4. Model-Free and Model-Based Estimation of Random Processes

4.1. Model-free estimation of random processes

In this section $\{X(n)\}$ is a WSS process with

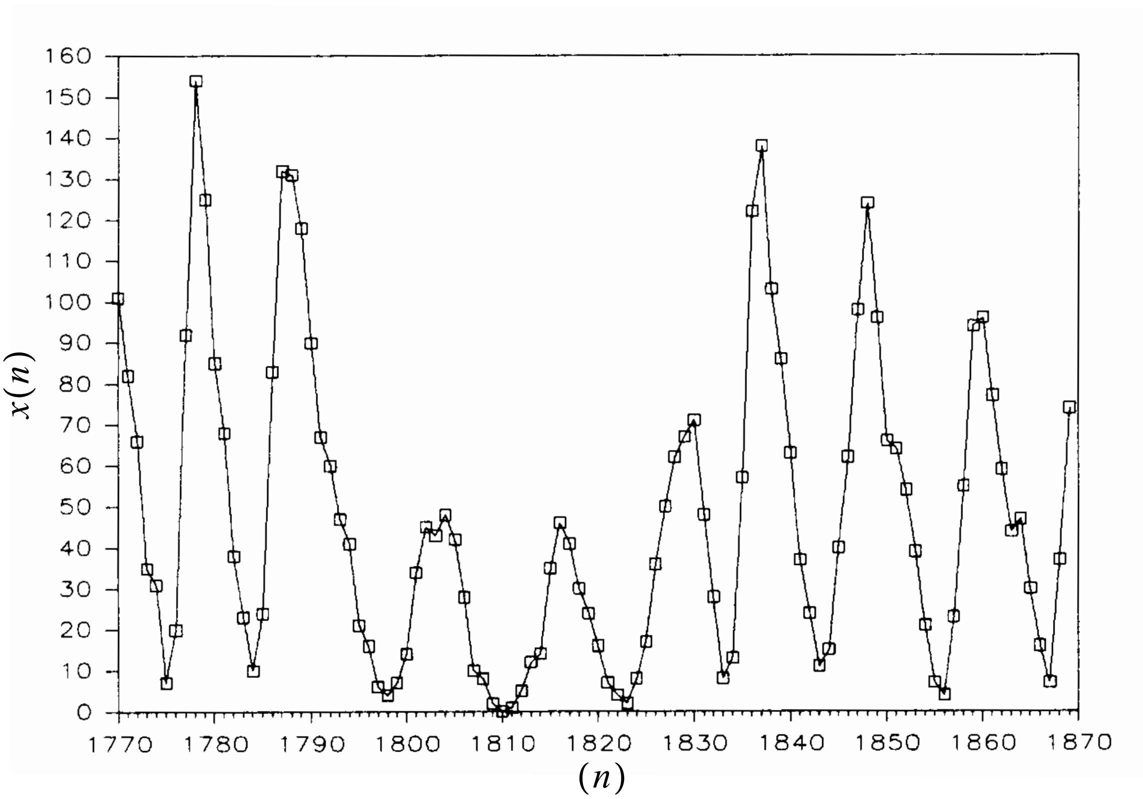
- mean value: $\mu_X \equiv \mathbf{E}[X(n)]$
- autocorrelation function: $R_{XX}(k) \equiv \mathbf{E}[X(n)X(n+k)]$

The autocovariance function of $\{X(n)\}$ is

$$C_{XX}(k) \equiv \mathbf{E}[(X(n)-\mu_X)(X(n+k)-\mu_X)] = R_{XX}(k)-\mu_X^2$$

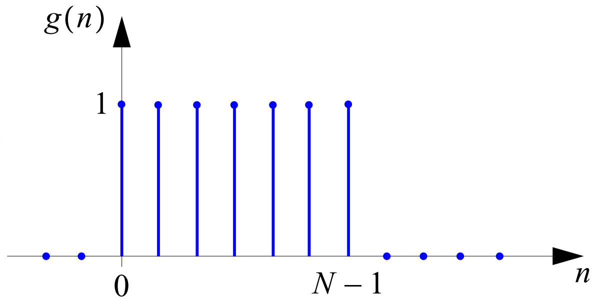
- ***Observed sequence:***
  We assume that $\{X(0), \ldots, X(N-1)\}$ can be observed.

Example 1: Wölfer sunspot numbers



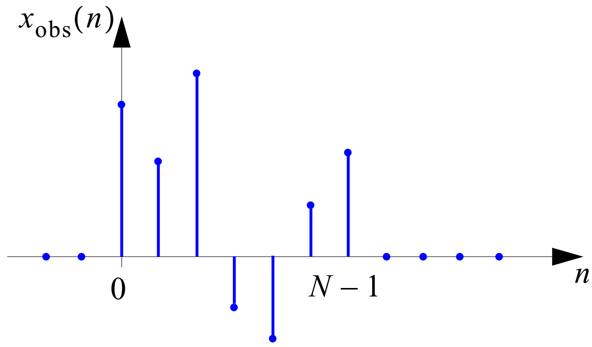
Defining the window function

$$g(n) \equiv \begin{cases} 1\ ; & n \in \{0, \ldots, N-1\} \\ 0\ ; & \text{otherwise} \end{cases}$$



the observed sequence reads:

$$X_{\text{obs}}(n) = g(n)X(n)$$



4.1.1. Estimation of the mean-value

- ***Arithmetic mean:***

$$\hat{\mu}_X \equiv \overline{X} = \frac{1}{N}\sum_{n=0}^{N-1} X(n)$$

- ***Mean and variance of $\overline{X}$:***

  - Mean: $\overline{X}$ is an unbiased estimator of μ_X:

$$\mu_{\overline{X}} = \mu_X$$

  - Variance:

$$\sigma_{\overline{X}}^2 = \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left[1-\frac{|k|}{N}\right]C_{XX}(k)$$

Special case: When $\{X(n)\}$ is an uncorrelated process:

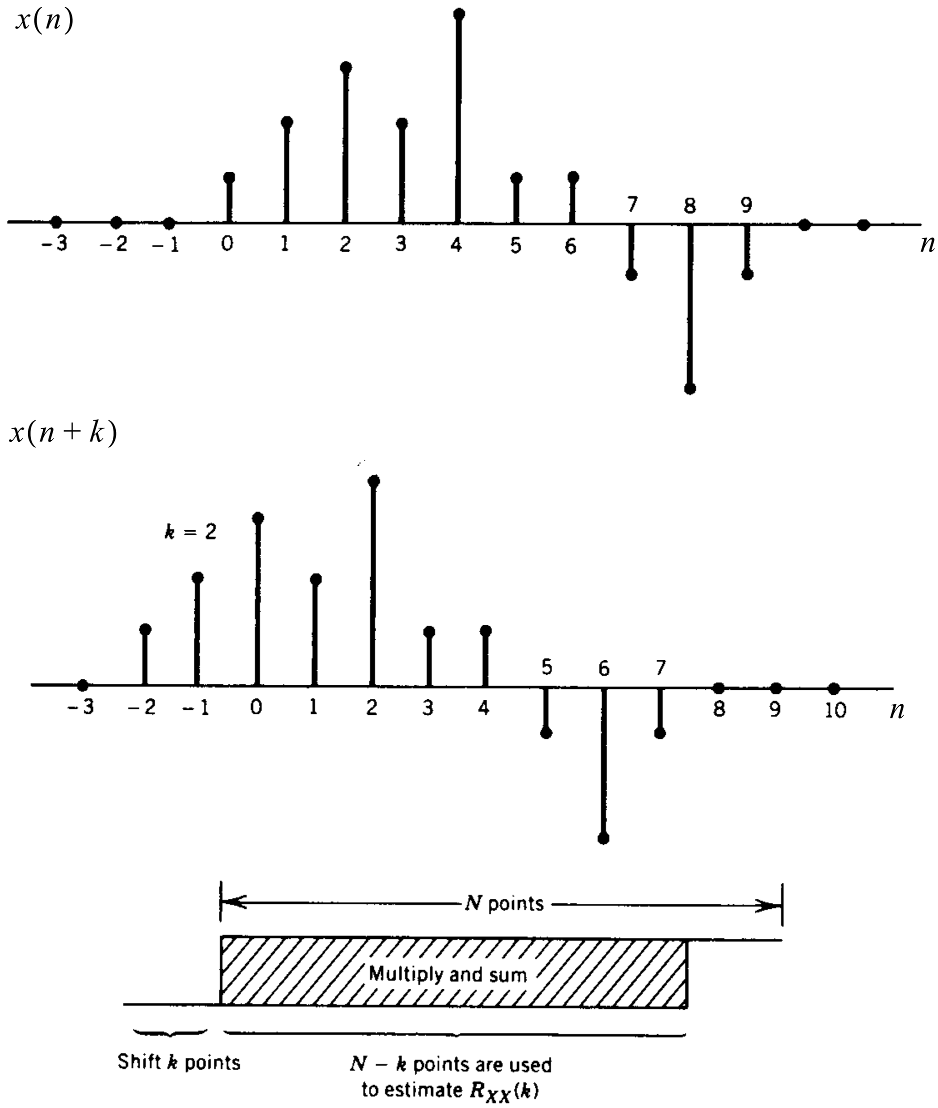
$$\sigma_{\overline{X}}^2 = \frac{1}{N}C_{XX}(0) = \frac{1}{N}\sigma_X^2$$

Proof: See Exercise 5.1.

4.1.2. Estimation of the autocorrelation function:

- ***Biased sample autocorrelation function:***

$$\hat{R}_{XX}(k) \equiv \begin{cases} \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-k-1} X(n)X(n+k) \;; & k = 0, \ldots, N-1 \\ \hat{R}_{XX}(-k) \;; & k = -(N-1), \ldots, -1 \\ 0 \;; & |k| \geq N \end{cases} \tag{4.1}$$



To show that the sample autocorrelation function $\hat{R}_{XX}(k)$ is biased we recast it as:

$$\begin{aligned}\hat{R}_{XX}(k) &= \frac{1}{N}\sum_{n=-\infty}^{\infty} X_{\text{obs}}(n)X_{\text{obs}}(n+k) \\ &= \frac{1}{N}\sum_{n=-\infty}^{\infty} g(n)g(n+k)X(n)X(n+k)\end{aligned}$$

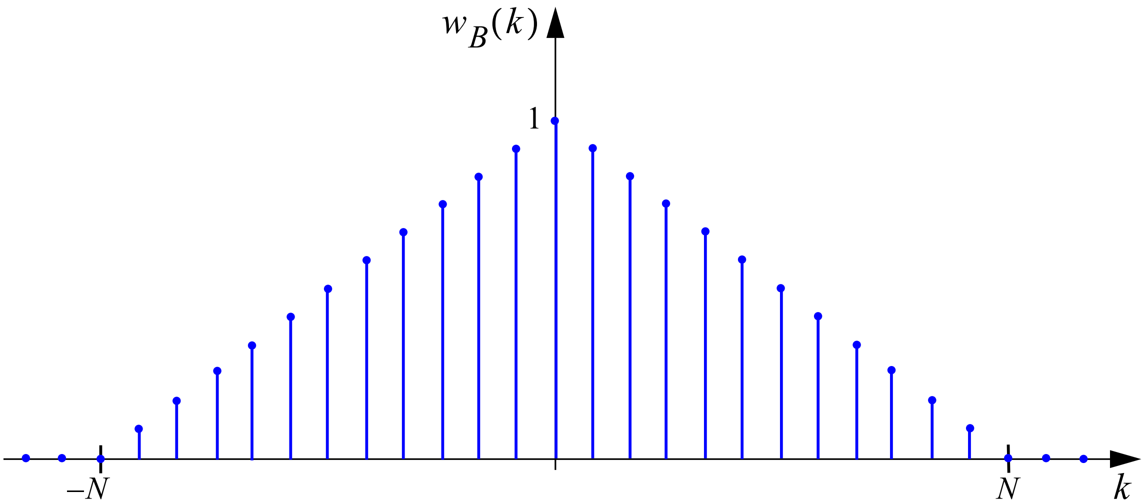
Taking the expectation on both side yields

$$\mathbf{E}[\hat{R}_{XX}(k)] = \frac{1}{N}R_{gg}(k)R_{XX}(k)$$

The function

$$w_B(k) \equiv \frac{1}{N}R_{gg}(k) = \begin{cases} 1-\dfrac{|k|}{N} & ; \quad |k| < N \\ 0 & ; \quad \text{otherwise} \end{cases}$$

is called the **Bartlett window**.



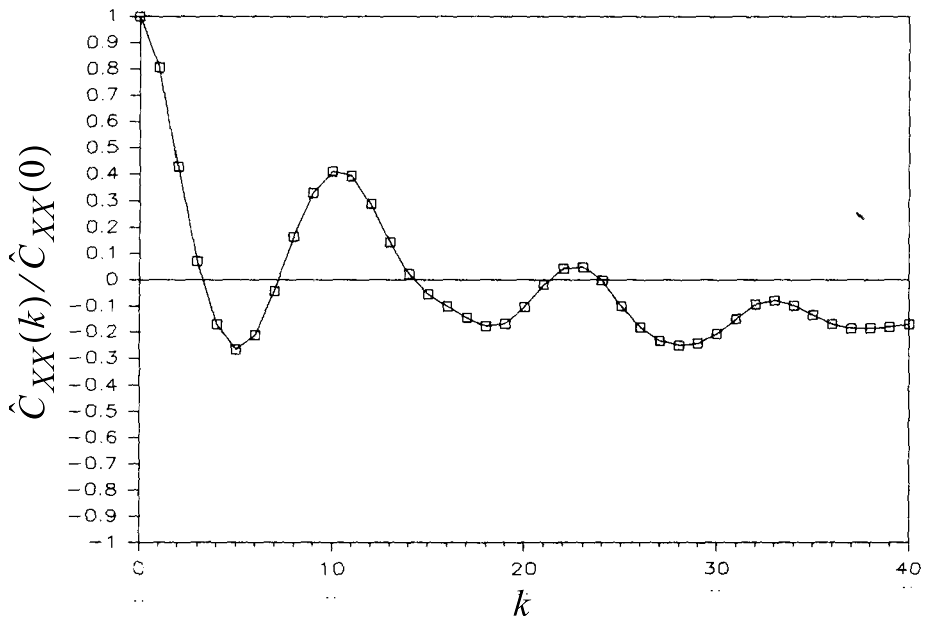
With this definition, the bias of $\hat{R}_{XX}(k)$ can be recast as

$$\mathbf{E}[\hat{R}_{XX}(k)] = w_B(k)R_{XX}(k) \tag{4.2}$$

- ***Biased sample autocovariance:***

$$\hat{C}_{XX}(k) = \hat{R}_{XX}(k) - \hat{\mu}_X^2$$

Example 1: Wölfer sunspot numbers



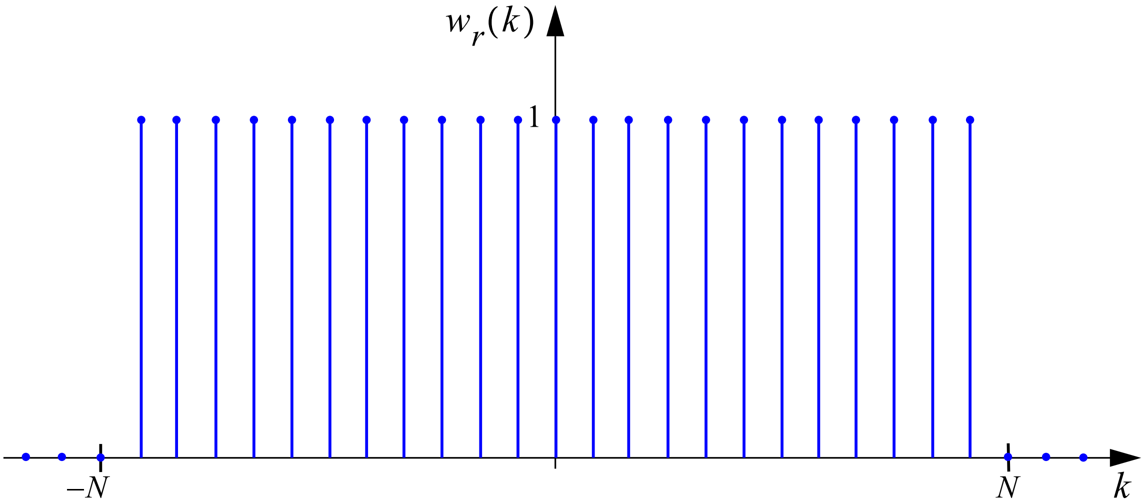
- ***Unbiased sample autocorrelation function:***

$$\widehat{R}_{XX}(k) \equiv \begin{cases} \frac{1}{N-k} \sum_{n=0}^{N-k-1} X(n)X(n+k) \ ; & k = 0, \ldots, N-1 \\ \widehat{R}_{XX}(-k) \ ; & k = -(N-1), \ldots, -1 \\ 0 \ ; & |k| \geq N \end{cases}$$

$\widehat{R}_{XX}(k)$ is unbiased for $|k| < N$:

$$\mathbf{E}[\widehat{R}_{XX}(k)] = w_r(k) R_{XX}(k)$$

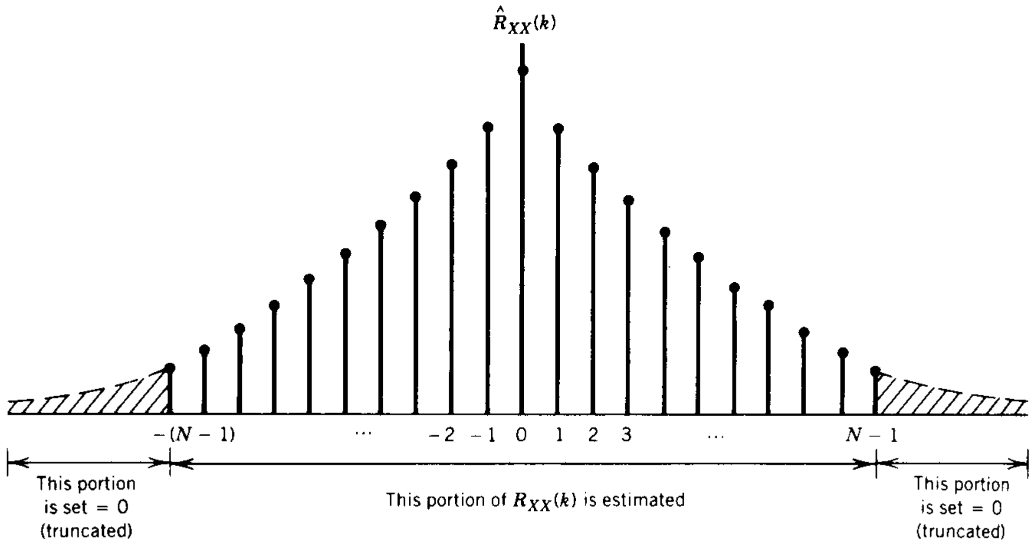
where $w_r(k)$ is the centered rectangular function:



- ***Properties of the sample autocorrelation functions:***

  - $\hat{R}_{XX}(k) = w_B(k)\overset{\frown}{R}_{XX}(k)$

  - With N observations, we can only estimate $R_{XX}(k)$ for $|k| < N$.



  - In general, it is difficult to calculate the variance of the sample autocorrelation functions since the computation involves fourth moments of the form $\mathbf{E}[X(n)X(n+m)X(k)X(k+m)]$.
    In the Gaussian case these moments can be evaluated and the variance of the sample autocorrelation functions can be calculated (See Exercise 9.8 of [Shanmugan]).

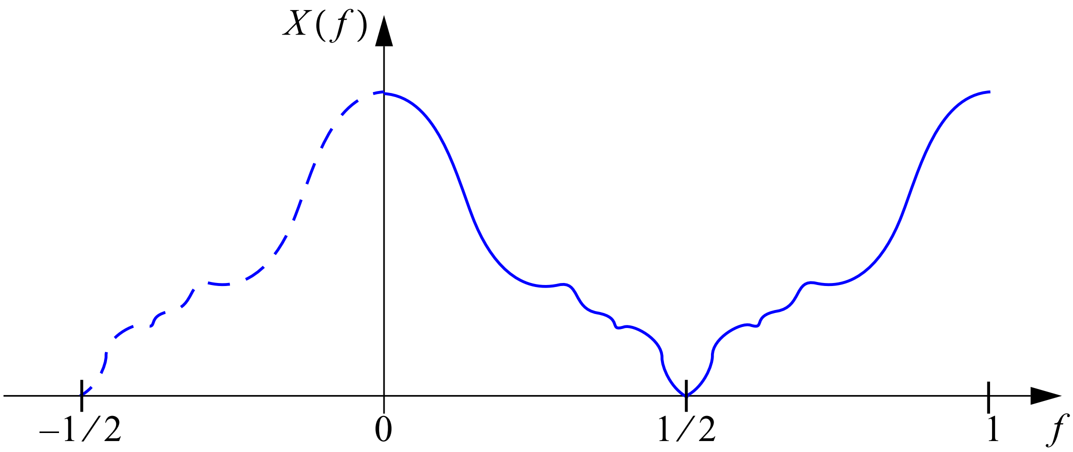
- A general conclusion is that the variance of $\hat{R}_{XX}(k)$ and $\widehat{R}_{XX}(k)$ increases with $|k|$ since the number of observations considered in the computation of these values is $N - |k|$.

4.1.3. Estimation of the power spectral density:

- ***Continuous-frequency periodogram:***

Let us start from the slightly differently reformulated Fourier transform:

$$X(f) = \sum_{n-0}^{N-1} x(n)\exp(-j2\pi nf) \qquad f \in [0, 1)$$



The periodogram of $X_{obs}(n)$ is defined to be

$$\hat{S}_{XX}(f) = F\{\hat{R}_{XX}(k)\}$$

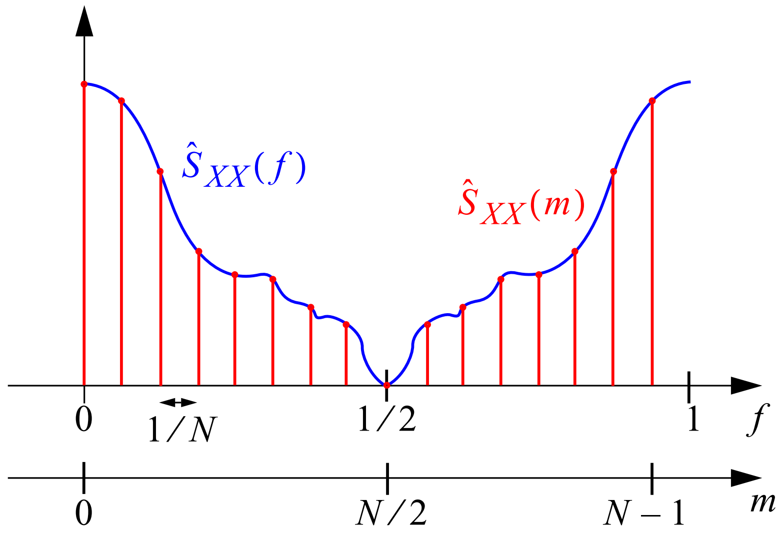
$$= \frac{1}{N}\left|\sum_{n-0}^{N-1} X(n)\exp(-j2\pi nf)\right|^2 = \frac{1}{N}|F\{X_{obs}(n)\}(f)|^2 \qquad f \in [0, 1)$$

Proof:

□

- ***Discrete-frequency periodogram:***

$$\hat{S}_{XX}(m) = \hat{S}_{XX}(f)\Big|_{f = m/N} \qquad m = 0, \ldots, N-1$$



- ***Discrete Fourier transform:***
The discrete Fourier transform and the inverse DFT are defined according to

$$X_d(m) = F_d\{x(n)\} \equiv \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \exp\left(-j2\pi\frac{nm}{N}\right)$$

$$x(n) = F_d^{-1}\{X_d(m)\} \equiv \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_d(m) \exp\left(j2\pi\frac{nm}{N}\right)$$

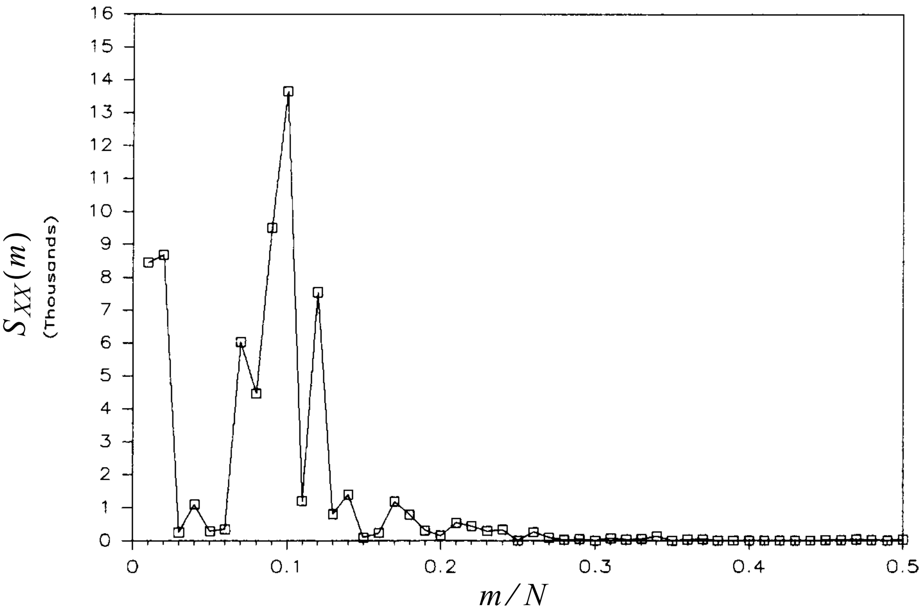
Relation between the discrete Fourier transform and the (continuous-frequency) Fourier transform:

$$X_d(m) = \frac{1}{\sqrt{N}} X(f)\Big|_{f = m/N} \qquad m = 0, \ldots, N-1$$

In particular, the discrete-frequency periodogram can be computed as

$$\hat{S}_{XX}(m) \equiv \left|F_d\{X_{obs}(n)\}(m)\right|^2$$

Example 1: Wölfer sunspot numbers



- ***Bias of the periodogram:***

  Because the Fourier transform is a linear operation, we have

$$\mathbf{E}[\hat{S}_{XX}(f)] = F\{\mathbf{E}[\hat{R}_{XX}(k)]\}$$

It follows from (4.2) that:

$$\mathbf{E}[\hat{S}_{XX}(f)] = F\{w_B(k)R_{XX}(k)\}$$

$$= W_F(f) * S_{XX}(f)$$

The Fourier transform

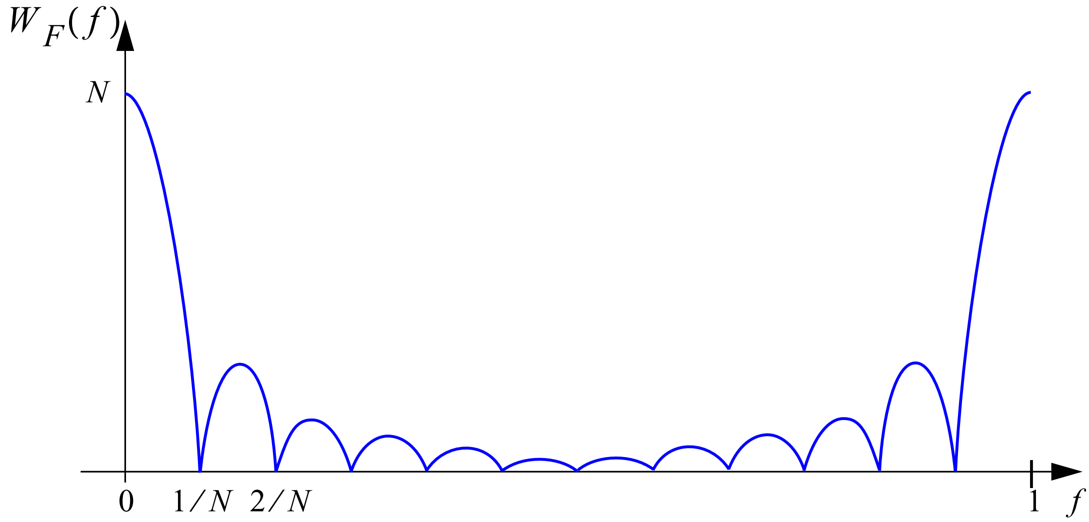
$$W_F(f) \equiv F\{w_B(k)\} = \frac{1}{N}\left(\frac{\sin(\pi f N)}{\sin(\pi f)}\right)^2$$

of the Bartlett window is called the Féjer kernel.

Proof: It can be easily shown that the Fourier spectrum of $R_{gg}(k)$ is

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

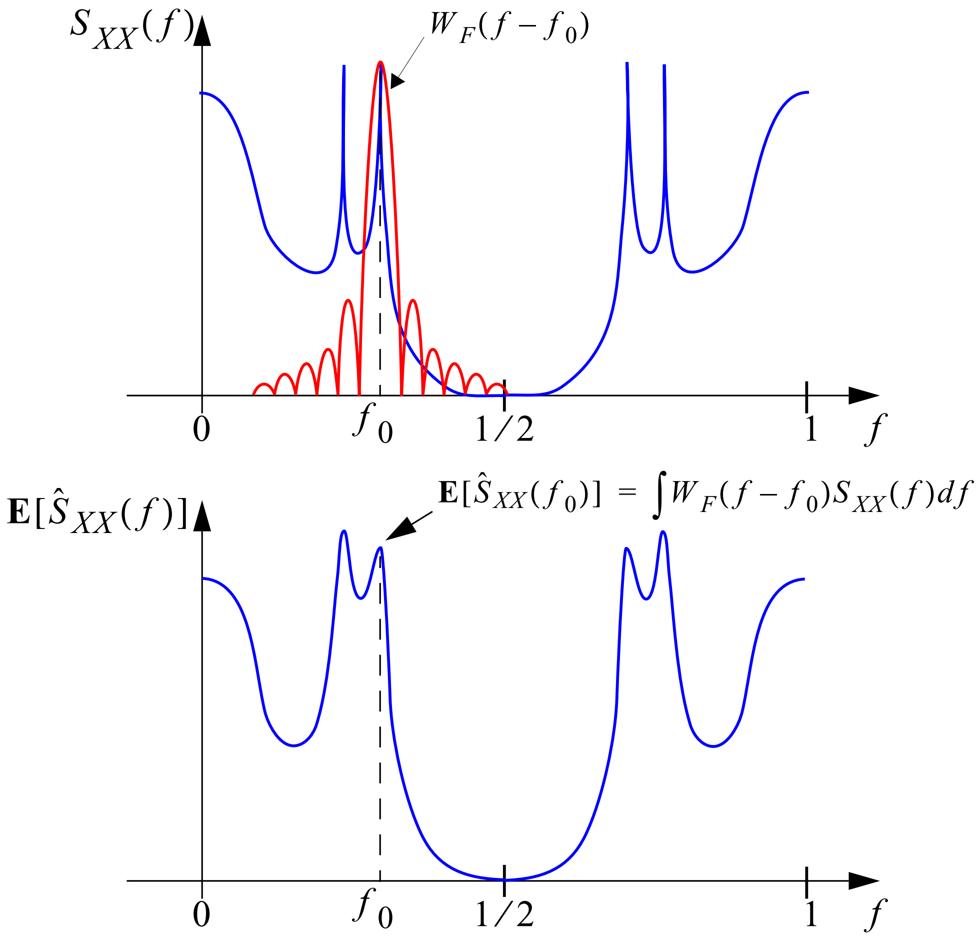
where $G(f) \equiv F\{g(n)\}$.



In summary, the bias of $\hat{S}_{XX}(f)$ and $\hat{S}_{XX}(m)$ are given by

$$\mathbf{E}[\hat{S}_{XX}(f)] = W_F(f) * S_{XX}(f)$$
$$\mathbf{E}[\hat{S}_{XX}(m)] = [W_F(f) * S_{XX}(f)]\Big|_{f = m/N}$$

- ***Spectral leakage:***



As N increases to infinity, $W_F(f) \to \delta(f)$, so that

$$\mathbf{E}[\hat{S}_{XX}(f)] \to S_{XX}(f),$$

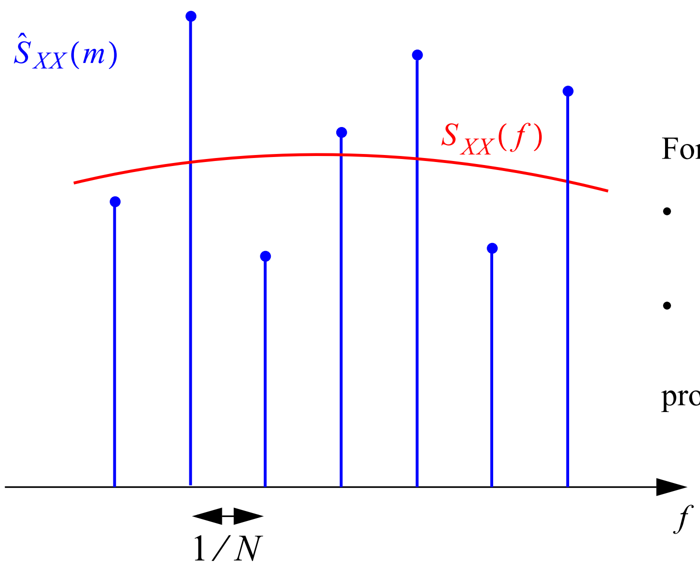
i.e. $\hat{S}_{XX}(f)$ and $\hat{S}_{XX}(m)$ are asymptotically unbiased.

- ***Variance of the periodogram:***
  The following asymptotic results are valid for a large classes of stochastic processes, and in particular for ARMA processes.

> As the number N of observations tends to infinity,
>
> $$\sigma^2_{\hat{S}_{XX}(f)} \to \begin{cases} 2S_{XX}(f)^2 & ; \quad f = 0, 1/2 \\ S_{XX}(f)^2 & ; \quad \text{otherwise} \end{cases}$$
>
> $$\Sigma_{\hat{S}_{XX}(f_1)\hat{S}_{XX}(f_2)} \to 0 \qquad \text{for any} \qquad f_1, f_2 \in \left[0, \frac{1}{2}\right], f_1 \neq f_2$$

Hence,

- Any two "different" samples of the periodogram are asymptotically uncorrelated.
  Remember that $\hat{S}_{XX}(f)$ and consequently $\hat{S}_{XX}(m)$ are even functions.
- As N increases the variance of the periodogram does not vanish but stabilizes to a value. This value coincides with the asymptotic mean of the periodogram when $f \neq 0, 1/2$.

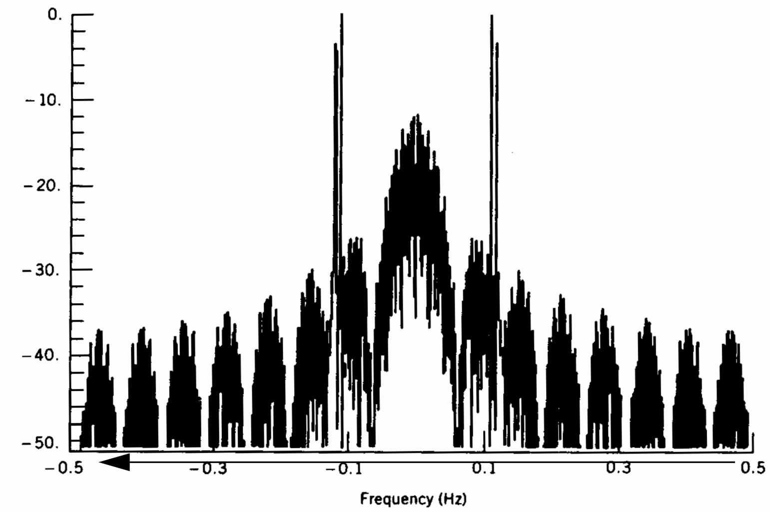


For large values of N:

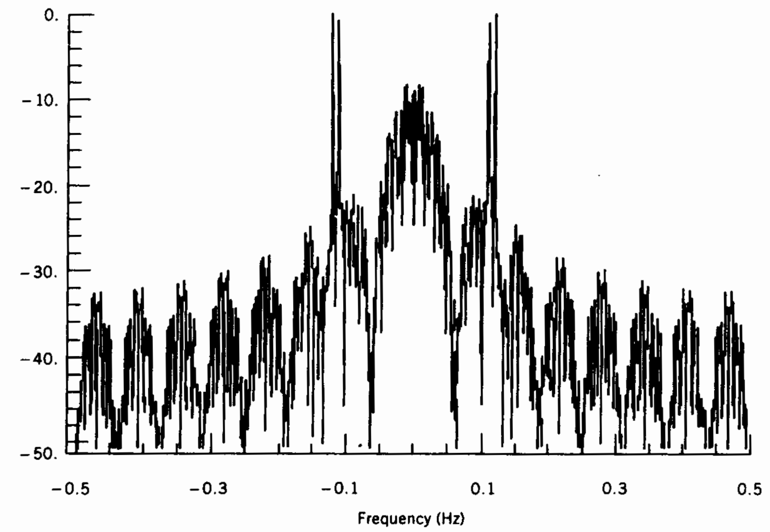
- $\mathbf{E}[\hat{S}_{XX}(m)] \approx S_{XX}(f)\big|_{f = \frac{m}{N}}$
- $\sigma_{\hat{S}_{XX}(m)} \approx S_{XX}(f)\big|_{f = \frac{m}{N}}$

provided $m \neq 0, N/2$

These two properties are responsible of the erratic nature of the periodogram (see the periodogram of the sunspot numbers).
Increasing the number of samples increases the spectral resolution only.



(a) $N = 4096$



(b) $N = 1024$

- ***Smoothing through windowing:***

Windowing aims at reducing the variability of the estimated spectrum.

A **lag window** $w(k)$ is a sequence satisfying the following properties:

- $w(k)$ is even, i.e $w(k) = w(-k)$.
- $w(k) = 0$ for $|k| > N$
- $w(0) = 1$

The **Blackman-Tukey estimator** of the spectrum is of the form

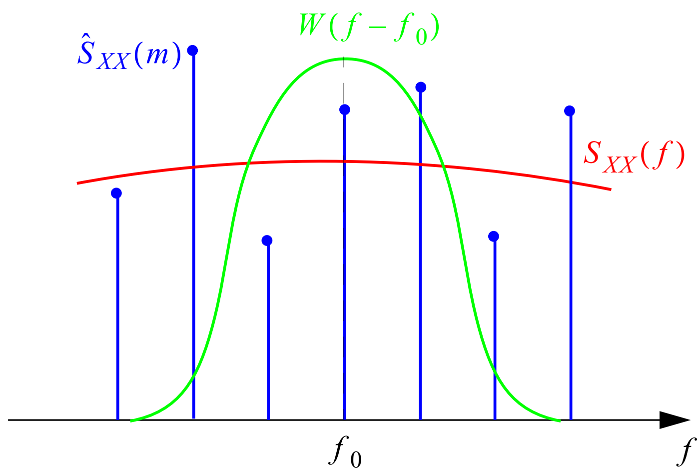
$$\hat{S}_{XX}^{(W)}(f) = F\{w(k)\hat{R}_{XX}(k)\}$$

where $w(k)$ is a given lag window with Fourier transform $W(f)$.

Making use of the property of the Fourier transform, we obtain

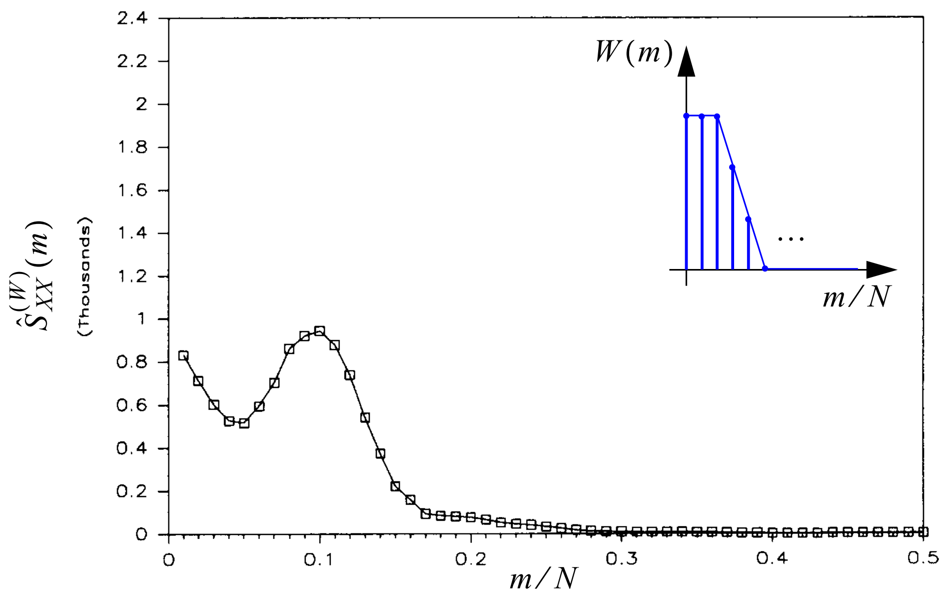
$$\hat{S}_{XX}^{(W)}(f) = W(f) * \hat{S}_{XX}(f)$$

Usually, the **spectral window** $W(f)$ is selected to have a narrow main lobe and low sidelobes. The above convolution corresponds to a local weighted averaging of $\hat{S}_{XX}(f)$.

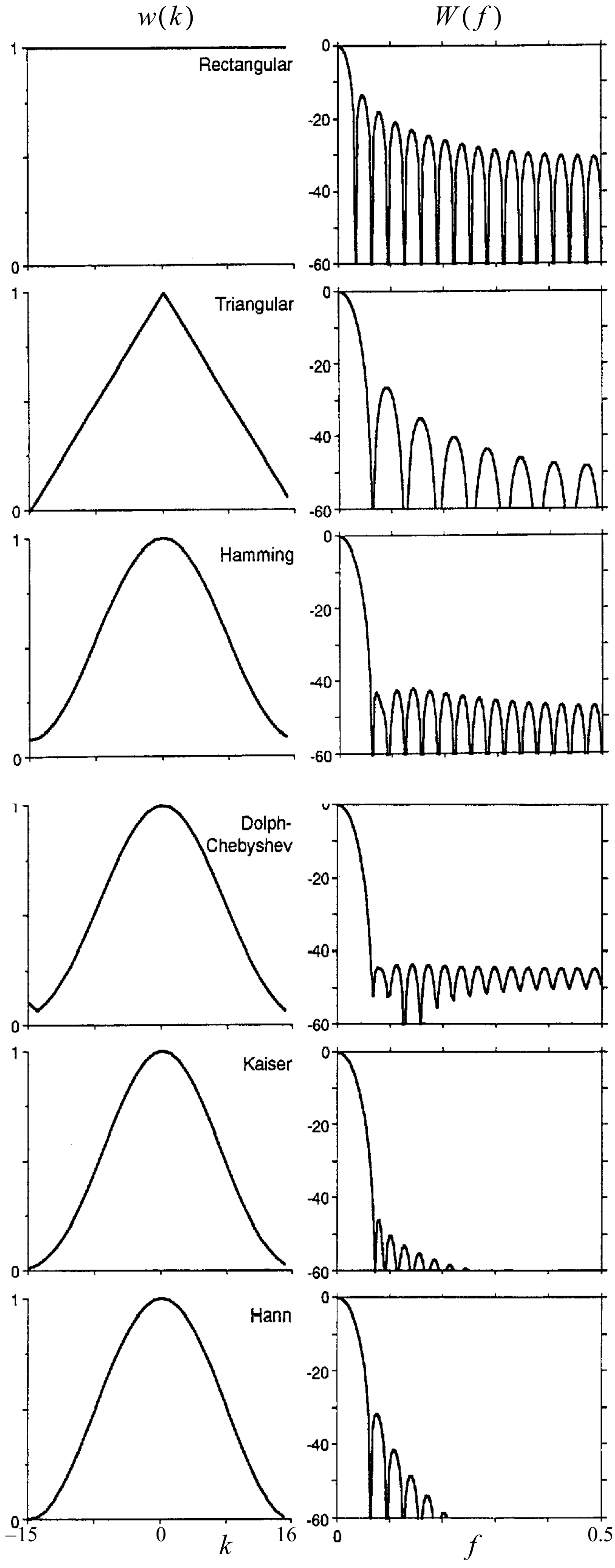


This averaging operation reduces the variability of $\hat{S}_{XX}^{(W)}(f)$ but also leads to a reduction of the spectral resolution.

Example 1: Wölfer sunspot numbers



Some well-known lag windows:

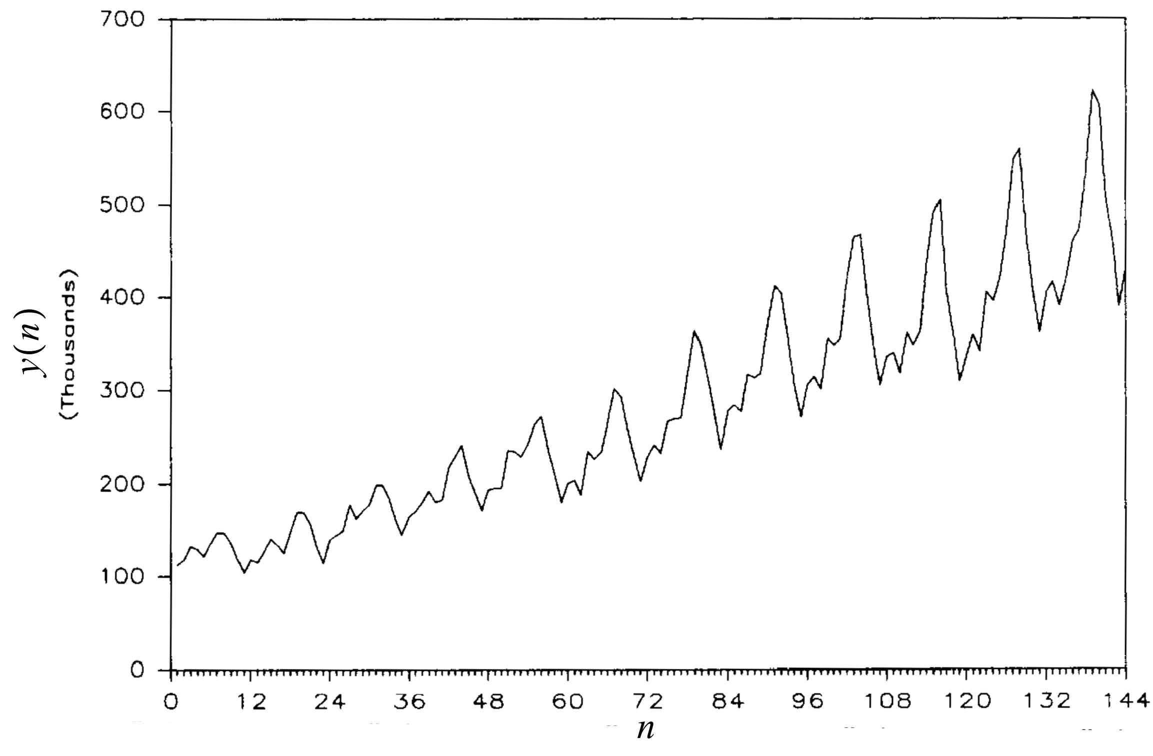


4.2. Parametric (model-based) estimation of random processes

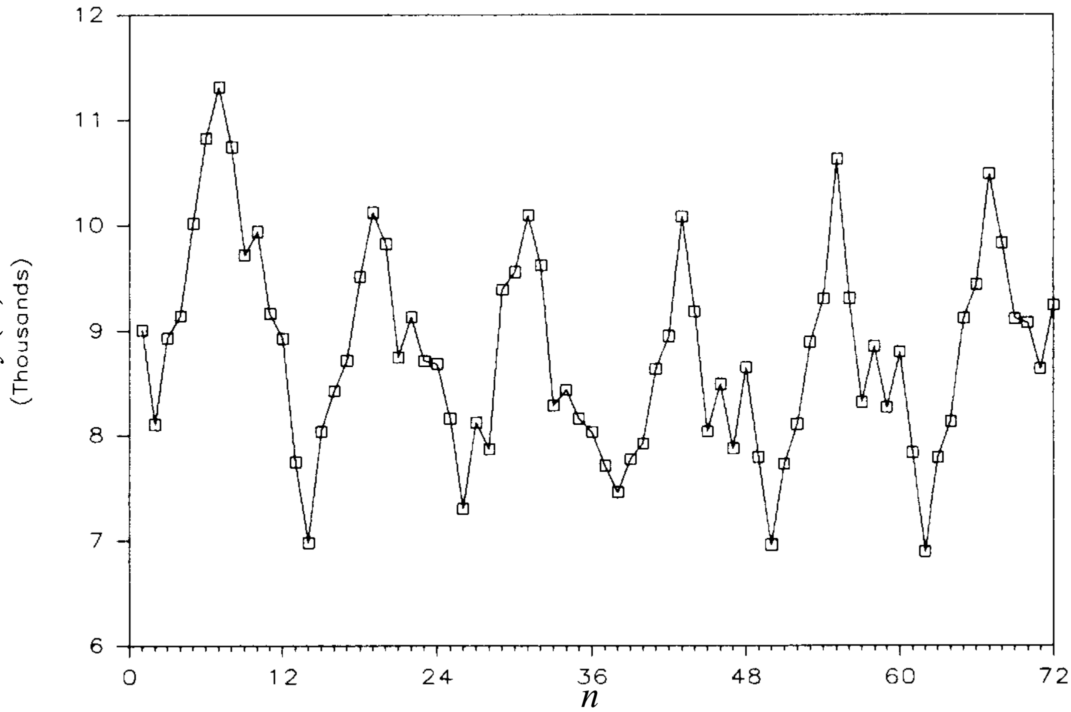
4.2.1. Box-Jenkins method:

- ***Key idea of the method:***
  - The observed sequence $\{y(0), \ldots, y(N'-1)\}$ is transformed in such a way that the transformed sequence $\{x(0), \ldots, x(N-1)\}$ can be reasonably assumed to be the realization of a WSS process $\{X(n)\}$.
  - An ARMA*(p,q)* process is fitted to $\{x(0), \ldots, x(N-1)\}$.
  - The estimated autocorrelation function and power spectrum are identified to the autocorrelation function and the power spectrum of the estimated ARMA*(p,q)* process.

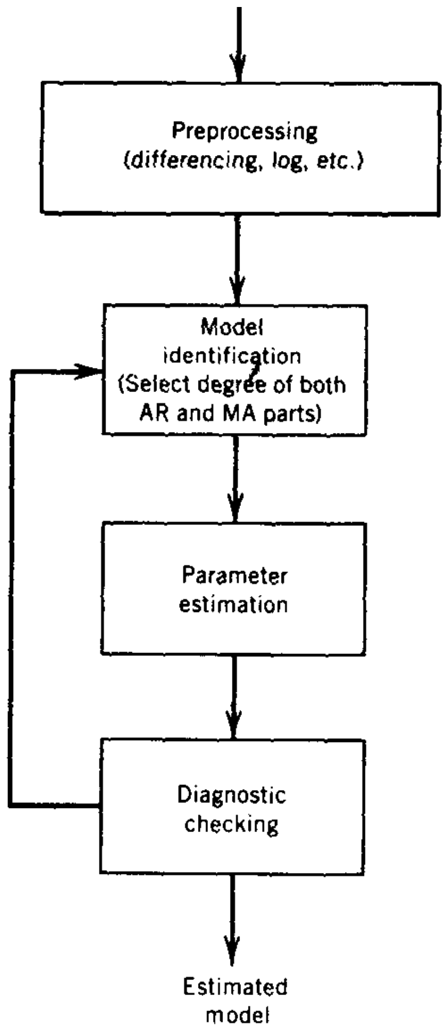
Example 2: International airline passengers.



Example 3: Monthly accidental deaths in the U.S.A.



- **The different steps of the Box-Jenkins method:**



4.2.2. Preprocessing:

- ***Objective:***

  The observed sequence $\{y(0), \ldots, y(N'-1)\}$ is transformed in such a way that the transformed sequence

  $$\{x(0), \ldots, x(N-1)\} = T[\{y(0), \ldots, y(N'-1)\}]$$

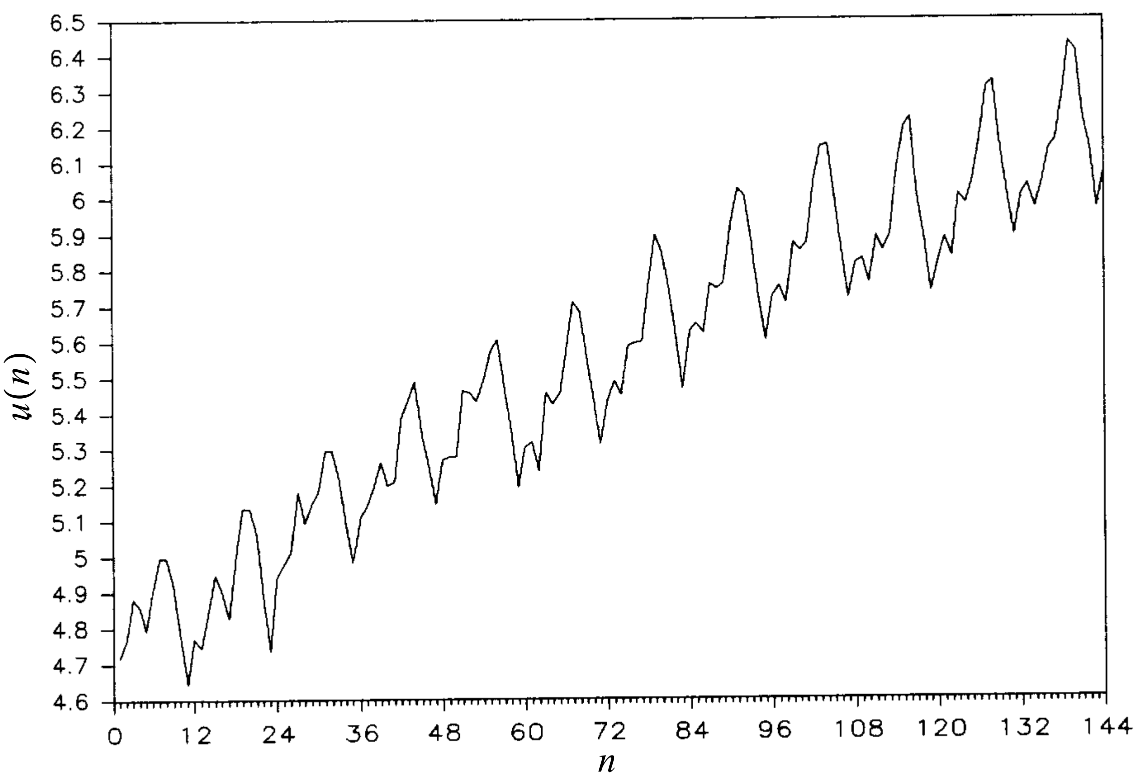
  can be reasonably assumed to be the realization of a WSS process $\{X(n)\}$.

- **Non-linear transformation to create stationarity:**

  Let $\{Y(n)\}$ be a sequence which exhibits some non-stationary features. We can apply a non-linear transformation T to $\{Y(n)\}$ to obtain a new sequence $\{X(n)\} = T[\{Y(n)\}]$ where these features are eliminated or at least reduced.

  Example 2: International airline passengers.
  The variability of the serie increases linearly as a function of the level of the serie. This variability is stabilized by applying the following transformation:

  $$U(n) = \ln(Y(n))$$



To understand how the transformation $Y(n) \to \ln(Y(n))$ stabilizes the variability, let us assume that the standard deviation of $\{Y(n)\}$ increases proportionally to its expectation:

$$\sigma_{Y(n)} = c\mu_{Y(n)}$$

Equivalently,

$$E\left[\left(\frac{Y(n)}{\mu_{Y(n)}} - 1\right)^2\right] = c^2.$$

We can rewrite $U(n) = \ln(Y(n))$ as

$$U(n) = \ln(\mu_{Y(n)}) + \ln\left(\frac{Y(n)}{\mu_{Y(n)}}\right) = \ln(\mu_{Y(n)}) + \ln\left(\frac{Y(n) - \mu_{Y(n)}}{\mu_{Y(n)}} + 1\right)$$

Considering the first order Taylor approximation $\ln(v+1) \approx v$ around 1, $U(n)$ can be approximated according to

$$U(n) \approx \ln(\mu_{Y(n)}) + \left(\frac{Y(n)}{\mu_{Y(n)}} - 1\right)$$

Approximation of the expectation and standard deviation of $U(n)$:

$$\mu_{U(n)} \approx \ln(\mu_{Y(n)})$$

$$\sigma_{U(n)} \approx c$$

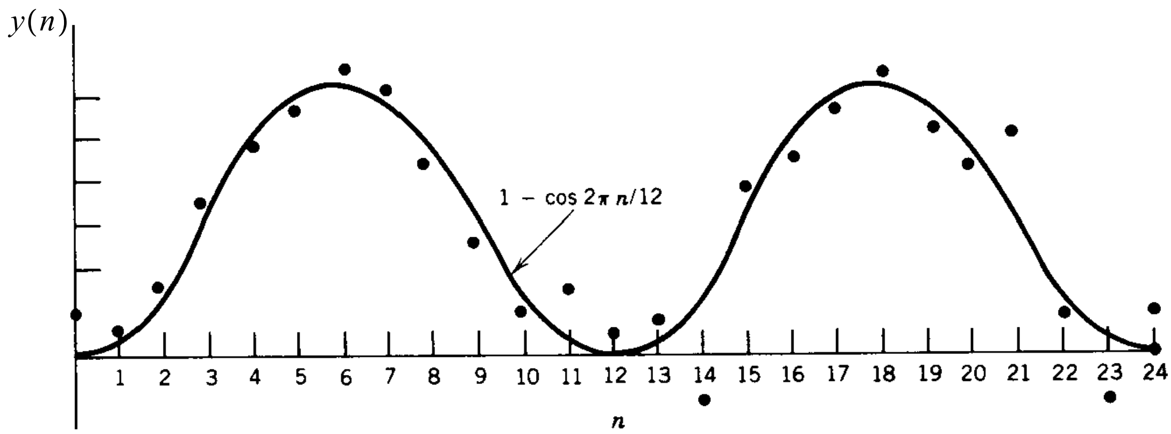
- ***Differentiating to remove periodicity (seasonality):***

Theoretical example 1:

Let consider the sequence $\{Y(n)\}$ where

$$Y(n) = \underbrace{\left[1 - \cos\left(2\pi\frac{n}{12}\right)\right]}_{\substack{\text{Periodic components} \\ \text{of period 12}}} + V(n)$$

where $\{V(n)\}$ is a WSS process.



For example, $\{Y(n)\}$ might represent a monthly average (see Examples 2 to 3). Let

$$\{X(n)\} = \Delta_{12}\{Y(n)\}$$

be the sequence obtained by transforming $\{Y(n)\}$ according to

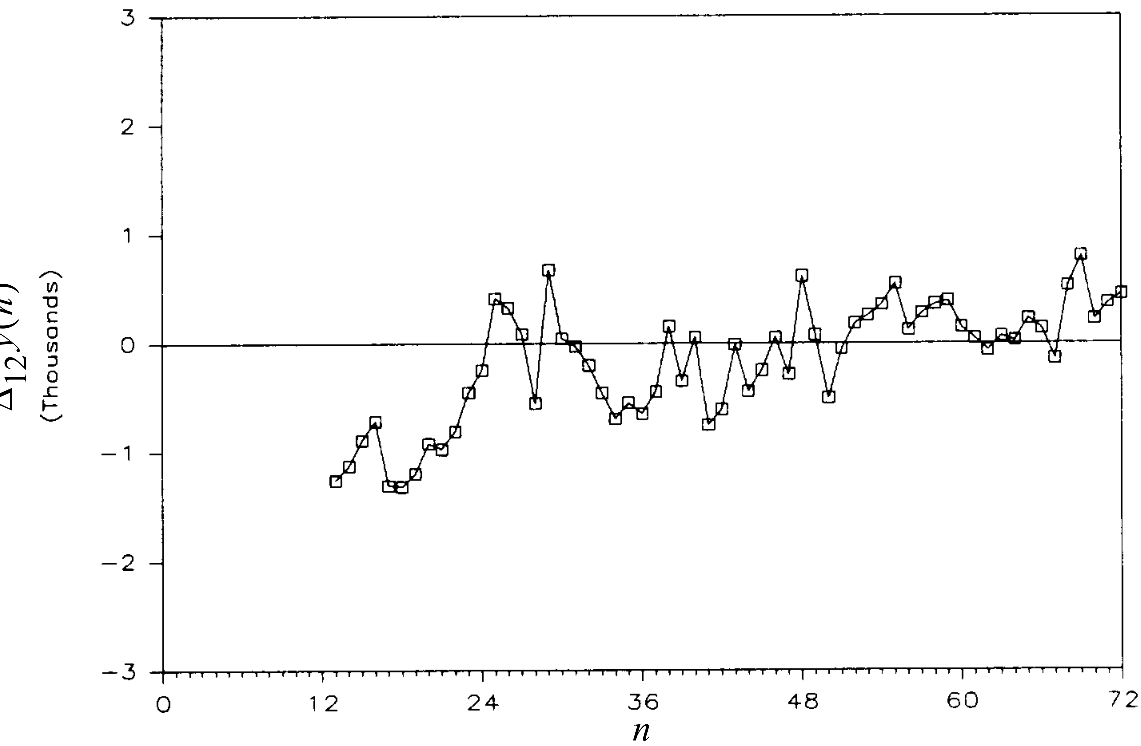
$$X(n) = Y(n) - Y(n-12)$$

Then

$$X(n) = V(n) - V(n-12)$$

Hence, the sequence $\{X(n)\}$ is stationary.

Example 3: Monthly accidental deaths in the U.S.A.



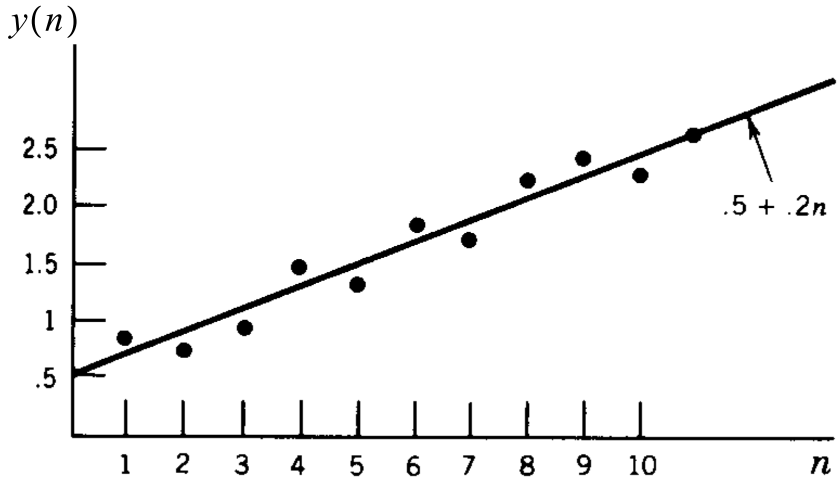
- ***Differentiating to remove trends:***

Theoretical example 2:

Let consider the sequence $\{Y(n)\}$ where

$$Y(n) = \underbrace{\left[\frac{1}{2} + \frac{1}{5}n\right]}_{\text{Trend}} + V(n)$$

where $\{V(n)\}$ is a WSS process.



Let us consider the transformation

$$X(n) = Y(n) - Y(n-1).$$

Then,

$$X(n) = V(n) - V(n-1) + \frac{1}{5} .$$

Hence, $\{X(n)\}$ is a WSS process, which can be modelled as an ARMA process.

- ***ARIMA(p,d,q) processes:***

Notice that the above process $\{X(n)\}$ is the "discrete derivative" of $\{V(n)\}$.

Let us introduce the following notation for discrete derivative:

$$\{X(n)\} = \Delta\{Y(n)\} \text{ if } X(n) = Y(n) - Y(n-1) \text{ for all } n .$$

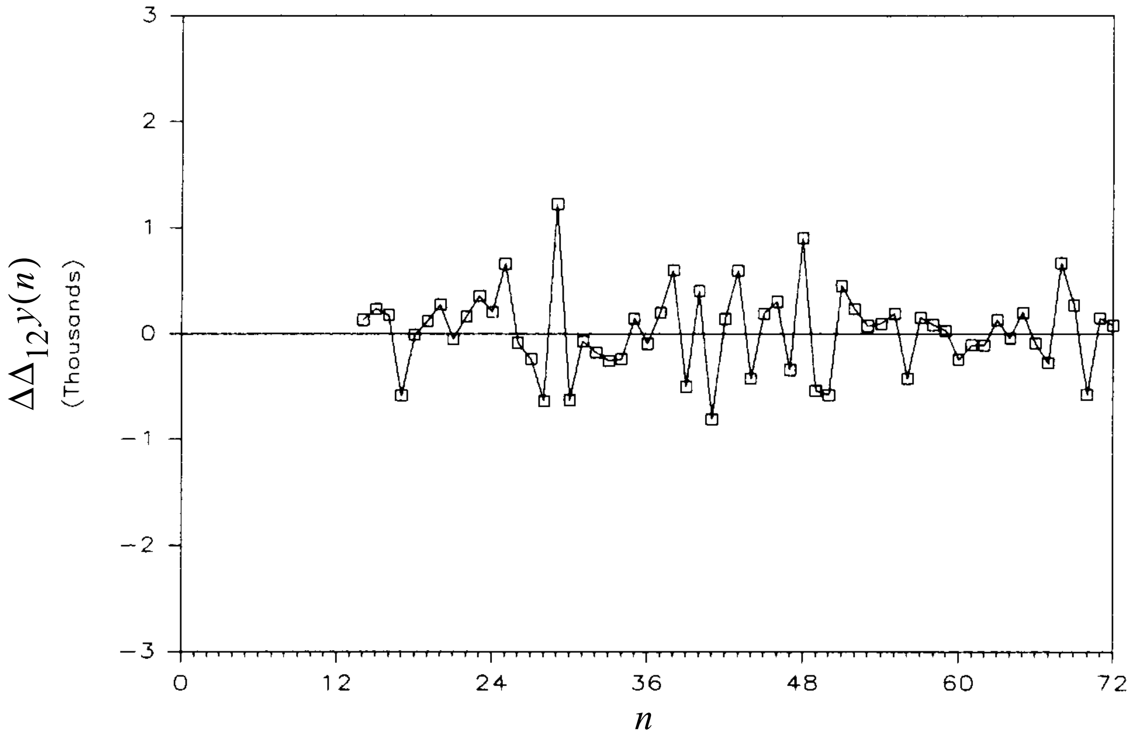
Notice that according to the previously introduced notation

$$\Delta\{Y(n)\} = \Delta_1\{Y(n)\} .$$

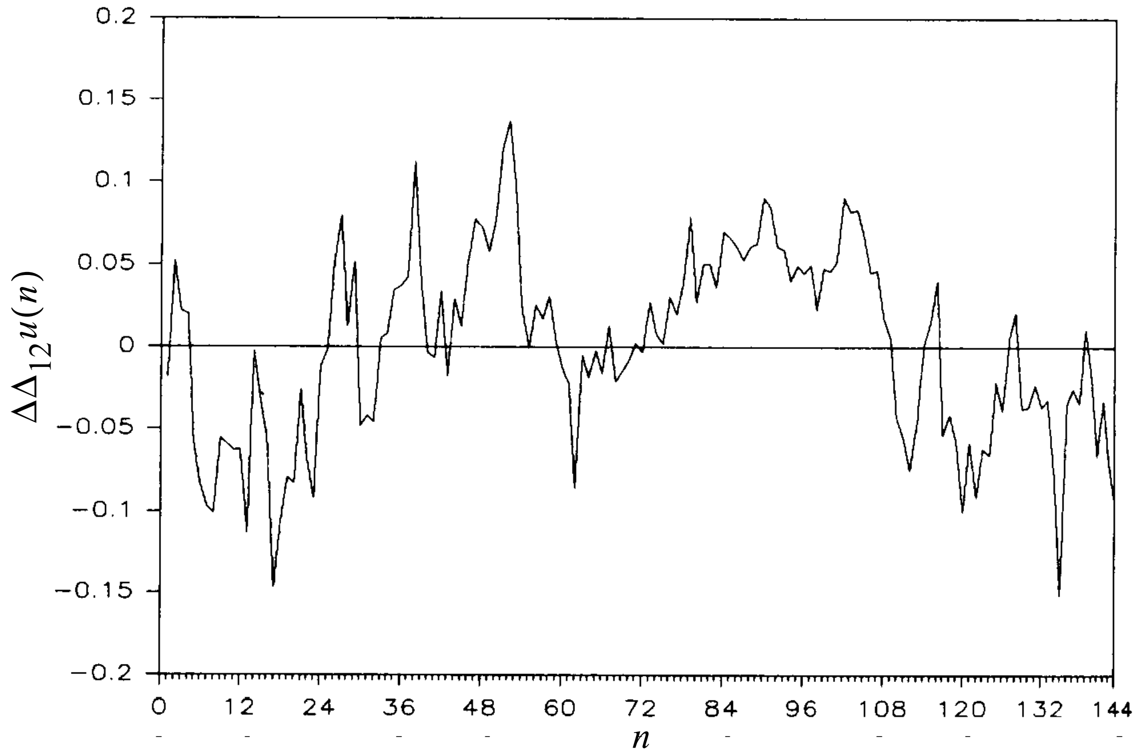
A process $\{Y(n)\}$ is an **ARIMA(*p,d,q*) process** if its *d*th discrete derivative $\{X(n)\} = \Delta^{(d)}\{Y(n)\}$ is an ARMA(*p,q*) process.

An ARIMA process reduces after differentiating finitely many times to an ARMA process. The letter **I** in ARIMA stands for **"integrated"**.
Notice that if $\{X(n)\} = \Delta\{Y(n)\}$ then $\{Y(n)\}$ can be obtained by carrying out a discrete integration of $\{X(n)\}$.

Example 3: Monthly accidental deaths in the U.S.A.



Example 2: International airline passengers.



4.2.3. Fitting ARMA(p,q) processes:

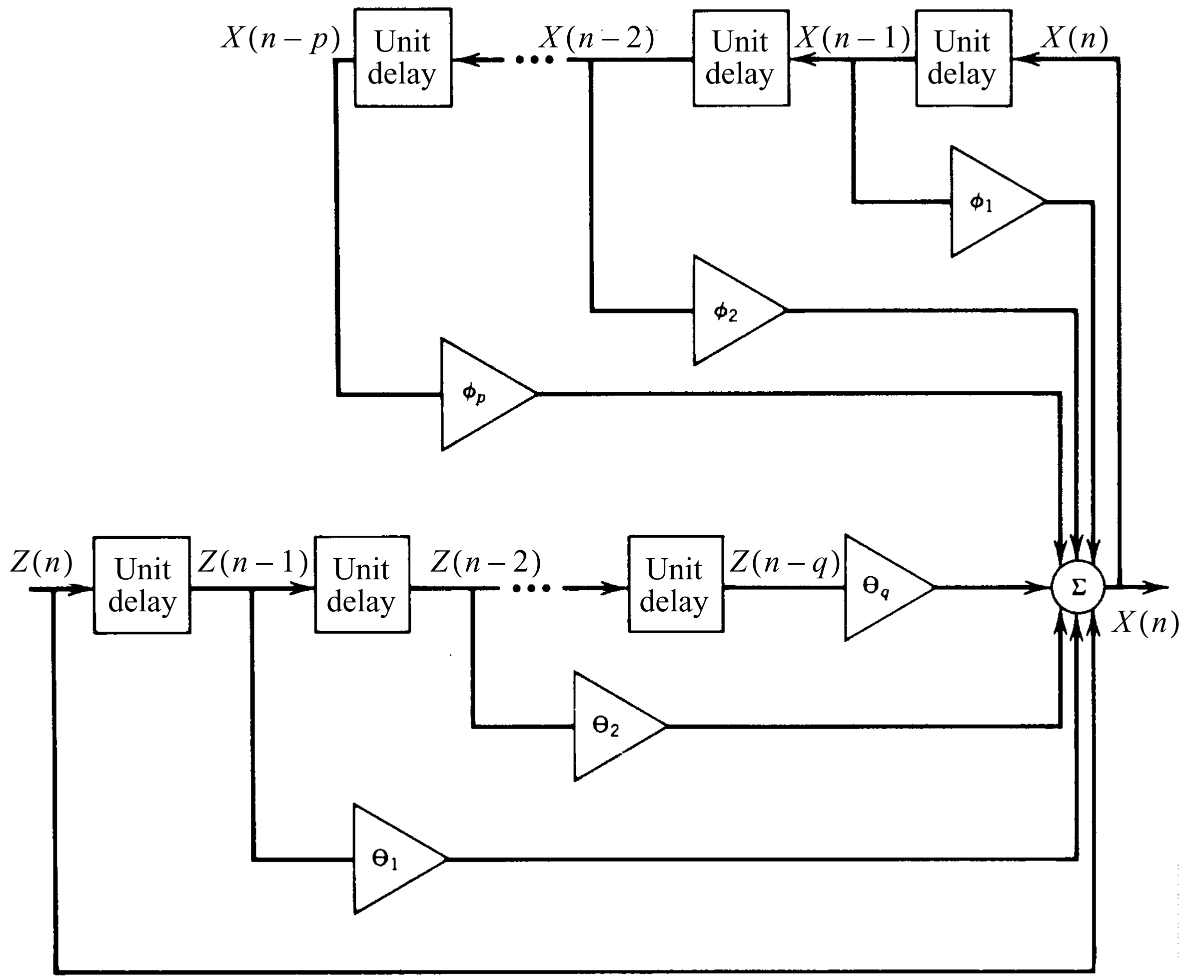
- ***Definition (review):***

  A random sequence $\{X(n)\}$ is an autoregressive moving average process (p, q) th order (ARMA((p, q))) if it is WSS and for any n:

$$X(n) = \sum_{i=1}^{p} \phi_i X(n-i) + \sum_{i=1}^{q} \theta_i Z(n-i) + Z(n)$$

  where $Z(n)$ is a white Gaussian process with variance σ_Z^2.

- ***Filter implementation:***



- ***Parameter estimation:***

  - **Model order** p, q**:**

    p and q are estimated by applying the Akaike information criterion (AIC) or the minimum description length (MDL) criterion.

- Coefficients $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$:

1. The parameters of an AR process can be estimated by solving the Yule-Walker equations:

$$\hat{\gamma} = \hat{\Gamma}\hat{\Phi}$$
$$\hat{R}_{XX}(0) = \hat{\gamma}^T\hat{\Phi} + \hat{\sigma}_Z^{\ 2}$$

where

$$\hat{\Phi} \equiv \begin{bmatrix} \hat{\phi}_1 \\ \ldots \\ \hat{\phi}_p \end{bmatrix} \qquad \hat{\gamma} \equiv \begin{bmatrix} \hat{R}_{XX}(1) \\ \hat{R}_{XX}(2) \\ \ldots \\ \hat{R}_{XX}(p) \end{bmatrix}$$

$$\hat{\Gamma} \equiv \begin{bmatrix} \hat{R}_{XX}(0) & \hat{R}_{XX}(1) & \ldots & \hat{R}_{XX}(p-1) \\ \hat{R}_{XX}(-1) & \hat{R}_{XX}(0) & \ldots & \hat{R}_{XX}(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ \hat{R}_{XX}(-(p-1)) & \hat{R}_{XX}(-(p-2)) & \ldots & \hat{R}_{XX}(0) \end{bmatrix}$$

Example 1: Wölfer sunspot numbers

The estimated AR model for the mean-corrected data is found to be

a) $\hat{p} = 3$,

b) $X(n) - \hat{\phi}_1 X(n-1) + \hat{\phi}_2 X(n-2) - \hat{\phi}_3 X(n-3) = Z(n)$

2. In the general case of an ARMA process, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ can be estimated by using the **maximum likelihood method**.

Example 1: Wölfer sunspot numbers

The estimated ARMA model for the mean-corrected data is found to be

a) $\hat{p} = 9, \hat{q} = 1$,

b) $X(n) - 1.475X(n-1) + 0.937X(n-2)$
$$-0.218X(n-3) + 0.134X(n-9) = Z(n)$$

- ***Estimate of the power spectrum:***

  - Estimate of the transfer function:

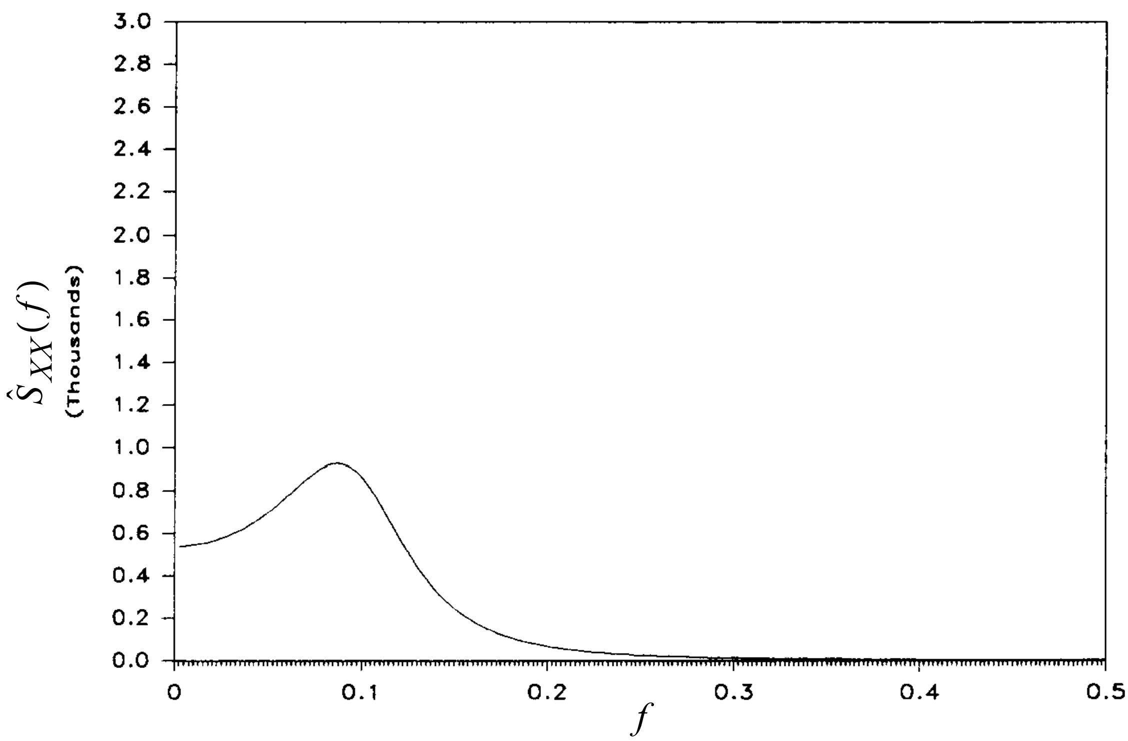
$$\hat{H}(f) = \frac{1 + \sum_{i=1}^{\hat{q}} \hat{\theta}_i \exp(-j2\pi i f)}{1 - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \exp(-j2\pi i f)}$$

  - Estimate of the power spectrum:

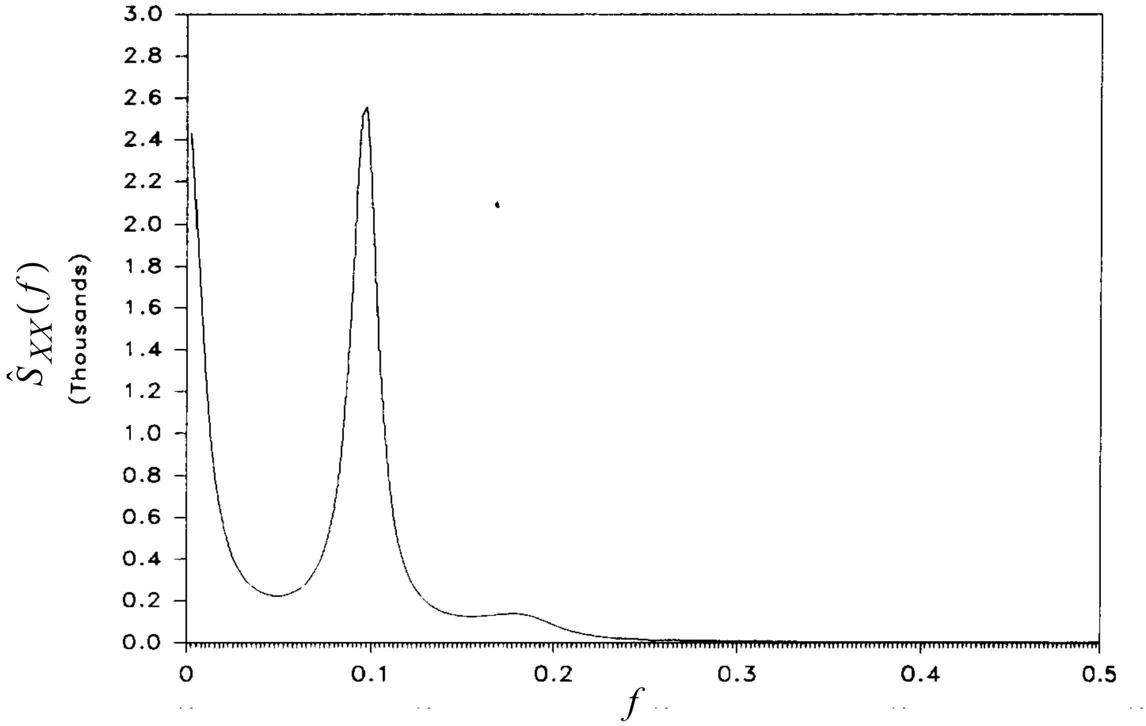
$$\hat{S}_{XX}(f) = \frac{\left|1 + \sum_{i=1}^{\hat{q}} \hat{\theta}_i \exp(-j2\pi i f)\right|^2}{\left|1 - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \exp(-j2\pi i f)\right|^2} \hat{\sigma}_Z^2$$

Example 1: Wölfer sunspot numbers:

- Estimate with the AR(3) model:



- Estimate with the ARMA(9,1) model:



- ***Estimate of the autocorrelation function:***

$$\hat{R}_{XX}(k) = F^{-1}\{\hat{S}_{XX}(f)\}$$