NTNU – Trondheim
Norwegian University of Science and Technology

TTT4120 Digital Signal Processing
Fall 2017

Estimation Basics and Periodogram

Prof. Stefan Werner
stefan.werner@ntnu.no
Office B329

Department of Electronic Systems

Lecture in course book*

- Proakis, Manolakis Digital Signal Processing, 4th Ed.
  - 12.1 Random signals, correlation functions, and power spectra
  - 14.1.2 Estimation of the autocorrelation and power spectrum of random signals: The periodogram
  - 14.1.3 The use of DFT in power spectrum estimation
  - 14.2.1 The Bartlett method: Averaging periodograms
- A comprehensive overview of topics treated in the lecture, see "Introdukjon til statistisk signalbehandling" on ItsLearning

*Level of detail is defined by lectures and problem sets

Contents and learning outcomes

- Basics of estimation theory
  - Simple example: estimating the mean
  - Properties of good estimators
- Estimating the autocorrelation sequence
- Periodogram: crude estimate of the PDS

Introduction

- Autocorrelation sequence of a random signal $X[n]$

$$\gamma_{XX}[l] = E\{X[n]X[n+l]\}$$

- Power spectrum density of a random signal $X[n]$

$$\Gamma_{XX}(f) = \sum_{l=-\infty}^{\infty} \gamma_{XX}[l]e^{-j2\pi fl}$$

- Statistical averages require knowledge of *all realizations* or *an infinitely long realization from an ergodic process*
- In practice, access to a single realization of finite duration
- Can we still estimate statistical quantities and to what accuracy?

Basics of estimation theory

- Our problem becomes to estimate an unknown quantity, θ, (e.g., a statistical average) from a discrete-time waveform or a data-set
- We have the N-point data set $\boldsymbol{x} = \{x[0], x[1], \dots, x[N-1]\}$, which is a realization of a random process containing information on θ
- Determine θ based on the data, or define an estimator

$$\hat{\theta} = g(\boldsymbol{x}) = g(x[0], x[1], \dots, x[N-1])$$

where $g(\cdot)$ is some function
- Since $x[n]$ is a realization of $X[n]$, $\hat{\theta}$ is related to random variable

$$\hat{\Theta} = g(\boldsymbol{X}) = g(X[0], X[1], \dots, X[N-1])$$

5

Basics of estimation theory…

- How good is a particular estimator? How good can *any* estimate be?
- How to measure goodness of an estimate?

6

Simple example: estimating the mean

- Example 1: Estimate the mean m_X from an N-point realization of i.i.d. sequence $X[n] \sim N(m_X, \sigma_W^2)$
- Based on the N-point data set $\{x[0], x[1], \ldots, x[N-1]\}$, we would like to estimate m_X. Reasonable to estimate m_X as

$$\widehat{m}_X = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

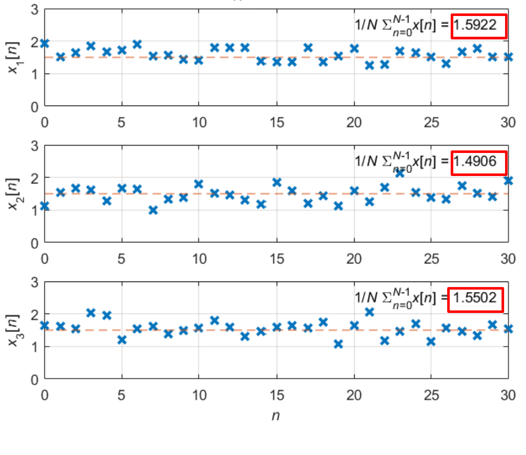
(which can be seen as an outcome of $\widehat{M}_X = \frac{1}{N} \sum_{n=0}^{N-1} X[n]$)

- How close is $\widehat{m}_X$ to m_X and what is the influence of N?

Simple example: estimating the mean…

- Three different realizations of $X[n] \sim N(1.5, \sigma_W^2 = 0.05)$

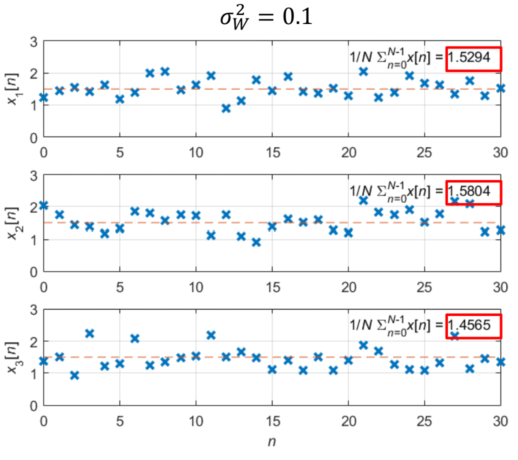
$\sigma_W^2 = 0.05$



Matlab
```
N = 31;
n = (0:N-1);
w = randn(1,N);
x = 1.5 + w;
plot(n,x,'x')
m_hat = mean(x)
```

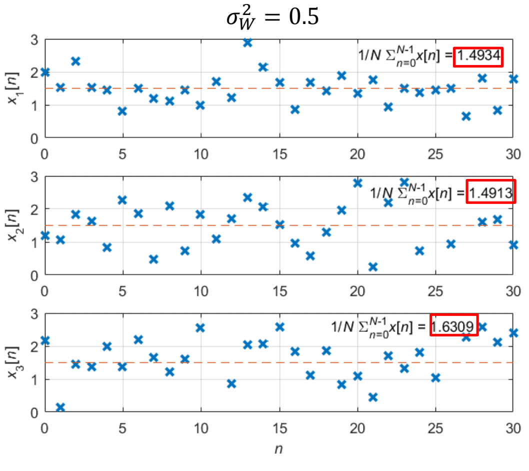
Simple example: estimating the mean…

- Three different realizations of $X[n] \sim N(1.5, \sigma_W^2 = 0.1)$



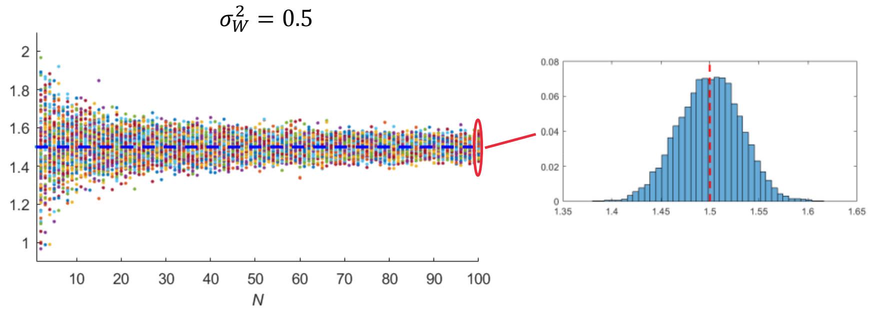
Simple example: estimating the mean…

- Three different realizations of $X[n] \sim N(1.5, \sigma_W^2 = 0.5)$



Simple example: estimating the mean…

- Varying number of data points N used for the estimation



- Each point (for a fixed N) corresponds to the estimate from a single realization

11

Simple example: estimating the mean…

- Observations from this simple example
  - Estimate depends on the realization (data available)
  - True value m_X is the mid point to all realizations of $\widehat{M}_X$
  - Variability of estimates increases with uncertainty
  - Variability of estimate across realizations decreases with N
  - Estimate approaches true value as N increases

- Let us calculate the mean and variance of $\widehat{M}_X$

12

Simple example: estimating the mean…

- Mean value of estimate

$$E\{\widehat{M}_X\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} X[n]\right\}$$
$$= \frac{1}{N}\sum_{n=0}^{N-1} E\{X[n]\}$$
$$= \frac{1}{N}\sum_{n=0}^{N-1} m_X = m_X$$

- On the average we get the true parameter

13

Simple example: estimating the mean…

- Variance of estimate

$$\sigma_{\widehat{M}_X}^2 = E\left\{\left(\widehat{M}_X - E\{\widehat{M}_X\}\right)^2\right\}$$
$$= E\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1} X[n] - m_X\right)^2\right\}$$
$$= \frac{1}{N^2} E\{\sum_{n=0}^{N-1}(X[n] - m_X)^2\}$$
$$= \frac{\sigma_W^2}{N}$$

- Variance of estimate goes to zero as N increases

14

Properties of good estimators

- An unbiased estimator provides the true value on average

$$m_{\widehat{\Theta}} = E\{\widehat{\Theta}\} = \theta$$

- A weaker requirement is asymptotic unbiasedness

$$\lim_{N\to\infty} m_{\widehat{\Theta}} = \lim_{N\to\infty} E\{\widehat{\Theta}\}$$
$$= \lim_{N\to\infty} E\{g(\boldsymbol{X})\} = \theta$$

- Small variance $\sigma_{\hat{\theta}}^2$: The estimates $\hat{\theta}$ are close to the true value θ irrespectively of the realization $\boldsymbol{x}$
- Variance decreasing for an increased number of observations, N

Properties of good estimators…

- An estimator is said to be consistent whenever, the estimate approaches the true value as $N \to \infty$, i.e.,

$$\lim_{N\to\infty} m_{\widehat{\Theta}} = \theta$$

$$\lim_{N\to\infty} \sigma_{\widehat{\Theta}}^2 = 0$$

- The simple averager in previous example is a consistent estimator

Estimation of autocorrelation

- Goal is to estimate the PDS of a signal from a single observation of the signal over a finite time interval
- The PDS is related to the autocorrelation sequence as

$$\Gamma_{XX}(f) = \sum_{l=-\infty}^{\infty} \gamma_{XX}[l]e^{-j2\pi fl}$$

with $\gamma_{XX}[l] = E\{X[n]X[n+l]\}$

- Given an N-point realization $\boldsymbol{x} = \{x[0]\ x[1] \ldots x[N-1]\}$, we would like to acquire a good estimate $\hat{\gamma}_{XX}[l]$ of $\gamma_{XX}[l]$

17

Estimation of autocorrelation…

- Approach 1: For lag l we can compute $N-|l|$ products. Compute the average over available products, i.e.,

$$\hat{\gamma}'_{XX}[l] = \frac{1}{N-|l|}\sum_{n=0}^{N-|l|-1} x[n]x[n+|l|]$$

- Is this estimator consistent?

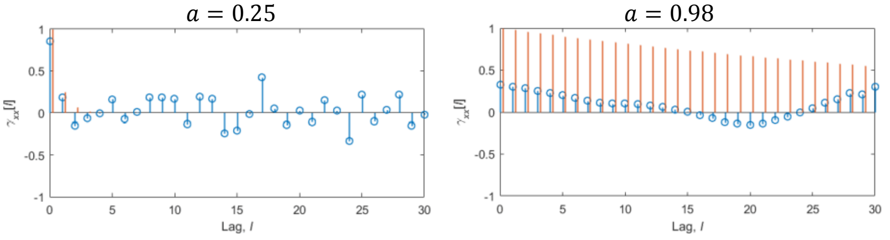
1. $E\left\{\frac{1}{N-|l|}\sum_{n=0}^{N-|l|-1} X[n]X[n+|l|]\right\} = \gamma_{XX}[l]$
2. $\lim_{N\to\infty} \mathrm{var}\left\{\frac{1}{N-|l|}\sum_{n=0}^{N-|l|-1} X[n]X[n+|l|]\right\} = 0$

Yes!

18

Estimation of autocorrelation…

- Estimate $\gamma_{XX}[l]$ from a realization of $X[n] = aX[n-1] + W[n]$ $0 \leq n \leq N-1 = 30$, $W[n] \sim N(0, \sigma_W^2)$

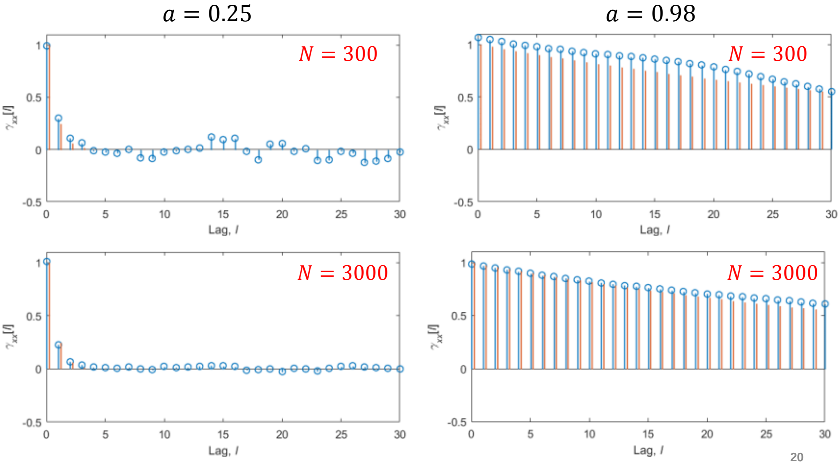


- Estimate: $\hat{\gamma}_{XX}[l] = \frac{1}{N-l}\sum_{n=0}^{N-1-l} x[n]x[n+l]\,, l = 0,1,\dots,N-1$
- As lag l increases, less products to average over ⇒ large errors
- Maximum lag to be estimated, $l_{\max}$, chosen such that $l_{\max} \ll N$

19

Estimation of autocorrelation…

- Increase the sample size: $N = 300$, and $N = 3000$



20

Estimation of autocorrelation…

```
Matlab
N = 30;
lmax = 21; % Max lag to compute < N
l = (-lmax:lmax);
a = 0.98;
varw = (1 - a^2);

% True autocorrelation function
gammaxx = varw/(1-a^2)*a.^abs(l);

% Generate data:
w = randn(N,1);
x = filter(1,[1 -a],sqrt(varw)*w);

% Compute ACF:
[gammaxx_est,lags] = xcorr(x,'unbiased',lmax);
stem(lags,gammaxx_est), hold on
stem(lags+.2,gammaxx,'Marker', 'none')
xlim([0 lmax])
```

21

Estimation of autocorrelation…

- Approach 2: For lag we can compute $N - |l|$ products. Compute the average over available products but normalize with N, i.e.,

$$\hat{\gamma}_{XX}[l] = \frac{1}{N}\sum_{n=0}^{N-|l|-1} x[n]x[n+|l|]$$

- Properties of this estimator
  1. Biased for $l \neq 0$
  2. Consistent for $|l| \ll N$

$$\lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{n=0}^{N-|l|-1} X[n]X[n+|l|]\right\} = \gamma_{XX}[l]$$

$$\lim_{N\to\infty} \operatorname{var}\left\{\frac{1}{N-|l|}\sum_{n=0}^{N-|l|-1} X[n]X[n+|l|]\right\} = 0$$

22

Estimation of autocorrelation…

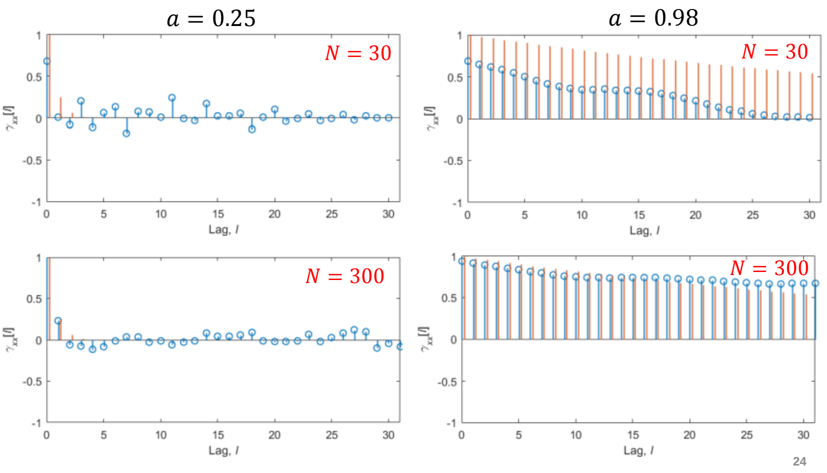
- Computing the bias

$$E\left\{\frac{1}{N}\sum_{n=0}^{N-|l|-1} X[n]X[n+|l|]\right\}$$
$$= \frac{1}{N}\sum_{n=0}^{N-|l|-1} E\{X[n]X[n+|l|]\}$$
$$= \frac{N-|l|}{N}\gamma_{XX}[l] = \left(1-\frac{|l|}{N}\right)\gamma_{XX}[l]$$
$$= w_B[l]\,\gamma_{XX}[l]$$

- Bias term disappears for fixed l when $N \to \infty$
- Triangular (Bartlett) window deemphasizes effects at lags $l \approx N$ $\Rightarrow$ lower variance

Estimation of autocorrelation…

- Revisit previous example: $N = 300$, and $N = 3000$



Estimation of autocorrelation…

```
Matlab
N = 30;
lmax = 21; % Max lag to compute < N
l = (-lmax:lmax);
a = 0.98;
varw = (1 - a^2);

% True autocorrelation function
gammaxx = varw/(1-a^2)*a.^abs(l);

% Generate data:
w = randn(N,1);
x = filter(1,[1 -a],sqrt(varw)*w);

% Compute ACF:
[gammaxx_est,lags] = xcorr(x,'biased',lmax);
stem(lags,gammaxx_est), hold on
stem(lags+.2,gammaxx,'Marker', 'none')
xlim([0 lmax])
```


Estimation of autocorrelation…

Comparing the of the two different estimators

- Approach 1: $\hat{\gamma}'_{XX}[l] = \frac{1}{N-|l|}\sum_{n=0}^{N-|l|-1} x[n]x[n+|l|]$
  - Consistent estimator (unbiased for any N and l)
- Approach 2: $\hat{\gamma}_{XX}[l] = \frac{1}{N}\sum_{n=0}^{N-|l|-1} x[n]x[n+|l|]$
  - Consistent estimator (asymptotically unbiased)
  - Lower variance than Approach 1
  - More effective for PDS estimation
  - Guarantees positive semidefinite autocorrelation sequence

Periodogram: crude estimate of the PDS

- We have the Fourier pair: $\hat{\gamma}_{XX}[l] \overset{\mathcal{F}}{\leftrightarrow} \Gamma_{XX}(f)$
- Periodogram:

$$\hat{\Gamma}_{XX}(f) = \sum_{l=-\infty}^{\infty} \hat{\gamma}_{XX}[l] e^{-j2\pi f l}$$

where $\hat{\gamma}_{XX}[l] = \frac{1}{N} \sum_{n=0}^{N-|l|-1} x[n]x[n+|l|]$

- Is the periodogram a good estimator for the PDS of $X[n]$?

27

Periodogram: crude estimate of the PDS

- With this choice of estimator, the periodogram becomes

$$\hat{\Gamma}_{XX}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2 = \frac{1}{N} |Y(f)|^2$$

where $Y(f)$ is the Fourier transform of

$$y[n] = \begin{cases} x[n] & \text{for } 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$

28

Periodogram: crude estimate of the PDS

- To see this, let us rewrite $\hat{\gamma}_{XX}[l]$

$$\hat{\gamma}_{XX}[l] = \begin{cases} \frac{1}{N}\sum_{n=0}^{N-|l|-1} x[n]x[n+|l|] & |l| < N \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{N}\sum_{n=0}^{N-|l|-1} y[n]y[n+|l|]$$

$$= \frac{1}{N}\sum_{n=-\infty}^{\infty} y[n]y[n+|l|]$$

$$= \frac{1}{N}\gamma_{YY}[l]$$

Periodogram: crude estimate of the PDS

- Putting the pieces together: take the DTFT of both sides

$$\hat{\Gamma}_{XX}(f) = \mathcal{F}\{\hat{\gamma}_{XX}[l]\}$$

$$= \mathcal{F}\left\{\frac{1}{N}\gamma_{YY}[l]\right\} = S_{YY}(f)$$

$$= \mathcal{F}\left\{\frac{1}{N}y[-l] * y[l]\right\}$$

$$= \frac{1}{N}|Y(f)|^2 = \frac{1}{N}\left|\underbrace{\sum_{n=-\infty}^{\infty} y[n]e^{-j2\pi f n}}_{\sum_{n=0}^{N-1} x[n]e^{-j2\pi f n}}\right|^2$$

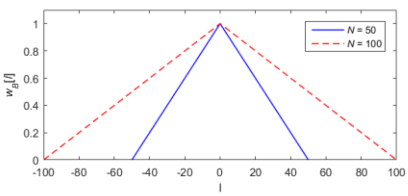
- Periodogram is obtained by taking the N-point DTFT of sequence $\{x[n]\}_{n=0}^{N-1}$

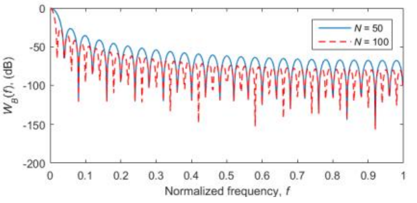
Periodogram: crude estimate of the PDS…

- Expected value of periodogram (bias)

$$E\{\hat{\Gamma}_{XX}(f)\} = E\left\{\mathcal{F}\{\hat{\gamma}_{XX}[l]\}\right\} = \mathcal{F}\left\{E\{\hat{\gamma}_{XX}[l]\}\right\}$$
$$= \mathcal{F}\{w_B[l]\gamma_{XX}[l]\} = W_B(f) * \Gamma_{XX}(f)$$

where $W_B(f)$ is the Fourier transform of the Bartlett window





$$w_B[l] = \begin{cases} 1 - \frac{|l|}{N}, & |l| \leq N-1 \\ 0, & \text{otherwise} \end{cases}$$

$$W_B(f) = \frac{1}{N}\left(\frac{\sin \pi f N}{\sin \pi f}\right)^2$$

Periodogram: crude estimate of the PDS…

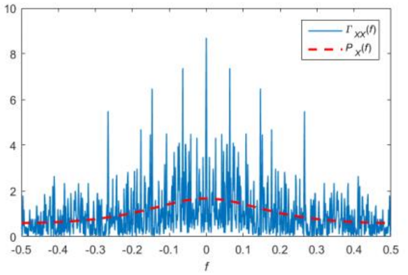
- Convolution with $W_B(f)$ results in spectrum spreading
  - Increasing window length reduces spectral leakage
- Frequency resolution is adequate for most situations
- Periodogram is asymptotically unbiased
- Peridogram is not a consistent estimator
  - That is, variance of estimate does not approach 0 as $N \rightarrow \infty$
  - For a Gaussian process $\text{var}\{\hat{\Gamma}_{XX}(f)\} \geq \Gamma_{XX}^2(f)$

∴ Periodogram is not a good estimator for the PDS

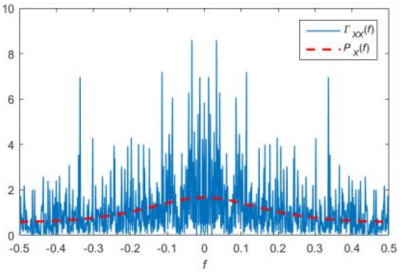
Periodogram: crude estimate of the PDS…

- Estimate $\Gamma_{XX}(f)$ from a realization of $X[n] = aX[n-1] + W[n]$
  $0 \leq n \leq N-1$, $W[n] \sim N(0, \sigma_W^2)$

$a = 0.25, N = 512$



$a = 0.25, N = 1024$



- Increasing N does not reduce variance

33

Improving the periodogram

- Use a different window function
  - Hamming, Kaiser
  - Reduces the spectral leakage and spread
  - Leads to a modified periodogram
- Take average of several periodograms
  - Split data into several blocks of length M
  - Compute periodogram for each block
  - Average over all computed periodograms
- Nonparametric methods: no assumptions made on how data were generated

34

Averaging periodogram: Bartlett method

$$\ldots, \underbrace{x[0], x[1], \ldots x[M-1]}_{M}, \underbrace{x[M], x[N+1], \ldots, x[2M-1]}_{M}, \underbrace{x[2M], x[2M+1] \ldots}_{M}.$$

- Break up $x[n]$ into K non-overlapping segments of length M

$$x_i[n] = x[n+iM], \quad \begin{matrix} i = 0, 1, \ldots, K-1 \\ n = 0, 1, \ldots, M-1 \end{matrix}$$

- Calculate the periodogram for each segment

$$\hat{\Gamma}_{XX}^{(i)}(f) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_i[n] e^{-j2\pi f n}\right|^2, \; i = 0, 1, \ldots, K-1$$

- Average the periodograms for the K segments

$$\hat{\Gamma}_{XX}^{B}(f) = \frac{1}{K}\sum_{n=0}^{K-1} \hat{\Gamma}_{XX}^{(i)}(f)$$

35

Averaging periodogram: Bartlett …

- Statistical properties
  - Mean value

$$E\{\hat{\Gamma}_{XX}^{B}(f)\} = \frac{1}{K}\sum_{n=0}^{K-1} E\left\{\hat{\Gamma}_{XX}^{(i)}(f)\right\} = W_B(f) * \Gamma_{XX}(f)$$

  - Variance

$$\mathrm{var}\{\hat{\Gamma}_{XX}^{B}(f)\} = \frac{1}{K}\mathrm{var}\{\hat{\Gamma}_{XX}(f)\}$$

- Bartlett window

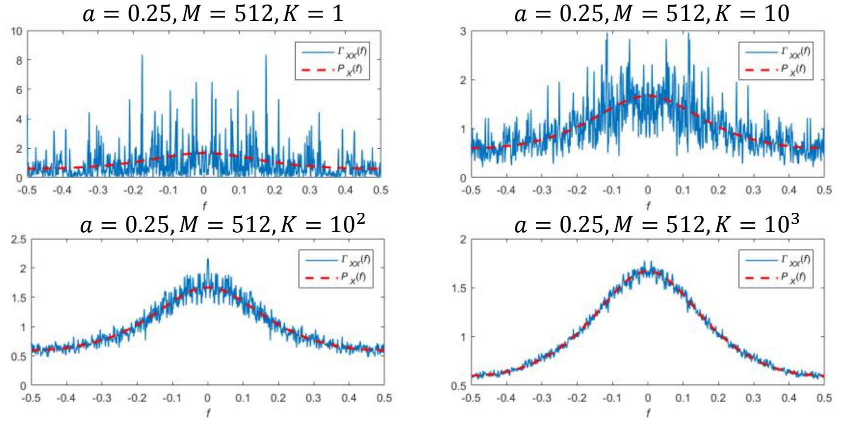
$$w_B[n] = \begin{cases} 1 - \frac{|m|}{M}, & |m| \le M-1 \\ 0, & \text{otherwise} \end{cases}$$

$$W_B(f) = \frac{1}{M}\left(\frac{\sin \pi f M}{\sin \pi f}\right)^2$$

36

Averaging periodogram: Bartlett …

- Estimate $\Gamma_{XX}(f)$ from a realization of $X[n] = aX[n-1] + W[n]$ $0 \leq n \leq N-1$, $W[n] \sim N(0, \sigma_W^2)$



37

Summary

- Today we discussed:
  - Basics of estimation theory
  - Nonparametric power density spectrum (PDS) estimation

- Next time:
  - Parametric PDS estimation

38