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

6.1. Model-free estimation of random processes

In this section \( \{X(n)\} \) is a WSS process with
- mean value: \( \mu_X = \mathbb{E}[X(n)] \)
- autocorrelation function: \( R_{XX}(k) = \mathbb{E}[X(n)X(n+k)] \)

The autocovariance function of \( \{X(n)\} \) is
\[
C_{XX}(k) = \mathbb{E}[(X(n) - \mu_X)(X(n+k) - \mu_X)] = R_{XX}(k) - \mu_X^2
\]

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

**Example 1: Wöller sunspot numbers**

![Graph of sunspot numbers](image)

Defining the window function
\[
g(n) = \begin{cases} 
1 & ; \quad n \in \{0, ..., N-1\} \\
0 & ; \quad \text{otherwise}
\end{cases}
\]

the observed sequence reads:
\[
X_{\text{obs}}(n) = g(n)X(n)
\]

6.1.1. Estimation of the mean-value

- **Arithmetic mean:**
  \[
  \hat{\mu}_X = \bar{X} = \frac{1}{N} \sum_{n=0}^{N-1} X(n)
  \]

- **Mean and variance of \( \bar{X} \):**
  - Mean: \( \bar{X} \) is an unbiased estimator of \( \mu_X \):
    \[
    \hat{\mu}_X = \mu_X
    \]
  - Variance:
    \[
    \sigma^2_{\bar{X}} = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} \left[1 - \frac{|k|}{N}\right] C_{XX}(k)
    \]

Special case: When \( \{X(n)\} \) is an uncorrelated process:
\[
\sigma^2_{\bar{X}} = \frac{1}{N} C_{XX}(0) = \frac{1}{N} \sigma^2_X
\]

**Proof:** See Exercise 9.1.
To show that the sample autocorrelation function \( \hat{R}_{XX}(k) \) is biased we recast it as:

\[
\hat{R}_{XX}(k) = \frac{1}{N} \sum_{n=0}^{\infty} X_{\text{obs}}(n)X_{\text{obs}}(n+k)
\]

\[
= \frac{1}{N} \sum_{n=\infty}^{\infty} g(n)g(n+k)X(n)X(n+k)
\]

Taking the expectation on both sides yields

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

The function

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

is called the Bartlett window.

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

\[
E[\hat{R}_{XX}(k)] = w_B(k) \hat{R}_{XX}(k)
\]
• **Biased sample autocovariance:**

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

**Example 1: Wolfer sunspot numbers**

![Graph of biased sample autocovariance](image)

• **Unbiased sample autocorrelation function:**

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

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

\[
\mathbb{E}[ \hat{R}_{XX}(k) ] = w_r(k) \hat{R}_{XX}(k)
\]

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

![Graph of rectangular function](image)

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

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

- 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 \(\mathbb{E}[X(n)X(n+m)X(k)X(k+m)]\).

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

**6.1.3. Estimation of the power spectral density:**

- **Continuous-frequency periodogram:**
  Let us start from the slightly differently reformulated Fourier transform:
  \[
  X(f) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi nf) \quad f \in [0, 1)
  \]

The periodogram of $X_{\text{obs}}(n)$ is defined to be
\[
\hat{S}_{XX}(f) = \mathcal{F}\{\hat{R}_{XX}(k)\}
= \frac{1}{N} \left| \sum_{n=0}^{N-1} X(n) \exp(-j2\pi nf) \right|^2 = \frac{1}{N} \left| \mathcal{F}\{X_{\text{obs}}(n)\}(f) \right|^2 \quad f \in [0, 1)
\]

**Proof:**

- **Discrete-frequency periodogram:**
  \[
  \hat{S}_{XX}(m) = \hat{S}_{XX}(f) \bigg|_{f = m/N} \quad m = 0, \ldots, N-1
  \]

<table>
<thead>
<tr>
<th>$f$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Discrete Fourier transform:**
  The discrete Fourier transform and the inverse DFT are defined according to
  \[
  X_d(m) = \mathcal{F}_d(x(n)) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \exp\left(-j2\pi \frac{nm}{N}\right)
  \]
  \[
  x(n) = \mathcal{F}_d^{-1}(X_d(m)) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} X_d(m) \exp\left(j2\pi \frac{nm}{N}\right)
  \]

Relation between the discrete Fourier transform and the (continuous-frequency) Fourier transform:
\[
X_d(m) = \frac{1}{\sqrt{N}} X(f) \bigg|_{f = m/N} \quad m = 0, \ldots, N-1
\]

In particular, the discrete-frequency periodogram can be computed as
\[
\hat{S}_{XX}(m) = \left| \mathcal{F}_d\{X_{\text{obs}}(n)\}(m) \right|^2
\]
Example 1: Wölf er sunspot numbers

- **Bias of the periodogram:**
  Because the Fourier transform is a linear operation, we have
  \[ \mathbb{E}[\hat{S}_{XX}(f)] = \mathcal{F}(\mathbb{E}[\hat{R}_{XX}(k)]) \]
  It follows from (6.2) that:
  \[ \mathbb{E}[\hat{S}_{XX}(f)] = \mathcal{F} \{ w_B(k) R_{XX}(k) \} \]
  \[ = W_p(f) \ast S_{XX}(f) \]
  The Fourier transform
  \[ W_p(f) = \mathcal{F} \{ w_B(k) \} = \frac{1}{N} \left( \frac{\sin(\pi k)}{\sin(\pi f)} \right)^2 \]
  of the Bartlett window is called the Féjer kernel.
  
  **Proof:** It can be easily shown that the Fourier spectrum of \( R_{gg}(k) \) is
  \[ |G(f)|^2 = \left( \frac{\sin(\pi N)}{\sin(\pi f)} \right)^2 \]
  where \( G(f) = \mathcal{F} \{ g(n) \} \).

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

\[ E[\hat{S}_{XX}(f)] = W_p(f) \ast S_{XX}(f) \]
\[ E[\hat{S}_{XX}(m)] = [W_p(f) \ast S_{XX}(f)] \big|_{f = m/N} \]

- **Spectral leakage:**

\[ W_p(f) \]

\[ 0 \quad 1/N \quad 2/N \quad f \]

\[ f \]

\[ 0 \quad 1/2 \quad f \]

\[ f \]

\[ f \]

\[ 0 \quad 1/2 \quad f \]
As \( N \) increases to infinity, \( W_P(f) \to \delta(f) \), so that
\[
E[\hat{S}_{XX}(f)] = S_{XX}(f),
\]
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,
  \[
  \sigma^2_{\hat{S}_{XX}(f)} \to \begin{cases} 2S_{XX}(f)^2 & ; f = 0, 1/2 \\ S_{XX}(f)^2 & ; \text{otherwise} \end{cases}
  \]
  \[
  \sum \hat{S}_{XX}(f_1)\hat{S}_{XX}(f_2) \to 0 \quad \text{for any} \quad f_1, f_2 \in \left[0, \frac{1}{2}\right], f_1 \neq f_2
  \]

  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 \( f \neq 0, 1/2 \).

  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.

  **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: Wolfer sunspot numbers

Some well-known lag windows:
6.2. Parametric (model-based) estimation of random processes

6.2.1. 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:
  1. Preprocessing (differencing, log, etc.)
  2. Model identification (Select degree of both AR and MA parts)
  3. Parameter estimation
  4. Diagnostic checking
  5. Estimated model
6.2.2. 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 series increases linearly as a function of the level of the series. 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 \( U(n) = \ln(T(n)) \) as

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

Considering the first order Taylor approximation \( \ln(v+1) = v \) 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 \( U(n) \):

\[
\mu_{U(n)} = \ln(\mu_{Y(n)})
\]

\[
\sigma_{U(n)} = c
\]

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

  **Theoretical example 1:**
  Let consider the sequence \( \{Y(n)\} \) where

  \[
  Y(n) = \left[ 1 - \cos\left( \frac{2\pi n}{12} \right) \right] + V(n)
  \]

  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.**

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:
  \[
  \{X(n)\} = \Delta\{Y(n)\} \quad \text{if} \quad X(n) = Y(n) - Y(n - 1) \quad \text{for all} \quad n.
  \]
  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 \( \{X(n)\} \).
6.2.3. 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 \(\sigma_Z^2\).

• **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:

\[
\hat{\Phi} = \begin{bmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{bmatrix}, \quad \hat{\gamma} = \begin{bmatrix} \hat{R}_{XX}(1) \\ \vdots \\ \hat{R}_{XX}(p) \end{bmatrix},
\]

where

\[
\hat{\Gamma} = \begin{bmatrix} \hat{R}_{XX}(0) & \hat{R}_{XX}(1) & \cdots & \hat{R}_{XX}(p-1) \\ \hat{R}_{XX}(-1) & \hat{R}_{XX}(0) & \cdots & \hat{R}_{XX}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{XX}(-(p-1)) & \hat{R}_{XX}(-(p-2)) & \cdots & \hat{R}_{XX}(0) \end{bmatrix}
\]

**Example 1: Wölf er sunspot numbers**

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

- \( p = 3 \),
  - \( X(n) - \hat{\phi}_1 X(n-1) - \hat{\phi}_2 X(n-2) - \hat{\phi}_3 X(n-3) = Z(n) \)

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: Wölf er sunspot numbers**

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

- \( p = 9, q = 1 \),
  - \( X(n) - 1.475 X(n-1) + 0.937 X(n-2) - 0.218 X(n-3) + 0.134 X(n-9) = Z(n) \)

- Estimate of the power spectrum:
  - Estimate of the transfer function:
    - Estimate of the power spectrum:
      - Estimate with the AR(3) model:

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

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

**Example 1: Wölf er sunspot numbers**

- Estimate with the AR(3) model:
- Estimate with the ARMA(9,1) model:

\[ \hat{S}_{XX}(f) = \gamma^{-1}(\hat{\xi}_{XX}(f)) \]