a stationary observer [Reid *et al.* (1997), Ferguson (1992), Ferguson & Quinn (1994)].

Problem: To estimate an aircraft's flight parameters using passive acoustic techniques and to determine confidence bounds given only a single acoustic realisation.

Solution: Estimate the aircraft flight parameters from the time varying phase of the observed acoustic signal. Use bootstrap techniques to estimate the confidence bounds [Reid *et al.* (1996), Zoubir & Boashash (1998)].

The Passive Acoustic Model

An observer phase model $\phi(t)$ describes the phase of the observed acoustic signal in terms of the flight parameters and is given by

$$\phi(t) = 2\pi \frac{f_a c^2}{c^2 - v^2} \left(t - \sqrt{\frac{h^2 c + v^2 t^2 c + 2v^2 t h}{c^3}} \right) + \phi_0, \quad -\infty < t < \infty \,,$$

where t_0 is the time when the aircraft is directly overhead, f_a is the source acoustic frequency, c is the speed of sound in the medium, v is the constant velocity of the aircraft, h is the constant altitude of the aircraft and ϕ_0 is an initial phase constant.

The Passive Acoustic Model (Cont'd)

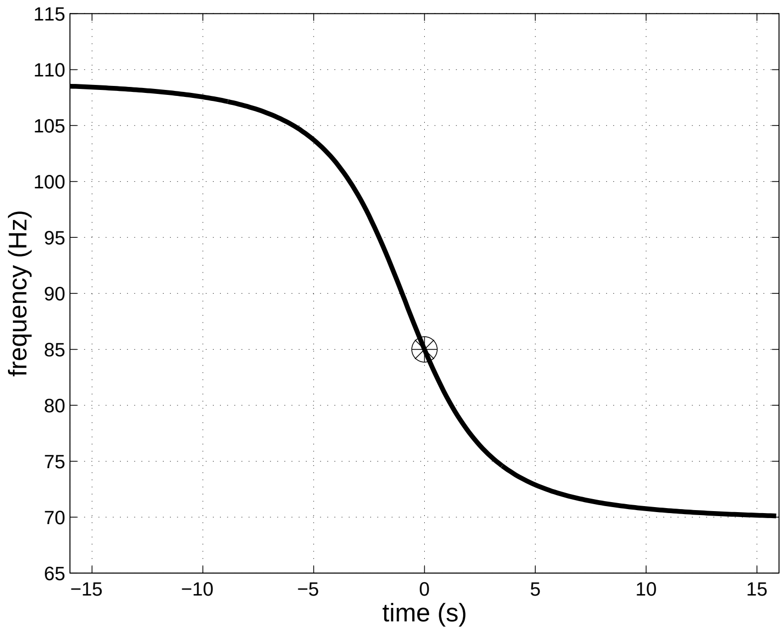
From the phase model, the instantaneous frequency (IF), relative to the stationary observer, can be expressed as

$$f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} = \frac{f_a c^2}{c^2 - v^2}\left(1 - \frac{v^2(t + h/c)}{\sqrt{h^2(c^2 - v^2) + v^2 c^2 (t + h/c)^2}}\right).$$

For a given $f(t)$, or $\phi(t)$, $-\infty < t < \infty$ and c, the aircraft parameters collected in the vector $\boldsymbol{\theta} = (f_a, h, v, t_0)'$ can be uniquely determined from the phase or observer IF model.

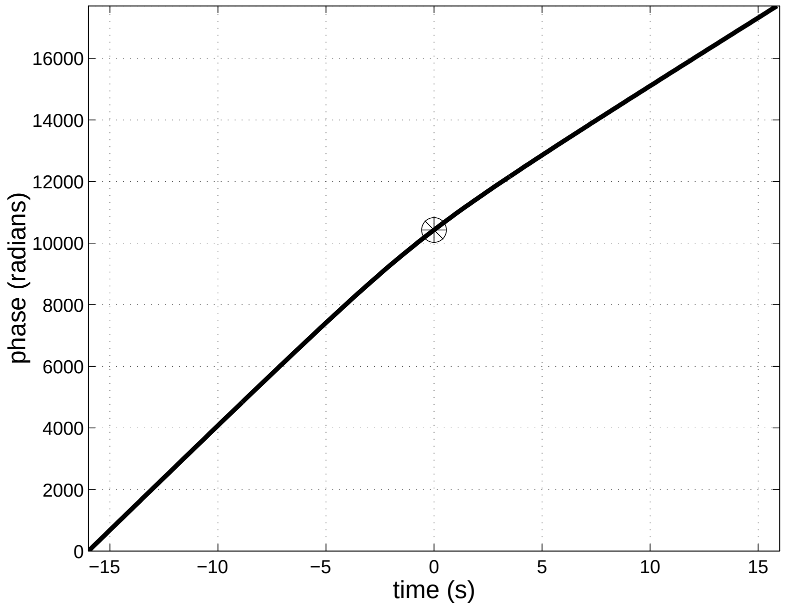
Consider appropriate sampled versions of the continuous-time signals and denote by ϕ_t and X_t, the phase, and the observed signal as functions of $t = 0, \pm 1, \pm 2, \ldots$.

The Instantaneous Frequency



Typical Time varying acoustic frequency profile of an over-flying aircraft as described by the observer frequency model. For this example, $v_a = 75$ m/s, $h = 300$ m, $f_a = 85$ Hz and $t_0 = 0$ s. The cartwheel symbol indicates the time at which the aircraft is directly overhead the observer.

The Instantaneous Phase



Typical Time varying acoustic phase profile of an over-flying aircraft as described by the observer phase model. For this example, $v_a = 75$ m/s, $h = 300$ m, $f_a = 85$ Hz and $t_0 = 0$ s. The cartwheel symbol marks the time at which the aircraft is directly overhead the observer.

A Bootstrap Approach

Step 0. Collect and sample the data to obtain X_t, $t = -n/2, \ldots, n/2-1$.

Step 1. Unwrap the phase of the signal X_t to provide a non-decreasing function $\hat{\phi}_t$ which approximates the true phase ϕ_t.

Step 2. Obtain $\hat{\boldsymbol{\theta}}$, an initial estimate of the aircraft parameters by fitting the non-linear observer phase model $\phi_{t;\boldsymbol{\theta}}$ to $\hat{\phi}_t$ in a least-squares sense.

Step 3. Compute the residuals

$$\hat{\varepsilon}_t = \phi_{t;\hat{\boldsymbol{\theta}}} - \hat{\phi}_t\,, \qquad t = -n/2, \cdots, n/2-1\,.$$

A Bootstrap Approach (Cont'd)

Step 4. Compute $\hat{\sigma}_i$, a bootstrap estimate of the standard deviation of $\hat{\theta}_i$, $i = 1, \ldots, 4$.

Step 5. Draw a random sample $\mathcal{X}^* = \{\hat{\varepsilon}^*_{-n/2}, \cdots, \hat{\varepsilon}^*_{n/2-1}\}$, with replacement, from $\mathcal{X} = \{\hat{\varepsilon}_{-n/2}, \cdots, \hat{\varepsilon}_{n/2-1}\}$ and construct

$$\hat{\phi}^*_t = \phi_{t;\hat{\boldsymbol{\theta}}} + \hat{\varepsilon}^*_t \,.$$

Step 6. Obtain and record the bootstrap estimates of the aircraft parameters $\hat{\boldsymbol{\theta}}^*$ by fitting the observer phase model to $\hat{\phi}^*_t$ in a least-squares sense.

A Bootstrap Approach (Cont'd)

Step 7. Estimate the standard deviation of $\hat{\theta}_i^*$ using a nested bootstrap step and compute and record

$$\hat{T}_i^* = \frac{\hat{\theta}_i^* - \hat{\theta}_i}{\hat{\sigma}_i^*}, \qquad i = 1, \ldots, 4 .$$

Step 8. Repeat Steps 5 through 7 a large number of times N.

Step 9. Order the bootstrap estimates as $\hat{T}^*_{i,(1)} \leq \hat{T}^*_{i,(2)} \leq \cdots \leq \hat{T}^*_{i,(N)}$ and compute the $(1-\alpha)100\%$ confidence interval

$$(\hat{\theta}_i - \hat{T}^*_{i,(q_1)}\hat{\sigma}_i, \;\; \hat{\theta}_i - \hat{T}^*_{i,(q_2)}\hat{\sigma}_i),$$

where $q_1 = N - \lfloor N\alpha/2 \rfloor + 1$ and $q_2 = \lfloor N\alpha/2 \rfloor$.

Simulation Results

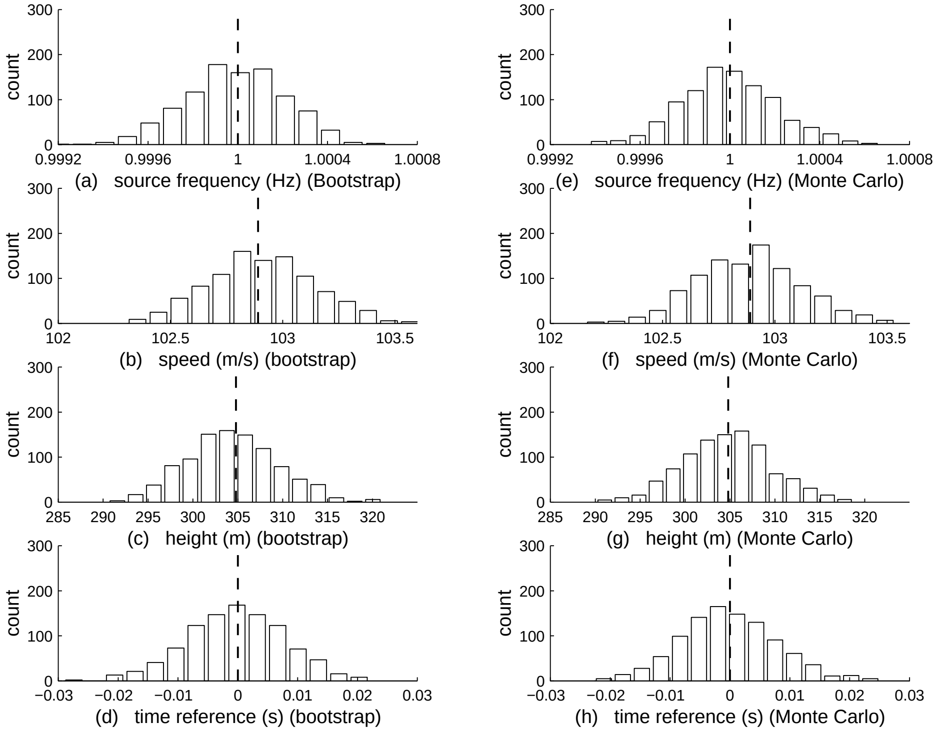
- An $n = 320$ point passive acoustic signal z_t is generated at three levels of SNR (15 dB, 20 dB and 30 dB).
- We set $h = 304.8$ m, $v = 102.89$ m/s, $f_a = 1$ Hz, $t_0 = 0$ s, sampling frequency $f_s = 8$ Hz and bootstrap variables $B_1 = 25$, $B_2 = 100$ and $B_3 = 1000$.
- The bootstrap confidence bounds are computed and compared with those obtained by Monte Carlo simulation where the aircraft parameters are calculated for 1000 independent realisations of z_t.

Simulation Results (Cont'd)

SNR dB		h [m]		v [m/s]		t_0 [s]		f_a [Hz]	
	Actual	304.8		102.89		0.000		1.000	
		BS	MC	BS	MC	BS	MC	BS	MC
30	Upper Bound	308.40	308.0	102.94	103.04	0.004	0.002	1.000	1.000
	Lower Bound	301.4	301.9	102.67	102.74	-0.005	-0.006	1.000	1.000
	Interval length	6.6	6.1	0.27	0.30	0.009	0.008	0.000	0.000
20	Upper Bound	309.5	314.5	103.17	103.34	0.011	0.016	1.000	1.000
	Lower Bound	290.8	295.2	102.26	102.44	-0.020	-0.016	0.999	0.999
	Interval length	18.7	19.3	0.91	0.90	0.031	0.032	0.001	0.001
15	Upper Bound	315.7	320.7	103.54	103.65	0.040	0.027	1.001	1.001
	Lower Bound	286.0	288.6	102.14	102.12	-0.008	-0.027	0.999	0.999
	Interval length	29.7	32.1	1.40	1.53	0.048	0.054	0.002	0.002

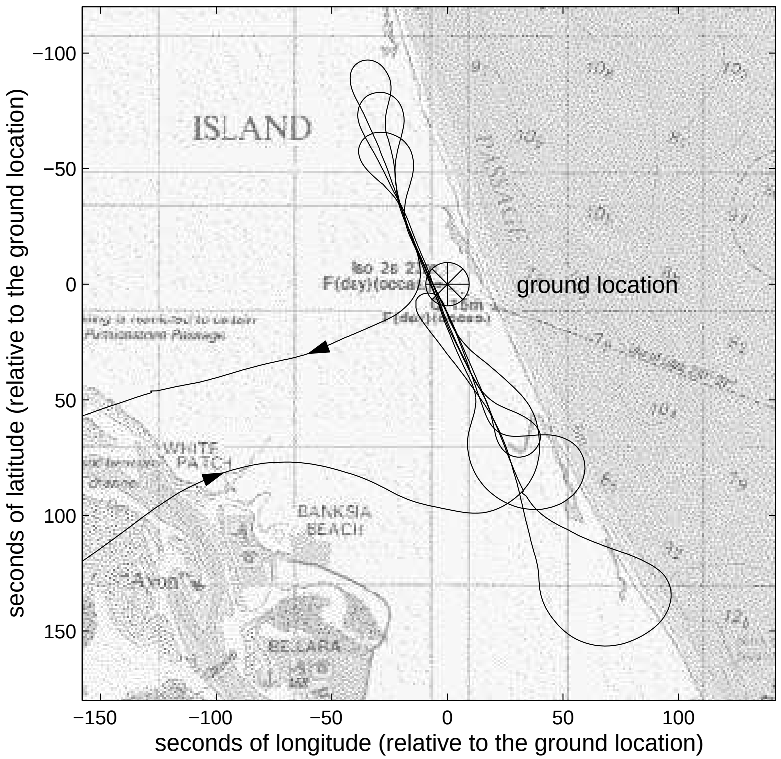
The bootstrap derived 95% confidence bounds for each of the parameters are compared with the 95% confidence bounds determined by Monte Carlo simulation.

Simulation Results (Cont'd)



The histograms of the bootstrap distribution (left column) are compared with the histograms obtained by Monte Carlo simulation (right column) for each of the parameters at 20 dB SNR. The true values are indicated by the dashed line.

Real Experiment



The complete flight path of an aircraft acoustic data experiment conducted at Bribie Island showing five fly-overs. The ground location is indicated by the cartwheel symbol.

Real Experiment (Cont'd)

		h [m]	v [m/s]	t_0 [s]	f_a [Hz]
	Nominal Value	149.05	36.61	0.00	76.91
Run 1.	Upper Bound	142.03	36.50	0.02	76.99
	Lower Bound	138.16	36.07	-0.03	76.92
	Interval length	3.87	0.43	0.05	0.07
	Nominal Value	152.31	52.94	0.00	77.90
Run 2.	Upper Bound	161.22	53.27	-0.04	78.18
	Lower Bound	156.45	52.64	-0.06	78.13
	Interval length	4.77	0.62	0.02	0.05
	Nominal Value	166.52	47.75	0.00	75.94
Run 3.	Upper Bound	232.30	57.29	0.10	76.48
	Lower Bound	193.18	53.47	-0.14	75.87
	Interval length	39.13	3.83	0.24	0.60
	Nominal Value	233.01	56.56	0.00	77.65
Run 6.	Upper Bound	243.02	57.15	0.74	77.96
	Lower Bound	209.72	53.69	0.68	77.80
	Interval length	33.30	3.46	0.06	0.16

The bootstrap derived 95% confidence bounds for four real test signals using the unwrapped phase based parameter estimator.

Interpretation of the Results

- The proposed bootstrap and unwrapped phase based estimation techniques can be used to provide a practical flight parameter estimation scheme.
- The techniques do not assume any statistical distribution for the parameter estimates and can be applied in the absence of multiple acoustic realisations.
- The obtained confidence bounds are in close agreement with those obtained from Monte Carlo simulations.
- Similar experiments with Central Finite Difference (CFD) estimates were performed. The confidence bounds presented here are much tighter than those of the CFD based estimates.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Modelling Polynomial Phase Signals

- Many non-stationary signals encountered in radar, sonar, telecommunications, seismology, or biomedical engineering can be expressed in the general form of a complex analytic signal

$$z(t) = a_0 \exp\{j\varphi(t)\}\,,$$

where a_0 and $\varphi(t)$, $t \in [T_1, T_2]$, $T_1, T_2 < \infty$, are the amplitude and the phase of the signal, respectively.

- In practice, $z(t)$ is observed in stationary complex noise $U(t)$ and sampled, yielding values x_t, $t = 0, \ldots, n-1$, from the model

$$X_t = a_0 \exp\{j\varphi_t\} + U_t, \qquad t = 0, \ldots, n-1.$$

Modelling Polynomial Phase Signals (Cont'd)

- Provided that certain regularity conditions are fulfilled [Kreyszig (1989)], $\varphi(t)$ can be modelled by

$$X_t = a_0 \exp\left\{ j \sum_{q=0}^{Q} b_q \psi_{t,q} \right\} + U_t, \qquad t = 0, \ldots, n-1,$$

  where b_q, $q = 0, \ldots, Q$, are unknown real valued parameters, and $\{\psi_{t,q}\}$ is an arbitrary set of basis sequences.

- The signal $z_t = a_0 \exp\left\{ j \sum_{q=0}^{Q} b_q \psi_{t,q} \right\}$, $t \in \mathbb{Z}$, is referred to as a frequency modulated (FM) signal.

Modelling Polynomial Phase Signals (Cont'd)

- In the special case where $\psi_{t,q} = t^q$, z_t is called a polynomial phase signal.
- <u>**Problem:**</u> Given a short segment of length n of noisy observations of the polynomial phase signal, model z_t. Modelling involves *selection* of the model and *conditional estimation* of the parameters [Akaike (1978)].
- model selection consists of choosing an appropriate set of sequences t^q, $q \in \beta$, $\beta \subseteq \{0, \ldots, Q\}$, where Q is an arbitrarily large model order.
- Conditional estimation of the model parameters refers to estimation conditioned on the unknown model.

Bootstrap Model Selection based on LSE

- Assuming that $\text{SNR} = a_0^2/\sigma_U^2$ is large, one can write for $t = 0, \ldots, n-1$ [Tretter (1985), Djuric & Kay (1990)]

$$X_t = a_0 \exp\left\{ j \sum_{q=0}^{Q} b_q t^q \right\} + U_t \approx a_0 \exp\left\{ j \left(\sum_{q=0}^{Q} b_q t^q + W_t \right) \right\},$$

with W_t, real, zero-mean, i.i.d. and $\sigma_W^2 = \sigma_U^2/(2a_0^2)$.

- The estimation problem is reduced to

$$\boldsymbol{\Phi} = \boldsymbol{H}\, \boldsymbol{b} + \boldsymbol{W},$$

where $\boldsymbol{\Phi} = (\phi_0, \ldots, \phi_{n-1})'$, $\boldsymbol{W} = (W_0, \ldots, W_{n-1})'$, $\boldsymbol{b} = (b_0, \ldots, b_Q)'$, $\boldsymbol{H} = (\boldsymbol{h}_0', \ldots, \boldsymbol{h}_{n-1}')'$, $\boldsymbol{h}_t' = (1, t, \ldots, t^Q)'$, $t = 0, \ldots, n-1$.

Residuals Based Bootstrap Model Selection

With β a subset of $\{0, \ldots, Q\}$, the *optimal model* is the model β_o such that $\boldsymbol{b}_{\beta_o}$ contains all non-zero components of $\boldsymbol{b}$, where $\boldsymbol{b}_\beta$ is a vector containing the components of $\boldsymbol{b}$ indexed by the integers in β.

Step 1. Select the largest possible model $\beta = \{0, \ldots, Q\}$, and find the least-squares estimate $\hat{\boldsymbol{b}}$ of $\boldsymbol{b} = (b_0, \ldots, b_Q)'$.

Step 2. Compute the residuals

$$\hat{w}_t = \phi_t - \boldsymbol{h}_t' \hat{\boldsymbol{b}}, \qquad t = 0, \ldots, n-1.$$

Step 3. Centre and scale the residuals, to obtain

$$\tilde{w}_t = \left(\hat{w}_t - \frac{1}{n} \sum_{t=0}^{n-1} \hat{w}_t \right) \Big/ \sqrt{1 - \frac{Q+1}{n}}, \qquad t = 0, \ldots, n-1.$$

Bootstrap Model Selection (Cont'd)

Step 4. For all models $\beta \subseteq \{0, \ldots, Q\}$,

(a) Draw $\hat{w}_t^*$, with replacement, from $\tilde{w}_t$, $t = 0, \ldots, n-1$.

(b) Compute

$$\phi_t^* = \boldsymbol{h}'_{\beta t}\hat{\boldsymbol{b}}_\beta + \hat{w}_t^*, \qquad t = 0, \ldots, l-1,$$

where l is such that $l/n \to 0$ and $l \to \infty$.

(c) Find the least-squares estimate $\hat{\boldsymbol{b}}^*_{\beta,l}$ from (b).

(d) Compute

$$\Gamma^*_{n,l}(\beta) = \frac{1}{n}\sum_{t=0}^{n-1}\left(\phi_t - \boldsymbol{h}'_{\beta t}\hat{\boldsymbol{b}}^*_{\beta,l}\right)^2 .$$

Bootstrap Model Selection (Cont'd)

Step 4. (*Continued*)

(e) Repeat steps (a)–(d) a number of times (e.g. 100), to obtain a total of N bootstrap statistics $\Gamma_{n,l}^{*(1)}(\beta), \ldots, \Gamma_{n,l}^{*(N)}(\beta)$ and compute

$$\hat{\Gamma}_{n,l}^{*}(\beta) = \frac{1}{B} \sum_{b=1}^{N} \hat{\Gamma}_{n,l}^{*(b)}(\beta).$$

Step 5. Choose $\hat{\beta}$ for which $\hat{\Gamma}_{n,l}^{*}(\beta)$ is minimum w.r.t. β.

The procedure above is consistent in that $\lim_{n\to\infty} \mathsf{Pr}\{\hat{\beta} = \beta_o\} = 1$, provided l is such that $l/n \to 0$ and $l \to \infty$ [Shao (1996), Zoubir & Iskander (1999)].

Simulation Results

- Consider a quadratic FM signal of the form

$$z_t = \exp\{j(0.5 + 0.05t + 0.0002t^3)\}, \qquad t = 0, \dots, n-1,$$

  embedded in i.i.d. noise U_t, $t = 0, \dots, n-1$. The true model of this quadratic FM signal is $\beta_o = \{b_0, b_1, b_3\}$.

- The noise is selected to be Gaussian, although the distribution of the noise is not relevant provided it has a finite variance.

- The signal-to-noise ratio ranges from 5 dB to 15 dB.

- 100 bootstrap resamples are used. The number of samples in each realisation is set to $n = 64$ and $l = 48$ (l should be such that Q/l is reasonably small).

Simulation Results (Cont'd)

Model β	$\hat{\Gamma}^*_{n,l}(\beta)$	AIC	MDL	HQ	AICC
$(b_0, 0, b_2, b_3, 0)$	0	0	0	0	0
$(b_0, 0, b_2, b_3, b_4)$	0	5.1	2.4	5.8	4.8
$(b_0, b_1, 0, b_3, 0)$	**100.0**	**87.5**	**96.0**	**82.7**	**88.9**
$(b_0, b_1, 0, b_3, b_4)$	0	3.4	0.9	5.1	3.0
$(b_0, b_1, b_2, b_3, 0)$	0	3.5	0.7	5.0	3.1
$(b_0, b_1, b_2, b_3, b_4)$	0	0.5	0	1.4	0.2

Empirical probability (in percent) of selecting the true model, $(b_0, b_1, 0, b_3, 0)$, of a quadratic FM signal embedded in Gaussian noise. SNR $= 15$dB, $n = 64$, $l = 48$. Models not selected by any of the methods are not shown.

Simulation Results (Cont'd)

Model β	$\hat{\Gamma}^*_{n,l}(\beta)$	AIC	MDL	HQ	AICC
$(b_0, 0, 0, b_3, 0)$	0.5	0	0	0	0
$(b_0, 0, 0, b_3, b_4)$	0	1.1	1.8	1	1.1
$(b_0, 0, b_2, 0, 0)$	0.2	0	0	0	0
$(b_0, 0, b_2, 0, b_4)$	0	1	1.8	0.7	1.2
$(b_0, 0, b_2, b_3, 0)$	0.3	13.8	14.3	13.7	13.9
$(b_0, 0, b_2, b_3, b_4)$	0	4.1	2.8	5	3.5
$(b_0, b_1, 0, 0, b_4)$	6.2	1	1.6	0.8	1.2
$(b_0, b_1, 0, b_3, 0)$	**84.9**	**56.9**	**62.8**	**53.4**	**58.9**
$(b_0, b_1, 0, b_3, b_4)$	0	1.6	0.3	2.5	1.2
$(b_0, b_1, b_2, 0, 0)$	5.3	1.1	1.4	1.1	1.1
$(b_0, b_1, b_2, 0, b_4)$	0	3.1	0.9	3.4	2.3
$(b_0, b_1, b_2, b_3, 0)$	2.6	3.1	2.2	3.5	3.1
$(b_0, b_1, b_2, b_3, b_4)$	0	13.2	10.1	14.9	12.5

Empirical probability (in percent) of selecting the true model, $(b_0, b_1, 0, b_3, 0)$, of a quadratic FM signal in Gaussian noise. SNR $= 5$dB, $n = 64$, $l = 48$.

Interpretation of Results

- Bootstrap techniques can be applied to constant amplitude polynomial phase modelling.
- Results have shown that the empirical probability of correctly selecting the model of a constant amplitude polynomial phase signal is high at a reasonable SNR.
- A comparison with other techniques demonstrates the superiority of bootstrap techniques.
- Other bootstrap techniques based on the discrete polynomial phase transform have also been devised. They consist of a two-stage approach involving hypothesis testing [Zoubir & Iskander (1999)].

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Summary

- It is often necessary to find the sampling distributions of parameter estimators, so that the respective means and variances can be calculated, and more generally, confidence intervals for the true parameters can be set.
- Most techniques for computing variances or confidence intervals 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 because, for example, the process is non-stationary and only small portions of stationary data are considered.

Summary (Cont'd)

- Bootstrap techniques are an alternative to asymptotic methods.
- The bootstrap does with a computer what the experimenter would do in practice, if it were possible: they would repeat the experiment.
- With the bootstrap, the observations are randomly re-assigned, and the estimates re-computed. This process is done thousands of times to simulate repeated experiments.
- In an era of exponentially increasing computational power, such computer-intensive methods are becoming increasingly attractive.

Summary (Cont'd)

- The tutorial provides the fundamental concepts and methods needed by the signal processing practitioner to decide when and how to apply the bootstrap successfully.
- The tutorial focuses on the independent data bootstrap. The assumption of i.i.d. data can break down in practice either because the data is not independent or because it is not identically distributed, or both.
- The bootstrap can still be invoked if we **knew the model** that generated the data. In other cases we can make the reasonable assumption that the data is identically distributed but not independent such as in autoregressive processes.

Summary (Cont'd)

- For confidence interval estimation or hypothesis testing, it is essential that the statistic used is asymptotically pivotal.
- It has been shown that working on a variance stable scale is better than studentising.
- When a variance stabilising transformation is not known, bootstrap can be used to estimate it.
- Many applications have been presented to demonstrate the power of the bootstrap. *Special care* is however required when applying the bootstrap in real-life situations.

Outline

- Model Selection
  - Non-Linear Models
  - Order Selection in Autoregressive Models
- More Applications in Signal processing
  - Confidence Intervals for Spectra
  - Bispectrum Based Gaussianity Tests
  - Noise Floor Estimation in Radar
  - Confidence Intervals for Flight Parameters in Passive Sonar
  - Model Selection of Polynomial Phase Signals
- Summary
- Acknowledgements

Acknowledgements

- Many of the examples and simulations were implemented by Dr. D. Robert Iskander and Mr. Hwa-Tung Ong who also designed and implemented the *bootstrap demo*.
- The passive acoustic problem was the subject of Dr. David Reid's PhD. He collected the data and implemented the algorithms.
- The data from the JINDALEE over-the-horizon radar system was provided by the Defence Science & Technology Organisation in Salisbury, South Australia.
- I wish to thank my colleagues and students for their support and the Australian Research Council which partly provided financial support for some of these projects.

Cited References

[1] H. Akaike. Comments on "On Model Structure Testing in System Identification". *International J. Control*, 27:323–324, 1978.

[2] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, 1984.

[3] T. Ankenbrand and M. Tomassini. Predicting Multivariate Financial Time Series Using Neural Networks: the Swiss Bond Case. In *Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr)*, pages 27–33, New York, USA, 1996. IEEE.

[4] G. Archer and K. Chan. Bootstrapping Uncertainty in Image Analysis. In *Proceedings of the 12th Symposium in Computational Statistics*, pages 193–198, Heidelberg, Germany, 1996. Physica-Verlag.

[5] C. Banga and F. Ghorbel. Optimal Bootstrap Sampling for Fast Image Segmentation: Application to Retina Image. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-93)*, Minnesota, 1993.

[6] R. Bhar and C. Chiarella. Bootstrap Algorithms and Applications. In *Proceedings of the IEEE/IAFE 1996 Conference on Computational*

Intelligence for Financial Engineering (CIFEr), pages 168–182, New York, USA, 1996. IEEE.

[7] R. N. Bhattacharya and R. R. Rao. *Normal Approximation and Asymptotic Expansions*. Wiley, New York, 1976.

[8] J. F. Böhme and D. Maiwald. Multiple Wideband Signal Detection and Tracking from Towed Array Data. In M. Blanke and T. Söderström, editors, *Proceedings of the 10th IFAC Symposium on System Identification (SYSID 94)*, volume 1, pages 107–112, Copenhagen, July 1994. Danish Automation Soc.

[9] A. Bose. Edgeworth Correction by Bootstrap in Autoregressions. *The Ann. of Statist.*, 16:1709–1722, 1988.

[10] D. R. Brillinger. *Time Series: Data Analysis and Theory*. Holden-Day, 1981.

[11] P. Ciarlini. Bootstrap Algorithms and Applications. In *Advanced Mathematical Tools in Metrology III*, pages 12–23, Singapore, 1997. World Scientific.

[12] M. G. Cox, P. M. Harris, M. J. T. Milton, and P. T. Woods. A method for evaluating trends in ozone-concentration data and its application to data from the UK Rural Ozone Monitoring network. In *Advanced*

Mathematical Tools in Metrology III, pages 171–177, Singapore, 1997. World Scientific.

[13] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1967.

[14] P. M. Djurić and S. M. Kay. Parameter Estimation of Chirp Signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(12):2118–2126, December 1990.

[15] B. Efron. Bootstrap Methods. Another Look at the Jackknife. *The Ann. of Statist.*, 7:1–26, 1979.

[16] B. Efron. Computers and the Theory of Statistics: Thinking the Unthinkable. *SIAM Review*, 4:460–480, 1979.

[17] B. Efron. Better Bootstrap Confidence Interval. *J. Amer. Statist. Assoc.*, 82:171–185, 1987.

[18] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.

[19] B. G. Ferguson. A Ground Based Narrow-Band Passive Acoustic Technique for Estimating the Altitude and Speed of a Propeller Driven Aircraft. *J. Acoust. Soc. Am.*, 92(3):1403–1407, 1992.

[20] B. G. Ferguson and B. G. Quinn. Application of the Short-Time Fourier Transform and the Wigner-Ville Distribution to the Acoustic Localization of Aircraft. *J. Acoust. Soc. Am.*, 96(2):821–827, 1994.

[21] N. I. Fisher and P. Hall. Bootstrap Confidence Regions for Directional Data. *J. Amer. Statist. Assoc.*, 84:996–1002, 1989.

[22] N. I. Fisher and P. Hall. New Statistical Methods for Directional Data – I. Bootstrap Comparison of Mean Directions and the Fold Test in Palaeomagnet. *Geophys. J. International*, 101:305–313, 1990.

[23] N. I. Fisher and P. Hall. Bootstrap Algorithms for Small Samples. *J. Statist. Plan. Infer.*, 27:157–169, 1991.

[24] N. I. Fisher and P. Hall. General Statistical Test for the Effect of Folding. *Geophys. J. International*, 105:419–427, 1991.

[25] R. A. Fisher. On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, 1:1–32, 1921.

[26] J. Franke and W. Härdle. On Bootstrapping Kernel Estimates. *The Ann. of Statist.*, 20:121–145, 1992.

[27] D. A. Freedman. Bootstrapping Regression Models. *The Ann. of Statist.*, 9:1218–1228, 1981.

[28] P. Hall. Theoretical Comparison of Bootstrap Confidence Intervals (with discussion). *The Ann. of Statist.*, 16:927–985, 1988.

[29] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag New York, Inc., 1992.

[30] P. Hall and D. M. Titterington. The Effect of Simulation Order on Level Accuracy and Power of Monte Carlo Tests. *J. R. Statist. Soc.* **B**, 51:459–467, 1989.

[31] S. R. Hanna. Confidence Limits for Air Quality Model Evaluations, as Estimated by Bootstrap and Jackknife Resampling Methods. *Atmospheric Environment*, 23:1385–1398, 1989.

[32] T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall, 1990.

[33] D. R. Haynor and S. D. Woods. Resampling Estimates of Precision in Emission Tomography. *IEEE Transactions on Medical Imaging*, 8:337–343, 1989.

[34] R. Herman. Using the Bootstrap Method for Evaluating Variations Due to Sampling in Voltage Regulation Calculations. In *Proceedings of the Sixth Southern African Universities Power Engineering Conference (SAUPEC-96)*, pages 235–238, Witwatersrand, South Africa, 1996. Univ.

Witwatersrand.

[35] A. O. Hero III, J. A. Fessler, and M. Usman. Exploring Estimator Bias-Variance Tradeoffs Using the Uniform CR Bound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, 1996.

[36] M. J. Hinich. Testing for Gaussianity and Linearity of a Stationary Time Series. *J. Time Series Anal.*, 3:169–176, 1982.

[37] L. B. Jaeckel. Some Flexible Estimates of Location. *Ann. Math. Statist.*, 42:1540–1552, 1971.

[38] K. Kanatani and N. Ohta. Optimal Robot Self-Localization and Reliability Evaluation. In *Proceedings of the 5th European Conference on Computer Vision (ECCV'98)*, volume 2, pages 796–808, Berlin, Germany, 1998. Springer-Verlag.

[39] S. Kay. *Modern Spectral Estimation. Theory and Application*. Prentice Hall, 1988.

[40] J. P. Kreiss and J. Franke. Bootstrapping Stationary Autoregressive Moving Average Models. *J. Time Ser. Anal.*, 13:297–319, 1992.

[41] J.-P. Kreiss and G. Lien. Bootstrapping autoregressions with infinite order. In *Proceedings of the IEEE International Conference on Acoustics,*

Speech and Signal Processing (ICASSP-94), volume 6, pages 97–100, Adelaide, South Australia, April 1994.

[42] E. Kreyszig. *Introductory Functional Analysis with Applications*. J. Wiley & Sons, 1989.

[43] J. Krolik, G. Niezgoda, and D. Swingler. A bootstrap approach for evaluating source localization performance on real sensor array data. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-91)*, volume 2, pages 1281–1284, Toronto, Canada, May 1991.

[44] H. R. Künsch. The Jackknife and the Bootstrap for General Stationary Observations. *The Ann. of Statist.*, 17:1217–1241, 1989.

[45] D. Lai and G. Chen. Computing the Distribution of the Lyapunov Exponent from Time Series: The One-dimensional Case Study. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 5:1721–1726, 1995.

[46] J. M. Lange, H. M. Voigt, S. Burkhardt, R. Gobel, S. Burkhardt, and R. Gobel. The Intelligent Inspection Engine—a Real-Time Real-World Visual Classifier System. In *IEEE International Joint Conference on Neural Networks Proceedings*, pages 1810–1815, New York, USA, 1998. IEEE.

[47] E. L. Lehmann. *Testing Statistical Hypotheses*. Wadsworth and Brooks, 1991.

[48] R. Y. Liu and K. Singh. Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. In LePage and Billard, editors, *Exploring the limits of bootstrap*, New York, 1992. John Wiley.

[49] S. L. Marple Jr. *Digital Spectral Analysis with Applications*. Prentice-Hall, 1987.

[50] R. G. Miller. The Jackknife - A Review. *Biometrika*, 61:1–15, 1974.

[51] S Nagaoka and O. Amai. A Method for Establishing a Separation in Air Traffic Control Using a Radar Estimation Accuracy of Close Approach Probability. *J. Japan Ins. Navigation*, 82:53–60, 1990.

[52] S Nagaoka and O. Amai. Estimation Accuracy of Close Approach Probability for Establishing a Radar Separation Minimum. *J. Navigation*, 44:110–121, 1991.

[53] H. Ong, D. R. Iskander, and A. M. Zoubir. Detection of a Common Non-Gaussian Signal in Two Sensors Using The Bootstrap. In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 463–467, Banff, Alberta, Canada, 1997.

[54] H. Ong and A. M. Zoubir. Non-Gaussian Signal Detection from Multiple Sensors using the Bootstrap. In *Proceedings of the International Conference on Information, Communications & Signal Processing (ICICS'97)*, volume 1, pages 340–344, Singapore, September 1997.

[55] E. Paparoditis. Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *J. Multivar. Anal.*, 57:277–296, 1996.

[56] D. N. Politis and J. P. Romano. A General Resampling Scheme for Triangular Arrays of α-Mixing Random Variables with Application to the Problem of Spectral Density Estimation. *The Ann. of Statist.*, 20:1985–2007, 1992.

[57] D. N. Politis and J. P. Romano. Bootstrap Confidence Bands for Spectra and Cross-Spectra. *IEEE Transactions on Signal Processing*, 40:1206–1215, 1992.

[58] D. N. Politis and J. P. Romano. The Stationary Bootstrap. *Journal Am. Stat. Assoc.*, 89:1303–1313, 1994.

[59] B. Porat and B. Friedlander. Performance Analysis of Parameter Estimation Algorithms Based on High-Order Moments. *International J. Adaptive Control and Signal Processing*, 3:191–229, 1989.

[60] M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

[61] D. C. Reid, A. M. Zoubir, and B. Boashash. The Bootstrap Applied to Passive Acoustic Aircraft Parameter Estimation. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-96)*, volume VI, pages 3153–3156, Atlanta, May 1996.

[62] D. C. Reid, A. M. Zoubir, and B. Boashash. Aircraft Parameter Estimation based on Passive Acoustic Techniques Using the Polynomial Wigner-Ville Distribution. *J. Acoust. Soc. Am.*, 102(1):207–223, 1997.

[63] M. Rosenblatt. *Stationary Sequences and Random Fields*. Birkhauser, Boston, 1985.

[64] Special Session. The Bootstrap and Its Applications. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-94)*, volume VI, pages 65–79, Adelaide, Australia, 1994.

[65] J. Shao. Bootstrap Model Selection. *J. Am. Statist. Assoc.*, 91:655–665, 1996.

[66] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, 1995.

[67] Q. Shao and H. Yu. Bootstrapping the Sample Means for Stationary Mixing Sequences. *Stochastic Processes and Their Appl.*, 48:175–190, 1993.

[68] R. H. Shumway. Replicated Time-Series Regression: An Approach to Signal Estimation and Detection. In D. R. Brillinger and P. R. Krishnaiah, editors, *Handbook of Statistics*, volume 3, pages 383–408. North-Holland, 1983.

[69] T. Subba Rao and M. M. Gabr. Testing for Linearity of a Stationary Time Series. *J. Time Series Anal.*, 1:145–158, 1980.

[70] A. Swami, J. M. Mendel, and C. L. Nikias. *Higher-Order Spectral Analysis Toolbox for use with MATLAB*. The MathWorks, Inc., 1995.

[71] L. Tauxe, N. Kylstra, and C. Constable. Bootstrap Statistics for Paleomagnetic Data. *J. Geophysical Research*, 96:11723–11740, 1991.

[72] S. Tretter. Estimating the Frequency of a Noisy Sinusoid. *IEEE Transactions on Information Theory*, 31:832–835, 1985.

[73] J. K. Tugnait. Two-Channel Test for Common Non-Gaussian Signal Detection. *IEE Proceedings-F*, 140:343–349, 1993.

[74] A. M. Yacout, S. Salvatores, and Y. G. Orechwa. Degradation Analysis

Estimates of the Time-to-Failure Distribution of Irradiated Fuel Elements. *Nuclear-Technology*, 113(2):177–189, 1996.

[75] G. A. Young. Bootstrap: More Than a Stab in the Dark? *Statistical Science*, 9:382–415, 1994.

[76] Y. Zhang, D. Hatzinakos, and A. N. Venetsanopoulos. Bootstrapping Techniques in the Estimation of Higher-Order Cumulants from Short Data Records . In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-93)*, volume IV, pages 200–203, Minnesota, 1993.

[77] A. M. Zoubir. Bootstrap: Theory and Applications. In F. T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures and Implementations*, volume 2027, pages 216–235, San Diego, July 1993. Proceedings of SPIE.

[78] A. M. Zoubir. Multiple Bootstrap Tests and Their Application. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-94)*, volume VI, pages 69–72, Adelaide, Australia, 1994.

[79] A. M. Zoubir and B. Boashash. Advanced Signal Processing for Over-The-Horizon Radar: Application of Bootstrap Methods. Technical

report, Signal Processing Research Centre, Queensland University of Technology, March 1996. DSTO Report ref. 96/2–Part I-IV.

[80] A. M. Zoubir and B. Boashash. The Bootstrap and Its Application in Signal Processing. *IEEE Signal Processing Magazine*, 15(1):56–76, 1998.

[81] A. M. Zoubir and J. F. Böhme. Bootstrap Multiple Hypotheses Tests for Optimizing Sensor Positions in Knock Detection. In *Proceedings of 13th World Congress on Computation and Applied Mathematics, IMACS'91*, pages 955–958, Dublin, 1991.

[82] A. M. Zoubir and J. F. Böhme. Application of Higher Order Spectra to the Analysis and Detection of Knock in Combustion Engines. In B. Boashash, E. J. Powers, and A. M. Zoubir, editors, *Higher-Order Statistical Signal Processing*, pages 269–290. John Wiley, 1995.

[83] A. M. Zoubir and J. F. Böhme. Multiple Bootstrap Tests: An Application to Sensor Location. *IEEE Transactions on Signal Processing*, 43:1386–1396, 1995.

[84] A. M. Zoubir and D. R. Iskander. A Bispectrum Based Gaussianity Test Using The Bootstrap. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-96)*, volume V, pages 3029–3032, Atlanta, 1996.

[85] A. M. Zoubir and D. R. Iskander. On Bootstrapping the Confidence Bands for Power Spectra. Technical report, Signal Processing Research Centre, Queensland University of Technology, January 1996.

[86] A. M. Zoubir and D. R. Iskander. Bootstrap Model Selection for Polynomial Phase Signals. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP-98)*, pages 2229–2232, Seattle, USA, May 1998.

[87] A. M. Zoubir and D. R. Iskander. Bootstrapping Bispectra: An Application to Testing for Departure from Gaussianity of Stationary Signals. *IEEE Transactions on Signal Processing*, March 1999.

[88] A. M. Zoubir and D. R. Iskander. Bootstrap Based Modelling of a Class of Non-Stationary Signals. *IEEE Transaction on Signal Processing*, (under review).