Stochastic Processes 2 - Preliminary Lecture Notes
Zuzana Prášková
Department of Probability and Mathematical Statistics
Charles University in Prague

Literature

- **Basic study text:** Z. Prášková: Základy náhodných procesů II, Karolinum 2006 (in Czech)

- **Supplementary texts:**
  
  
1 Definitions and basic characteristics

1.1 Definition of a stochastic process

Definition 1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(S, \mathcal{E})$ a measurable space, and $T \subset \mathbb{R}$. A family of random variables $\{X_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $S$ is called a stochastic (random) process. If $S = \mathbb{R}$, $\{X_t, t \in T\}$ is called a real stochastic process.

If $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ or $T \subset \mathbb{Z}$, $\{X_t, t \in T\}$ is called discrete time stochastic process, time series. If $T = [a, b]$, $-\infty \leq a < b \leq \infty$, $\{X_t, t \in T\}$ is continuous time process.

For any $\omega \in \Omega$ fixed, $X_t(\omega)$ is a function on $T$ with values in $S$ which is called a trajectory of the process.

Definition 2. A pair $(S, \mathcal{E})$, where $S$ is a set of values of random variables $X_t$ and $\mathcal{E}$ is a $\sigma$-algebra of subsets of $S$, is called the state space of the process $\{X_t, t \in T\}$.

Definition 3. A real stochastic process $\{X_t, t \in T\}$ is said to be measurable, if the mapping $(\omega, t) \rightarrow X_t(\omega)$ is $\mathcal{A} \otimes \mathcal{B}_T$-measurable, where $\mathcal{B}_T$ is a $\sigma$-algebra of Borel subsets of $T$ and $\mathcal{A} \otimes \mathcal{B}_T$ is a product $\sigma$-algebra.

Finite-dimensional distributions:

For all $n \in \mathbb{N}_0$ and any finite subset $\{t_1, \ldots, t_n\} \subset T$ there is a system of random variables $X_{t_1}, \ldots, X_{t_n}$, with the joint distribution function

$$P [X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n] = F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)$$

for all $x_1, \ldots, x_n$.

A system of distribution functions is said to be consistent, if

- $F_{t_1, \ldots, t_n}(x_{i_1}, \ldots, x_{i_n}) = F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)$ for any permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ (symmetry)

- $\lim_{x_n \to \infty} F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = F_{t_1, \ldots, t_{n-1}}(x_1, \ldots, x_{n-1})$ (consistency)

The characteristic function of a random vector $X = (X_1, \ldots, X_n)$ is

$$\varphi_X(u) := Ee^{iu^\top X} = Ee^{\sum_{j=1}^n u_j X_j}, \quad u = (u_1, \ldots, u_n).$$
A system of characteristic functions is said to be consistent if

- \( \varphi(u_{i_1}, \ldots, u_{i_n}) = \varphi(u_1, \ldots, u_n) \) for any permutation \((i_1, \ldots, i_n)\) of \((1, \ldots, n)\), (symmetry)

- \( \lim_{u_n \to 0} \varphi_{X_{t_1}, \ldots, X_{t_n}}(u_1, \ldots, u_n) = \varphi_{X_{t_1}, \ldots, X_{t_{n-1}}}(u_1, \ldots, u_{n-1}) \) (consistency)

### 1.2 Daniell-Kolmogorov theorem

For any stochastic process there exists a consistent system of distribution functions.

**Theorem 1.** Let \( \{F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)\} \) be a consistent system of distribution functions. Then there exists a stochastic process \( \{X_t, t \in T\} \) such that for any \( n \in \mathbb{N} \), any \( t_1, \ldots, t_n \in T \) and any real \( x_1, \ldots, x_n \) it holds

\[
P[X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n] = F_{t_1, \ldots, t_n}(x_1, \ldots, x_n).
\]

**Proof.** Štěpán (1987), Theorem I.10.3.

### 1.3 Autocovariance and autocorrelation function

**Definition 4.** A complex-valued random variable \( X \) is defined by \( X = Y + iZ \), where \( Y \) and \( Z \) are real random variables, \( i = \sqrt{-1} \).

The mean value of a complex-valued random variable \( X = Y + iZ \) is defined by \( \mathbb{E}X = \mathbb{E}Y + i\mathbb{E}Z \) provided the mean values \( \mathbb{E}Y \) and \( \mathbb{E}Z \) exist.

The variance of a complex-valued random variable \( X = Y + iZ \) is defined by \( \text{var}X := \mathbb{E}[|X - \mathbb{E}X|^2] = \mathbb{E}|X - \mathbb{E}X|^2 \geq 0 \) provided the second moments of random variables \( Y \) and \( Z \) exist.

**Definition 5.** A complex-valued stochastic process is a family of complex-valued random variables on \( (\Omega, \mathcal{A}, \mathbb{P}) \).

**Definition 6.** Let \( \{X_t, t \in T\} \) be a stochastic process such that \( \mathbb{E}X_t \) exists for all \( t \in T \). Then the function \( \{\mu_t, t \in T\} \) where \( \mu_t = \mathbb{E}X_t \) defined on \( T \) is called the mean value of the process \( \{X_t, t \in T\} \). We say that the process is centered if its mean value is zero for all \( t \in T \).
Definition 7. Let \( \{X_t, t \in T\} \) be a process with finite second moments, i.e., \( \mathbb{E}|X_t|^2 < \infty, \forall t \in T \). Then a (complex-valued) function defined on \( T \times T \) by
\[
R(s, t) = \mathbb{E} [(X_s - \mu_s)(X_t - \mu_t)]
\]
is called the autocovariance function of the process \( \{X_t, t \in T\} \). The value \( R(t, t) \) is the variance of the process at time \( t \).

Definition 8. Autocorrelation function of the process \( \{X_t, t \in T\} \) with positive variances is defined by
\[
r(s, t) = \frac{R(s, t)}{\sqrt{R(s, s)R(t, t)}}, \quad s, t \in T.
\]

Definition 9. Stochastic process \( \{X_t, t \in T\} \) is called Gaussian, if for any \( n \in \mathbb{N} \) and \( t_1, \ldots, t_n \in T \), the vector \( (X_{t_1}, \ldots, X_{t_n})^\top \) is normally distributed \( \mathcal{N}_n(m_t, V_t) \), where
\[
m_t = (\mathbb{E}X_{t_1}, \ldots, \mathbb{E}X_{t_n})^\top
\]
and
\[
V_t = \begin{pmatrix}
\text{var}X_{t_1} & \text{cov}(X_{t_1}, X_{t_2}) & \ldots & \text{cov}(X_{t_1}, X_{t_n}) \\
\text{cov}(X_{t_2}, X_{t_1}) & \text{var}X_{t_2} & \ldots & \text{cov}(X_{t_2}, X_{t_n}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(X_{t_n}, X_{t_1}) & \text{cov}(X_{t_n}, X_{t_2}) & \ldots & \text{var}X_{t_n}
\end{pmatrix}.
\]

1.4 Strict and weak stationarity

Definition 10. Stochastic process \( \{X_t, t \in T\} \) is said to be strictly stationary, if for any \( n \in \mathbb{N} \), for any \( x_1, \ldots, x_n \) real and for any \( t_1, \ldots, t_n \) a \( h \) such that \( t_k \in T, t_k + h \in T, 1 \leq k \leq n \),
\[
F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = F_{t_1+h, \ldots, t_n+h}(x_1, \ldots, x_n).
\]

Definition 11. Stochastic process \( \{X_t, t \in T\} \) with finite second moments is said to be weakly stationary or second order stationary, if its mean value is constant, \( \mu_t = \mu, \forall t \in T \) and if its autocovariance function \( R(s, t) \) is a function of \( s - t \), only. If only the latter condition is satisfied, the process is called covariance stationary.

Autocovariance function of a weakly stationary process is a function of one variable:
\[
R(t) := R(t, 0), \quad t \in T,
\]

Autocorrelation function in such case:
\[
r(t) = \frac{R(t)}{R(0)}.
\]
Theorem 2. Strictly stationary stochastic process \( \{X_t, t \in T\} \) with finite second moments is also weakly stationary.

Proof. \( \{X_t, t \in T\} \) strictly stationary with finite second moments \( \Rightarrow X_t \) are equally distributed for all \( t \in T \) and thus with the mean value
\[
EX_t = EX_{t+h}, \forall t \in T, \forall h : t + h \in T
\]
especially, for \( h = -t \) : \( EX_t = EX_0 = \text{const} \)

Similarly, \( (X_t, X_s) \) are equally distributed and
\[
E[X_t X_s] = E[X_{t+h} X_{s+h}] \forall s, t \in T, \forall h : s + h \in T, t + h \in T
\]
especially, for \( h = -t \) : \( E[X_t X_s] = E[X_0 X_{s-t}] \) is a function of \( s - t \).

Example 1. Let \( \{X_t, t \in T\} \) be a sequence of iid random variables with a distribution function \( F \)
\[
F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = P[X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n] = \prod_{i=1}^{n} P[X_{t_i} \leq x_i] = \prod_{i=1}^{n} F(x_i),
\]
\[
F_{t_1+h, \ldots, t_n+h}(x_1, \ldots, x_n) = P[X_{t_1+h} \leq x_1, \ldots, X_{t_n+h} \leq x_n] = \prod_{i=1}^{n} P[X_{t_i+h} \leq x_i] = \prod_{i=1}^{n} F(x_i),
\]
\( \Rightarrow \{X_t, t \in T\} \) is strictly stationary.

Example 2. \( \{X_t, t \in \mathbb{Z}\} \) - a sequence defined by \( X_t = (-1)^t X \), where \( X \) is a random variable:
\[
X = \begin{cases} 
-\frac{1}{2} & \text{with probability } \frac{3}{4}, \\
\frac{3}{4} & \text{with probability } \frac{1}{4}.
\end{cases}
\]
Then \( \{X_t, t \in \mathbb{Z}\} \) is weakly stationary, since
\[
EX_t = 0,
\]
\[
\text{var } X_t = \sigma^2 = \frac{3}{16},
\]
\[
R(s,t) = \sigma^2(-1)^{s+t} = \sigma^2(-1)^{s-t},
\]
but it is not strictly stationary (variables \( X \) and \( -X \) are not equally distributed).
Theorem 3. A weakly stationary Gaussian process \( \{X_t, t \in T\} \) is also strictly stationary.

Proof. Weak stationarity of the process \( \{X_t, t \in T\} \) implies \( \mathbf{E}X_t = \mu, \ \text{cov}(X_t, X_s) = R(t - s) = \text{cov}(X_{t+h}, X_{s+h}), t, s \in T \), thus

\[
\mathbf{E}(X_{t_1}, \ldots, X_{t_n}) = \mathbf{E}(X_{t_1+h}, \ldots, X_{t_n+h}) = (\mu, \ldots, \mu) := \mu
\]

\[
\text{var}(X_{t_1}, \ldots, X_{t_n}) = \text{var}(X_{t_1+h}, \ldots, X_{t_n+h}) := \Sigma
\]

\[
\Sigma = \begin{pmatrix}
R(0) & R(t_2 - t_1) & \cdots & R(t_n - t_1) \\
R(t_2 - t_1) & R(0) & \cdots & R(t_n - t_2) \\
\vdots & \vdots & \ddots & \vdots \\
& & & R(0)
\end{pmatrix}
\]

Since the normal distribution is uniquely defined by the mean value vector and the variance matrix, \( (X_{t_1}, \ldots, X_{t_n}) \sim \mathcal{N}(\mu, \Sigma) \), and \( (X_{t_1+h}, \ldots, X_{t_n+h}) \sim \mathcal{N}(\mu, \Sigma) \Rightarrow \{X_t, t \in T\} \) is strictly stationary. \( \square \)

1.5 Properties of autocovariance function

Theorem 4. Let \( \{X_t, t \in T\} \) be a process with finite second moments. Then its autocovariance function satisfies

\[
R(t, t) \geq 0,
\]

\[
|R(s, t)| \leq \sqrt{R(s, s)} \sqrt{R(t, t)}.
\]

Proof. The first assertion follows from the definition of the variance. The second one follows from the Schwarz inequality, since

\[
|R(s, t)| = |\mathbb{E}(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)| \leq (\mathbb{E}|X_s - \mathbb{E}X_s|^2)^{\frac{1}{2}} (\mathbb{E}|X_t - \mathbb{E}X_t|^2)^{\frac{1}{2}}
\]

\[
= \sqrt{R(s, s)} \sqrt{R(t, t)}.
\]

Thus, for weakly stationary process \( R(0) \geq 0 \) a \( |R(t)| \leq R(0) \). \( \square \)
**Definition 12.** Let \( f(s, t) \) be a complex-valued function defined on \( T \times T, T \subset \mathbb{R} \). We say that \( f \) is positive semidefinite, if \( \forall n \in \mathbb{N}, \) any complex numbers \( c_1, \ldots, c_n \) and any \( t_1, \ldots, t_n \in T \) it holds
\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} f(t_j, t_k) \geq 0.
\]

We say that a complex-valued function \( g \) on \( T \) is positive semidefinite, if \( \forall n \in \mathbb{N}, \) any complex numbers \( c_1, \ldots, c_n \) and any \( t_1, \ldots, t_n \in T \), such that \( t_j - t_k \in T \), it holds
\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} g(t_j - t_k) \geq 0.
\]

**Definition 13.** We say that a complex-valued function \( f \) on \( T \times T \) is Hermitian, if \( f(s, t) = \overline{f(t, s)} \) for all \( s, t \in T \). A complex-valued function \( g \) of one variable is called Hermitian, if \( g(-t) = \overline{g(t)} \) for all \( t \in T \).

**Theorem 5.** Any positive semidefinite function is also Hermitian.

**Proof.** Use the definition of positive semidefiniteness and for \( n = 1 \) choose \( c_1 = 1 \); for \( n = 2 \) choose \( c_1 = 1, c_2 = 1 \) and \( c_1 = 1, c_2 = i (= \sqrt{-1}) \).

**Remark 1.** A positive semidefinite real-valued function \( f \) on \( T \times T \), is symmetric, i.e., \( f(s, t) = f(t, s) \) for all \( s, t \in T \). A positive semidefinite real-valued function \( g \) on \( T \) is symmetric, i.e, \( g(t) = g(-t) \) for all \( t \in T \).

**Theorem 6.** Let \( \{X_t, t \in T\} \) be a process with finite second moments. Then its autocovariance function is positive semidefinite on \( T \times T \).

**Proof.** Suppose wlog that the process is centered. Then for any \( n \in \mathbb{N} \), complex constants \( c_1, \ldots, c_n \) and \( t_1, \ldots, t_n \in T \)
\[
0 \leq \mathbb{E} \left[ \sum_{j=1}^{n} c_j X_{t_j} \right]^2 = \mathbb{E} \left[ \sum_{j=1}^{n} c_j X_{t_j} \sum_{k=1}^{n} c_k X_{t_k} \right] = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} \mathbb{E}(X_{t_j} X_{t_k}) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} R(t_j, t_k).
\]
Theorem 7. To any positive semidefinite function \( R \) on \( T \times T \) there exists a stochastic process \( \{X_t, t \in T\} \) with finite second moments such that its autocovariance function is \( R \).

Proof. The proof will be given for real-valued function \( R \), only. For the proof with complex-valued \( R \) see, e.g., Loève (1955), Chap. X, Par. 34.

Since \( R \) is positive semidefinite, then for any \( n \in \mathbb{N} \) and any real \( t_1, \ldots, t_n \in T \) the matrix

\[
V_t = \begin{pmatrix}
R(t_1, t_1) & R(t_1, t_2) & \ldots & R(t_1, t_n) \\
R(t_2, t_1) & R(t_2, t_2) & \ldots & R(t_2, t_n) \\
& \ddots & \ddots & \ddots \\
R(t_n, t_1) & R(t_n, t_2) & \ldots & R(t_n, t_n)
\end{pmatrix}
\]

is positive semidefinite. Function

\[
\varphi(u) = \exp \left\{-\frac{1}{2} u^\top V_t u \right\}, \quad u \in \mathbb{R}^n
\]

is the characteristic function of the normal distribution \( \mathcal{N}_n(0, V_t) \). In this way, \( \forall n \in \mathbb{N} \) and any real \( t_1, \ldots, t_n \in T \) we get a consistent system of characteristic functions. The corresponding system of the distribution functions is also consistent. Thus according to Daniell-Kolmogorov theorem (Theorem 1) there exists a Gaussian stochastic process, covariances of which are given the values of the function \( R(s, t) \); hence, function \( R \) is the autocovariance function of this process. \( \square \)

Example 3. Decide whether function \( \cos t, t \in T = (-\infty, \infty) \) is an autocovariance function of a stochastic process.

Solution:
It suffices to show that \( \cos t \) is a positive semidefinite function. Consider \( n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{C} \) a \( t_1, \ldots, t_n \in \mathbb{R} \). Then we have

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \bar{c}_k \cos(t_j - t_k) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \bar{c}_k (\cos t_j \cos t_k + \sin t_j \sin t_k) \\
= \left| \sum_{j=1}^{n} c_j \cos t_j \right|^2 + \left| \sum_{k=1}^{n} c_k \sin t_k \right|^2 \geq 0.
\]

Function \( \cos t \) is positive semidefinite, and according to Theorem 6 there exists a (Gaussian) stochastic process \( \{X_t, t \in T\} \), autocovariance function of which is \( R(s, t) = \cos(s - t) \).
Theorem 8. The sum of two positive semidefinite functions is a positive semidefinite function.

Proof. It follows from the definition of the positive semidefinite function. If \( f \) and \( g \) are positive semidefinite and \( h = f + g \), then for any \( n \in \mathbb{N} \), complex \( c_1, \ldots, c_n \) and \( t_1, \ldots, t_n \in T \)

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} h(t_j, t_k) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} [f(t_j, t_k) + g(t_j, t_k)]
\]

\[
= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} f(t_j, t_k) + \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} g(t_j, t_k) \geq 0.
\]

\[
\square
\]

Corollary 1. Sum of two autocovariance functions is an autocovariance function of a stochastic process with finite second moments.

Proof. It follows from Theorems 6 - 8. \( \square \)

2 Some important classes of stochastic processes

2.1 Markov processes

Definition 14. We say that \( \{X_t, t \in T\} \) is a Markov process with state space \((S, \mathcal{E})\), if for any \( t_0, t_1, \ldots, t_n \), \( 0 \leq t_0 < t_1 < \cdots < t_n \), it holds

\[
P(X_{t_n} \leq x | X_{t_{n-1}}, \ldots, X_{t_0}) = P(X_{t_n} \leq x | X_{t_{n-1}}) \text{ a.s.} \quad (1)
\]

for all \( x \in \mathbb{R} \).

The relation (1) is called Markovian property. Simple cases: discrete state Markov processes, i.e., discrete and continuous time Markov chains.

Example 4. Consider Markov chain \( \{X_t, t \geq 0\} \) with state space \( S = \{0, 1\} \), initial distribution \( P(X_0 = 0) = 1, P(X_0 = 1) = 0 \) and the intensity matrix

\[
Q = \begin{pmatrix}
-\alpha & \alpha \\
\beta & -\beta 
\end{pmatrix}, \quad \alpha > 0, \ \beta > 0
\]

Treat the stationarity of this process.
We know: $P(t) = \exp(Qt)$,

$$P(t) = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha e^{-(\alpha+\beta)t} & \alpha - \alpha e^{-(\alpha+\beta)t} \\ \beta - \beta e^{-(\alpha+\beta)t} & \alpha + \beta e^{-(\alpha+\beta)t} \end{pmatrix}.$$

$$p(t)^T = p(0)^T P(t) = (1, 0)^T P(t) = (p_{00}(t), p_{01}(t))^T$$

Due to the initial distribution we have

$$E[X_t] = P(X_t = 1) = p_{01}(t) = \frac{1}{\alpha + \beta} \cdot (\alpha - \alpha e^{-(\alpha+\beta)t}),$$

which depends on $t$, the process is neither strictly nor weakly stationary.

One the other hand, if the initial distribution is the stationary distribution $\pi$ of the Markov chain, then $\{X_t, t \geq 0\}$ is a strictly stationary process with the mean $E[X_t] = \frac{\alpha}{\alpha + \beta}$ and the autocovariance function

$$R(s, t) = \frac{\alpha \beta}{(\alpha + \beta)^2} e^{-(\alpha+\beta)|s-t|}.$$ 

It means it is also weakly stationary.

Proof of strict stationarity: see the Course Stochastic Processes 1.

Stationary distribution - solution of $\pi^T = \pi^T P(t)$ gives

$$\pi_0 = \frac{\beta}{\alpha + \beta}, \pi_1 = \frac{\alpha}{\alpha + \beta}.$$ 

Further, stationary distribution implies: $p(t) = \pi \Rightarrow E[X_t] = P(X_t = 1) = \frac{\alpha}{\alpha + \beta}$.

To compute $R(s, t)$, notice that for $t < s$, we have

$$E(X_t X_s) = 1 \cdot P(X_t = 1, X_s = 1)$$

$$= P(X_s = 1|X_t = 1)P(X_t = 1)$$

$$= p_{11}(s - t)P(X_t = 1)$$

### 2.2 Independent increment processes

**Definition 15.** A process $\{X_t, t \in T\}$, where $T$ is an interval, has independent increments, if for any $t_1, t_2, \ldots, t_n \in T$ such that $t_1 < t_2 < \cdots < t_n$, the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

If for any $s, t \in T$, $s < t$, the distribution of the increments $X_t - X_s$ depends only on $t - s$, we say that $\{X_t, t \in T\}$ has stationary increments.
Example 5. Poisson process with intensity $\lambda$ is a continuous time Markov chain $\{X_t, t \geq 0\}$ such that $X_0 = 0$ a.s. and for $t > 0$, $X_t$ has the Poisson distribution with parameter $\lambda t$. Increments $X_t - X_s$, $s < t$ have the Poisson distribution with the parameter $\lambda(t - s)$. Poisson process is neither strictly nor weakly stationary.

Example 6. Wiener process (Brownian motion process) is a Gaussian stochastic process $\{W_t, t \geq 0\}$ with the properties

1. $W_0 = 0$ a.s. and $\{W_t, t \geq 0\}$ has continuous trajectories
2. For any $0 \leq t_1 < t_2 < \cdots < t_n$, $W_{t_1}, W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent random variables (independent increments).
3. For any $0 \leq t < s$, the increments $W_s - W_t$ have normal distribution with zero mean and the variance $\sigma^2(s - t)$, where $\sigma^2$ is a positive constant. Especially, for any $t \geq 0$, $\text{E}W_t = 0$ and $\text{var}W_t = \sigma^2 t$.

The Wiener process is Markovian but it is neither strictly nor weakly stationary.

2.3 Martingales

Definition 16. Let $\{\Omega, \mathcal{A}, P\}$ be a probability space, $T \subset \mathbb{R}, T \neq \emptyset$. Let for any $t \in T$, $\mathcal{F}_t \subset \mathcal{A}$ be a $\sigma-$ algebra ($\sigma-$field). The system of $\sigma$-fields $\{\mathcal{F}_t, t \in T\}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for any $s, t \in T, s < t$ is called a filtration.

Definition 17. Let $\{X_t, t \in T\}$ be a stochastic process defined on $\{\Omega, \mathcal{A}, P\}$, and let $\{\mathcal{F}_t, t \in T\}$ be a filtration. We say that $\{X_t, t \in T\}$ is adapted to $\{\mathcal{F}_t, t \in T\}$ if for any $t \in T$, $X_t$ is $\mathcal{F}_t$ measurable.

Definition 18. Let $\{X_t, t \in T\}$ be adapted to $\{\mathcal{F}_t, t \in T\}$ and $E|X_t| < \infty$ for all $t \in T$. Then $\{X_t, t \in T\}$ is said to be a martingale if $EX_t|\mathcal{F}_s = X_s$ a.s. for any $s < t, s, t \in T$.

3 Hilbert space

3.1 Inner product space

Definition 19. A complex vector space $H$ is said to be an inner product space, if for any $x, y \in H$ there exists a number $\langle x, y \rangle \in \mathbb{C}$, called the inner product of elements $x, y$ such that

1. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
2. $\forall \alpha \in \mathbb{C}, \forall x, y \in H : \langle \alpha x, y \rangle = \alpha \langle x, y \rangle$. 

11
3. \( \forall x, y, z \in H \ (x + y, z) = (x, z) + (y, z) \).

4. \( \forall x \in H \) is \( (x, x) \geq 0; \ (x, x) = 0 \iff x = 0 \) (0 is zero element in \( H \)).

The number 
\[
\|x\| := \sqrt{(x, x)}, \quad \forall x \in H
\]
is called the norm of an element \( x \).

**Theorem 9.** For \( \|x\| := \sqrt{(x, x)} \) it holds

1. \( \|x\| \geq 0 \ \forall x \in H \) and \( \|x\| = 0 \iff x = 0 \).
2. \( \forall \alpha \in \mathbb{C}, \forall x \in H: \ |\alpha x| = |\alpha| \cdot \|x\| \).
3. \( \forall x, y \in H: \ |x + y| \leq \|x\| + \|y\| \).
4. \( \forall x, y \in H \)
   \[
   |(x, y)| \leq \|x\| \cdot \|y\| = \sqrt{(x, x)} \sqrt{(y, y)}. 
   \]
   (the Cauchy-Schwarz inequality)

**Proof.** It can be found in any textbook on Functional Analysis, see, e.g., Rudin (2003), Chap. 4. \( \square \)

### 3.2 Convergence in norm

**Definition 20.** We say that a sequence \( \{x_n, n \in \mathbb{N}\} \) of elements of a inner product space \( H \) converges in norm to element \( x \in H \), if \( \|x_n - x\| \to 0 \) as \( n \to \infty \).

**Theorem 10.** (The inner product continuity)

Let \( \{x_n, n \in \mathbb{N}\} \) and \( \{y_n, n \in \mathbb{N}\} \) be sequences of elements of \( H \). Let \( x, y \in H \) and \( x_n \to x, y_n \to y \) in norm as \( n \to \infty \). Then

\[
\|x_n\| \to \|x\| \quad \text{and} \quad (x_n, y_n) \to (x, y).
\]
Proof. From the triangular inequality we get

\[ ||x|| \leq ||x - y|| + ||y|| \]
\[ ||y|| \leq ||y - x|| + ||x|| \]
\[ ||x|| - ||y|| \leq ||x - y|| \]

From here we get the first assertion since

\[ ||x_n|| - ||x|| \leq ||x_n - x||. \]

The second assertion is obtained by using the Cauchy-Schwarz inequality

\[ |\langle x_n, y_n \rangle - \langle x, y \rangle| = |\langle x_n - x + x, y_n - y + y \rangle - \langle x, y \rangle| \]
\[ \leq |\langle x_n - x, y_n - y \rangle| + |\langle x, y_n - y \rangle| + |\langle x_n - x, y \rangle| \]
\[ \leq ||x_n - x|| \cdot ||y_n - y|| + ||x|| \cdot ||y_n - y|| + ||x_n - x|| \cdot ||y||. \]

\[ \Box \]

**Definition 21.** A sequence \( \{x_n, n \in \mathbb{N}\} \) of elements of \( H \) is said to be a Cauchy sequence, if \( ||x_n - x_m|| \to 0 \) as \( n, m \to \infty \).

**Definition 22.** An inner product space \( H \) is defined to be a Hilbert space, if it is complete, i.e., if any Cauchy sequence of elements of \( H \) converges in norm to some element of \( H \).

4 Space \( L_2(\Omega, \mathcal{A}, P) \)

4.1 Construction

Let \( \mathcal{L} \) be a set of all random variables with finite second moments defined on a probability space \( (\Omega, \mathcal{A}, P) \).

\( \mathcal{L} \) is a vector space:

- \( X + Y \in \mathcal{L} \forall X, Y \in \mathcal{L} \), since

\[ \mathbb{E}|X + Y|^2 \leq 2 (\mathbb{E}|X|^2 + \mathbb{E}|Y|^2) < \infty, \]
\[ \alpha X \in \mathcal{L} \quad \forall X \in \mathcal{L}, \quad \text{and} \quad \forall \alpha \in \mathbb{C}, \quad \text{since} \quad E|\alpha X|^2 = |\alpha|^2 \cdot E|X|^2 < \infty \]

- The null element of \( \mathcal{L} \) is the random variable identically equal to zero.

On the space \( \mathcal{L} \) we define classes of equivalent random variables that satisfy

\[ X \sim Y \iff P [X = Y] = 1 \]

and on the set of classes of equivalent random variables from \( \mathcal{L} \) define the relation

\[ \langle X, Y \rangle = E [XY], \quad \forall X \in \widetilde{X}, Y \in \widetilde{Y}, \]

where \( \widetilde{X}, \widetilde{Y} \) denote classes of equivalence.

The space of classes of equivalence on \( \mathcal{L} \) with the above relation \( \langle \ldots, \ldots \rangle \) is denoted \( L_2(\Omega, \mathcal{A}, P) \).

The relation \( \langle X, Y \rangle \) satisfies the properties of the inner product on \( L_2(\Omega, \mathcal{A}, P) \):

1. \( \langle \alpha X, Y \rangle = E [\alpha XY] = \alpha E [XY] = \alpha \langle X, Y \rangle \).
2. \( \langle X + Y, Z \rangle = E [(X + Y)Z] = E [XZ] + E [YZ] = \langle X, Z \rangle + \langle Y, Z \rangle \).
3. \( \langle X, X \rangle = E |X|^2 \geq 0 \).
4. \( \langle X, X \rangle = E|X|^2 = 0 \iff X \sim 0 \).

### 4.2 Mean square convergence

\( L_2(\Omega, \mathcal{A}, P) \) is the space of equivalence classes on \( \mathcal{L} \) with the inner product

Norm is defined by

\[ ||X|| := \sqrt{E|X|^2} \]

convergence in \( v \) \( L_2(\Omega, \mathcal{A}, P) \) is the convergence in this norm.

**Definition 23.** We say that a sequence of random variables \( X_n \) such that \( E|X_n|^2 < \infty \) converges in the mean square (or in the squared mean) to a random variable \( X \), if it converges to \( X \) in \( L_2(\Omega, \mathcal{A}, P) \), i.e.,

\[ ||X_n - X||^2 = E|X_n - X|^2 \to 0 \quad \text{as} \quad n \to \infty. \]

Notation: \( X = l.i.m. \ X_n \) (limit in the (squared) mean).
Theorem 11. The space $L_2(\Omega, \mathcal{A}, P)$ is complete.

Proof. See, e.g., Brockwell, Davis (1991), Par. 2.10, or Rudin (2003), Theorem 3.11. □

The space $L_2(\Omega, \mathcal{A}, P)$ is the Hilbert space.

Convention:
A stochastic process $\{X_t, t \in T\}$ such that $E|X_t|^2 < \infty$ will be called a second order process.

4.3 Hilbert space generated by a stochastic process

Definition 24. Let $\{X_t, t \in T\}$ be a stochastic process with finite second moments on an $(\Omega, \mathcal{A}, P)$. The set $\mathcal{M}\{X_t, t \in T\}$ of all finite linear combinations of random variables from $\{X_t, t \in T\}$ is a linear span of the process $\{X_t, t \in T\}$, i.e.,

$$\mathcal{M}\{X_t, t \in T\} = \left\{ \sum_{k=1}^{n} c_k X_{t_k}, \; n \in \mathbb{N}, \; c_1, \ldots, c_n \in \mathbb{C}, \; t_1, \ldots, t_n \in T \right\}.$$  

Equivalence classes in $\mathcal{M}\{X_t, t \in T\}$ and the inner product $\langle X, Y \rangle$ are defined as above.

Definition 25. A closure $\overline{\mathcal{M}}\{X_t, t \in T\}$ of the linear span $\mathcal{M}\{X_t, t \in T\}$ consists of all the elements of $\mathcal{M}\{X_t, t \in T\}$ and mean square limits of all convergent sequences of elements of $\mathcal{M}\{X_t, t \in T\}$.

Then $\overline{\mathcal{M}}\{X_t, t \in T\}$ is a closed subspace of the complete space $L_2(\Omega, \mathcal{A}, P)$ and thus a complete inner product space. It is called the Hilbert space generated by a stochastic process $\{X_t, t \in T\}$, notation $\mathcal{H}\{X_t, t \in T\}$.

Definition 26. Let $\{X^h_t, t \in T\}_{h \in S}, T \subset \mathbb{R}, S \subset \mathbb{R}$, be a collection of stochastic processes in $L_2(\Omega, \mathcal{A}, P)$ (shortly: second order processes). We say that processes $\{X^h_t, t \in T\}_{h \in S}$ converge in mean square to a second order process $\{X_t, t \in T\}$ as $h \to h_0$, if

$$\forall t \in T : \; X^h_t \overset{h \to h_0}{\longrightarrow} X_t \; \text{in mean square}.$$  

Briefly:

$$\{X^h_t, t \in T\}_{h \in S} \overset{h \to h_0}{\longrightarrow} \{X_t, t \in T\} \; \text{in mean square}.$$
Theorem 12. Centered second order processes \( \{X^h_t, t \in T\}_{h \in S} \) converge in mean square to a centered second order process \( \{X_t, t \in T\} \) as \( h \to h_0 \) if and only if

\[
E \left[ X^h_t X^{h'}_t \right] \to b(t) \quad \text{as} \quad h, h' \to h_0,
\]

where \( b(.) \) is a finite function on \( T \).

When processes \( \{X^h_t, t \in T\}_{h \in S} \) converge to a process \( \{X_t, t \in T\} \) in mean square as \( h \to h_0 \), the autocovariance functions of the processes \( \{X^h_t, t \in T\}_{h \in S} \) converge to the autocovariance function of \( \{X_t, t \in T\} \) as \( h \to h_0 \).

Proof. 1. Let \( \{X^h_t, t \in T\}_{h \in S} \xrightarrow{h \to h_0} \{X_t, t \in T\} \) in mean square. Then \( \forall t, t' \in T \)

\[
X^h_t \xrightarrow{h \to h_0} X_t \quad \text{in mean square}
\]

\[
X^{h'}_{t'} \xrightarrow{h' \to h_0} X_{t'} \quad \text{in mean square}.
\]

It follows from the continuity of the inner product that as \( h, h' \to h_0 \),

\[
E \left[ X^h_t X^{h'}_{t'} \right] \to E \left[ X_t X_{t'} \right].
\]

Thus for \( t = t' \) a \( h, h' \to h_0 \) we have

\[
E \left[ X^h_t X^{h'}_t \right] \to E \left[ X_t X_t \right] = E|X_t|^2 := b(t) < \infty,
\]

since \( \{X_t, t \in T\} \) is a second order process. For \( h = h' \), we get

\[
E \left[ X^h_t X^{h'}_{t'} \right] \xrightarrow{h \to h_0} E \left[ X_t X_{t'} \right] \quad \text{as} \quad t, t' \in T,
\]

where \( E \left[ X^h_t X^{h'}_{t'} \right] = R_h(t, t') \) is the autocovariance function of the process \( \{X^h_t, t \in T\} \) and \( E \left[ X_t X_{t'} \right] = R(t, t') \) is the autocovariance function of the process \( \{X_t, t \in T\} \).

2. Let \( \{X^h_t, t \in T\}_{h \in S} \) be centered second order processes for which

\[
E \left[ X^h_t X^{h'}_t \right] \to b(t) < \infty \quad \text{as} \quad h, h' \to h_0 \quad \forall t \in T.
\]

Then

\[
\|X^h_t - X^{h'}_t\|^2 \to 0, \quad \text{as} \quad h, h' \to h_0, \quad \forall t \in T
\]
since \( \forall t \in T \)

\[
\|X_t^h - X_t^{h'}\|^2 = E \left[ (X_t^h - X_t^{h'}) (X_t^h - X_t^{h'}) \right] \\
= E \left[ X_t^h X_t^h \right] - E \left[ X_t^{h'} X_t^h \right] - E \left[ X_t^h X_t^{h'} \right] + \\
+ E \left[ X_t^{h'} X_t^{h'} \right] \rightarrow b(t) - b(t) - b(t) + b(t) = 0
\]

as \( h, h' \to h_0 \).

For all \( t, h, X_t^h \in L_2(\Omega, \mathcal{A}, P) \) which is complete \( \Rightarrow \exists X_t \in L_2(\Omega, \mathcal{A}, P) : X_t^h \to X_t \) in mean square as \( h \to h_0 \), thus \( E|X_t|^2 < \infty \ \forall t \in T, \Rightarrow \) there exist a limit process \( \{X_t, t \in T\} \in L_2(\Omega, \mathcal{A}, P) \). We prove that \( \{X_t, t \in T\} \) is centered:

\[
EX_t = EX_t - EX_t^h + EX_t^h = E \left[ X_t - X_t^h \right].
\]

Then

\[
|EX_t| = |E \left[ X_t - X_t^h \right]| \leq \sqrt{E \left| X_t - X_t^h \right|^2} \to 0
\]

as \( h \to h_0, \forall t \in T \).

\[\square\]

5 Continuous time processes in \( L_2(\Omega, \mathcal{A}, P) \)

5.1 Mean square continuity

**Definition 27.** Let \( \{X_t, t \in T\} \) be a second order process, \( T \subset \mathbb{R} \) an open interval. We say that the process \( \{X_t, t \in T\} \) is mean square continuous (or \( L_2 \)-continuous at a point \( t_0 \in T \), if

\[
E|X_t - X_{t_0}|^2 \to 0 \ \text{pro} \ t \to t_0.
\]

We say that the process \( \{X_t, t \in T\} \) is mean square continuous, if it is continuous at each point of \( T \).

**Remark 2.** A second order process that is mean square continuous is also stochastically continuous (continuous in probability), since

\[
P \left[ |X_t - X_{t_0}| > \varepsilon \right] \leq \varepsilon^{-2} \cdot E|X_t - X_{t_0}|^2.
\]

**Theorem 13.** Let \( \{X_t, t \in T\} \) be a centered second order process, \( T \subset \mathbb{R} \) be an interval. Then \( \{X_t, t \in T\} \) is mean square continuous if and only if its autocovariance function \( R(s,t) \) is continuous at points \( [s,t] \), such that \( s = t \).
which follows from the continuity of the inner product, since

$X$ and an autocovariance function

Due to the weak stationarity,

Proof.

Let $\{X_t, t \in T\}$ be a centered mean square continuous process with a mean value $\{\mu_t, t \in T\}$ and an autocovariance function $R(s,t)$ defined on $T \times T$. Then $\{X_t, t \in T\}$ is mean square continuous if $\{\mu_t, t \in T\}$ is continuous on $T$ and $R(s,t)$ is continuous in points $[s,t]$, such that $s = t$.

Proof.

$$E|X_t - X_{t_0}|^2 = E|X_t - \mu_t + \mu_t - X_{t_0}|^2 =$$

$$= E[|X_t - \mu_t - (X_{t_0} - \mu_{t_0}) + \mu_t - \mu_{t_0}|^2]$$

Put $Y_t := X_t - \mu_t$, $\forall t \in T$. Then $\{Y_t, t \in T\}$ is centered process with the same autocovariance function $R(s,t)$ and

$$E|X_t - X_{t_0}|^2 = E|Y_t - Y_{t_0} + \mu_t - \mu_{t_0}|^2 \leq$$

$$\leq 2E|Y_t - Y_{t_0}|^2 + 2|\mu_t - \mu_{t_0}|^2$$

The limit on the right hand side is zero as $t \to t_0$, thus the limit on the left hand side is zero.

Theorem 15. Let $\{X_t, t \in T\}$ be a centered weakly stationary process with an autocovariance function $R(t)$. Then $\{X_t, t \in T\}$ is mean square continuous if and only if $R(t)$ is continuous at zero.

Proof. Due to the weak stationarity, $R(s,t) = R(s-t)$. Then the assertion follows from the previous theorem.
Example 7. A centered weakly stationary process with the autocovariance function \( R(t) = \cos(t), \ t \in \mathbb{R} \), is mean square continuous.

Example 8. Let \( \{X_t, t \in T\}, \ T = \mathbb{R} \), be a process of uncorrelated random variables with \( \mathbb{E}X_t = 0, \ t \in \mathbb{R} \) and the same variance \( 0 < \sigma^2 < \infty \). The autocovariance function is \( R(s,t) = \sigma^2 \delta_{s-t} \) where

\[
\delta(x) = \begin{cases} 
1 & \text{if } x = 0, \\
0 & \text{if } x \neq 0.
\end{cases}
\]

The process is weakly stationary, but not mean square continuous (the autocovariance function is not continuous at zero).

Example 9. Wiener process \( \{W_t, t \geq 0\} \) is a Gaussian process with independent and stationary increments, \( \mathbb{E}W_t = 0, \ R(s,t) = \mathbb{E}W_sW_t = \sigma^2 \cdot \min\{s,t\} \Rightarrow \) is not weakly stationary.

The process is centered, \( R(s,t) \) is continuous (thus at \([s,t]\) with \( s = t \) \( \Rightarrow \) \( \{W_t\} \) is mean square continuous.

Example 10. Poisson process \( \{X_t, t \geq 0\} \) with intensity \( \lambda > 0 \) is a process with stationary and independent increments, \( X_t \sim \text{Po} (\lambda) \). Since \( \mathbb{E}X_t = \mu_t = \lambda t, \ t \geq 0 \) and \( \text{cov}(X_s, X_t) = \lambda \cdot \min\{s,t\} \Rightarrow \) the process is not weakly stationary.

Since \( \mu_t \) is continuous, \( R(s,t) \) is continuous \( \Rightarrow \) the process is mean square continuous.

5.2 Mean square derivative of the process

Definition 28. Let \( \{X_t, t \in T\} \) be a second order process, \( T \subset \mathbb{R} \) an open interval. We say that the process is mean square differentiable (\( L_2 \)-differentiable) at point \( t_0 \) if there exists the mean square limit

\[
\lim_{h \to 0} \mathbb{E} \frac{X_{t_0+h} - X_{t_0}}{h} := X_{t_0}'.
\]

This limit is called the mean square derivative (\( L_2 \)-derivative) of the process at \( t_0 \).

We say that the process \( \{X_t, t \in T\} \) is mean square differentiable, if it is mean square differentiable at every point \( t \in T \).

Theorem 16. A centered second order process \( \{X_t, t \in T\} \) is mean square differentiable if and only if there exists a finite generalized second-order partial derivative of its autocovariance function \( R(s,t) \) at points \([s,t]\), where \( s = t \), i.e., if at these points there exists finite limit

\[
\lim_{h,h' \to 0} \frac{1}{hh'} \left[ R(s+h,t+h') - R(s,t+h') - R(s+h,t) + R(s,t) \right].
\]
Figure 1: A trajectory of a Wiener process

Figure 2: The autocovariance function of a Wiener process
Proof. According to Theorem 12 the necessary and sufficient condition for the mean square convergence of \( (X_{t+h} - X_t)/h \) is the existence of the finite limit
\[
\lim_{h,h' \to 0} E \left[ \frac{X_{t+h} - X_t}{h} \cdot \frac{X_{t+h'} - X_t}{h'} \right] = \\
\lim_{h,h' \to 0} \frac{1}{hh'} \left[ R(t+h,t+h') - R(t,t+h') - R(t+h,t) + R(t,t) \right].
\]

Remark 3. A sufficient condition the generalized second-order partial derivative of \( R(s,t) \) to exist:

Let \([s,t]\) be an interior point in \( T \times T \). If there exist
\[
\frac{\partial^2 R(s,t)}{\partial s \partial t} \quad \text{and} \quad \frac{\partial^2 R(s,t)}{\partial t \partial s}
\]
and they are continuous, then there exists the generalized second-order partial derivative of \( R(s,t) \) and is equal to \( \frac{\partial^2 R(s,t)}{\partial s \partial t} \) (Anděl (1976), p. 20).

Theorem 17. A second order process \( \{X_t, t \in T\} \) with a mean value \( \{\mu_t, t \in T\} \) is mean square differentiable, if \( \{\mu_t, t \in T\} \) is differentiable and the generalized second-order partial derivative of the autocovariance function exists and is finite at points \([s,t]\), such that \( s = t \).

Proof. A sufficient condition for the mean square limit of \( \frac{X_{t+h} - X_t}{h} \) to exists is the Cauchy condition
\[
E \left[ \left| \frac{X_{t+h} - X_t}{h} - \frac{X_{t+h'} - X_t}{h'} \right|^2 \right] \to 0 \quad \text{as} \quad h \to 0, \ h' \to 0
\]
\( \forall t \in T \). It holds since
\[
E \left| \frac{X_{t+h} - X_t}{h} - \frac{X_{t+h'} - X_t}{h'} \right|^2 \leq 2E \left| \frac{Y_{t+h} - Y_t}{h} - \frac{Y_{t+h'} - Y_t}{h'} \right|^2 + 2 \left| \frac{\mu_{t+h} - \mu_t}{h} - \frac{\mu_{t+h'} - \mu_t}{h'} \right|^2,
\]
where \( Y_t = X_t - \mu_t \). According to Theorem 16, the process \( \{Y_t, t \in T\} \) is mean square differentiable, and the first term on the right hand side of the previous inequality converges to zero as \( h \to 0, h' \to 0 \). The second term converges to zero, since function \( \{\mu_t, t \in T\} \) is differentiable. \( \square \)
Example 11. A centered weakly stationary process with the autocovariance function \( R(s, t) = \cos(s - t) \), \( s, t \in \mathbb{R} \), is mean square differentiable, since

\[
\frac{\partial^2 \cos(s - t)}{\partial s \partial t} \quad \text{and} \quad \frac{\partial^2 \cos(s - t)}{\partial t \partial s}
\]

exist and they are continuous.

Example 12. Poisson process \( \{X_t, t > 0\} \) has the mean value \( \mu_t = \lambda t \), which is continuous and differentiable for all \( t > 0 \) and the autocovariance function \( R(s, t) = \lambda \min(s, t) \). The generalized second-order partial derivative of \( R(s, t) \) however is not finite: for \( s = t \) we have

\[
\lim_{h \to 0+} \frac{1}{h^2} [s + h - \min(s + h, s) - \min(s, s + h) + s] = +\infty,
\]

\[
\lim_{h \to 0-} \frac{1}{h^2} [s + h - \min(s + h, s) - \min(s, s + h) + s] = +\infty.
\]

Poisson process is not mean square differentiable.

5.3 Riemann integral

Definition 29. Let \( T = [a, b] \) be a closed interval, \(-\infty < a < b < +\infty\). Let \( D_n = \{t_{n,0}, t_{n,1}, \ldots, t_{n,n}\} \), where \( a = t_{n,0} < t_{n,1} < \ldots < t_{n,n} = b \), \( \forall n \in \mathbb{N} \) be a partition of the interval \([a, b]\). Denote the norm of the partition \( D_n \) to be

\[
\Delta_n := \max_{0 \leq i \leq n-1} (t_{n,i+1} - t_{n,i})
\]

and define partial sums \( I_n \) of a centered second order process \( \{X_t, t \in [a, b]\} \) by

\[
I_n := \sum_{i=0}^{n-1} X_{t_{n,i}} (t_{n,i+1} - t_{n,i}), \quad n \in \mathbb{N}.
\]

If the sequence \( \{I_n, n \in \mathbb{N}\} \) has the mean square limit \( I \) for any partition of the interval \([a, b]\) such that \( \Delta_n \to 0 \) as \( n \to \infty \), we call it Riemann integral of the process \( \{X_t, t \in [a, b]\} \) and write

\[
I = \int_a^b X_t dt.
\]

If the process \( \{X_t, t \in T\} \) has mean value \( \{\mu_t, t \in T\} \), we define Riemann integral of the process \( \{X_t, t \in [a, b]\} \) to be

\[
\int_a^b X_t dt = \int_a^b (X_t - \mu_t) dt + \int_a^b \mu_t dt,
\]

if the centered process \( \{X_t - \mu_t\} \) is Riemann integrable and \( \int_a^b \mu_t dt \) exists and is finite.

22
Theorem 18. Let \( \{X_t, t \in [a, b]\} \) be a centered second order process with the autocovariance function \( R(s, t) \). Then the Riemann integral \( \int_a^b X_t \, dt \) exists, if the Riemann integral \( \int_a^b \int_a^b R(s, t) \, ds \, dt \) exists and is finite.

Proof. Let \( D_m = \{s_{m,0}, \ldots, s_{m,m}\} \), \( D_n = \{t_{n,0}, \ldots, t_{n,n}\} \) be partitions of interval \([a, b]\), the norms \( \Delta_m, \Delta_n \) of which converge to zero as \( m, n \to \infty \). Put

\[
I_m := \sum_{j=0}^{m-1} (s_{j+1,m} - s_{j,m}) X_{s_{m,m}}
\]

\[
I_n := \sum_{k=0}^{n-1} (t_{k+1,n} - t_{k,n}) X_{t_{n,n}}.
\]

Similarly as in the proof of Theorem 12 we can see that \( \int_a^b X_t \, dt \) exist if there exist the finite limit

\[
\mathbb{E} \left[ I_m I_n \right] = \lim \mathbb{E} \left\{ \left[ \sum_{j=0}^{m-1} X_{s_{m,j}} (s_{m,j+1} - s_{m,j}) \right] \cdot \left[ \sum_{k=0}^{n-1} X_{t_{n,k}} (t_{n,k+1} - t_{n,k}) \right] \right\}
\]

\[
= \lim \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} R(s_{m,j}, t_{n,k})(s_{m,j+1} - s_{m,j})(t_{n,k+1} - t_{n,k})
\]

as \( m, n \to \infty \), \( \Delta_m, \Delta_n \to 0 \), which follows from the existence of \( \int_a^b \int_a^b R(s, t) \, ds \, dt \).

Example 13. Riemann integral \( \int_a^b X_t \, dt \) of a centered continuous time process with the autocovariance function \( R(s, t) = \cos(s - t) \) exists, since \( R(s, t) \) is continuous on \([a, b] \times [a, b]\).

Example 14. Let \( \{X_t, t \in \mathbb{R}\} \) be a centered second order process. We define

\[
\int_{-\infty}^\infty X_t \, dt := \text{l. i. m.} \int_a^b X_t \, dt \quad \text{as} \quad a \to -\infty, \ b \to \infty,
\]

if the limit and the Riemann integral on the right hand side exist.

Example 15. Poisson process \( \{X_t, t \geq 0\} \) is Riemann integrable on any finite interval \([a, b] \subset [0, \infty)\), since its autocovariance function is continuous on \([a, b] \times [a, b]\).
6 Spectral decomposition of autocovariance function

6.1 Auxiliary assertions

**Lemma 1.** 1. Let $\mu, \nu$ be finite measures on Borel subsets of the interval $[-\pi, \pi]$. If for every $t \in \mathbb{Z}$,
$$
\int_{-\pi}^{\pi} e^{it\lambda} d\mu(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} d\nu(\lambda),
$$
then $\mu(B) = \nu(B)$ for every $B \subset (-\pi, \pi)$ and $\mu(\{-\pi\} \cup \{\pi\}) = \nu(\{-\pi\} \cup \{\pi\})$.

2. Let $\mu, \nu$ be finite measures on $(\mathbb{R}, \mathcal{B})$. If for every $t \in \mathbb{R}$
$$
\int_{-\infty}^{\infty} e^{it\lambda} d\mu(\lambda) = \int_{-\infty}^{\infty} e^{it\lambda} d\nu(\lambda),
$$
then $\mu(B) = \nu(B)$ for all $B \subset \mathcal{B}$.

**Proof.** See Andéel (1976), III.1, Theorems 5 and 6.

**Lemma 2** (Helly theorem). Let $\{F_n, n \in \mathbb{N}\}$ be a sequence on non-decreasing uniformly bounded functions. Then there exists a subsequence $\{F_{n_k}\}$, that, as $k \to \infty$, $n_k \to \infty$, converges weakly to a non-decreasing right-continuous function $F$, i.e., on the continuity set of $F$.

**Proof.** Rao (1978), Theorem 2c.4, I.

**Lemma 3** (Helly-Bray). Let $\{F_n, n \in \mathbb{N}\}$ be a sequence of non-decreasing uniformly bounded functions that, as $n \to \infty$ converges weakly to a non-decreasing bounded right-continuous function $F$, and $\lim F_n(-\infty) = F(-\infty), \lim F_n(+\infty) = F(+\infty)$. Let $f$ be a continuous bounded function. Then
$$
\int_{-\infty}^{\infty} f(x) dF_n(x) \longrightarrow \int_{-\infty}^{\infty} f(x) dF(x) \text{ as } n \to \infty.
$$

**Proof.** Rao (1978), Theorem 2c.4, II.

**Remark 4.** The integral at the Helly-Bray theorem is the Riemann-Stieltjes integral of a function $f$ with respect to a function $F$. If $[a, b]$ is a bounded interval and $F$ is right-continuous, we will understand that
$$
\int_{a}^{b} f(x) dF(x) := \int_{[a,b]} f(x) dF(x).
$$
6.2 Spectral decomposition of autocovariance function

**Theorem 19.** A complex-valued function \( R(t) \), \( t \in \mathbb{Z} \), is an autocovariance function of a stationary random sequence if and only if

\[
R(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda) \quad \text{for all } t \in \mathbb{Z},
\]

(2)

where \( F \) is a right-continuous non-decreasing bounded function on \([-\pi, \pi]\), \( F(-\pi) = 0 \). The function \( F \) is determined by formula (2) uniquely.

Formula (2) is called the **spectral decomposition** (or representation) of an autocovariance function of a stationary random sequence. The function \( F \) is called the **spectral distribution function** of a random sequence.

**Proof.** 1. Suppose that (2) holds for any complex-valued function \( R \) on \( \mathbb{Z} \). Then \( R \) is positive semidefinite since for any \( n \in \mathbb{N} \), any constants \( c_1, \ldots, c_n \in \mathbb{C} \) and all \( t_1, \ldots, t_n \in \mathbb{Z} \)

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c}_k R(t_j - t_k) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c}_k \int_{-\pi}^{\pi} e^{it_j - it_k} e^{i\lambda} dF(\lambda) = \int_{-\pi}^{\pi} \left[ \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c}_k e^{it_j} e^{-it_k} \right] e^{i\lambda} dF(\lambda) = \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} c_j e^{it_j} \right|^2 dF(\lambda) \geq 0,
\]

because \( F \) is non-decreasing in \([-\pi, \pi]\). It means that \( R \) is an autocovariance function of a stationary random sequence.

2. Let \( R \) be an autocovariance function of a stationary random sequence; it is positive semidefinite, i.e.,

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c}_k R(t_j - t_k) \geq 0 \quad \text{for all } n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{C} \text{ and } t_1, \ldots, t_n \in \mathbb{Z}.
\]

Put \( t_j = j, \ c_j = e^{-ij\lambda} \) for a \( \lambda \in [-\pi, \pi] \).
Then for every \( n \in \mathbb{N}, \lambda \in [-\pi, \pi] \),

\[
\varphi_n(\lambda) := \frac{1}{2\pi n} \sum_{j=1}^{n} \sum_{k=1}^{n} e^{-i(j-k)\lambda} R(j - k) \geq 0.
\]
From here we get

\[
\varphi_n(\lambda) = \frac{1}{2\pi n} \sum_{j=1}^{n} \sum_{k=1}^{n} e^{-i(j-k)\lambda} R(j-k)
\]

\[
= \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} \sum_{j=\max(1,\kappa+1)}^{\min(n,\kappa+n)} e^{-i\kappa\lambda} R(\kappa)
\]

\[
= \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda} R(\kappa)(n - |\kappa|).
\]

For any \( n \in \mathbb{N} \) let us define function

\[
F_n(x) = \begin{cases} 
0, & x \leq -\pi, \\
\int_{-\pi}^{x} \varphi_n(\lambda) d\lambda, & x \in [-\pi, \pi], \\
F_n(\pi), & x \geq \pi.
\end{cases}
\]

Obviously, \( F_n(-\pi) = 0 \) and \( F_n(x) \) is non-decreasing on \([-\pi, \pi]\). Compute \( F_n(\pi) \):

\[
F_n(\pi) = \int_{-\pi}^{\pi} \varphi_n(\lambda) d\lambda
\]

\[
= \frac{1}{2\pi n} \int_{-\pi}^{\pi} \left[ \sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda} R(\kappa)(n - |\kappa|) \right] d\lambda
\]

\[
= \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} R(\kappa)(n - |\kappa|) \int_{-\pi}^{\pi} e^{-i\kappa\lambda} d\lambda = R(0),
\]

since the last integral is \( 2\pi \delta(\kappa) \). \{F_n, n \in \mathbb{N}\} is a sequence of non-decreasing functions, \( 0 \leq F_n(x) \leq R(0) < \infty \) for all \( x \in \mathbb{R} \) and all \( n \in \mathbb{N} \).

According to the Helly theorem there exists a subsequence \( \{F_{n_k}\} \subset \{F_n\} \), \( F_{n_k} \to F \) weakly as \( k \to \infty, n_k \to \infty \), where \( F \) is a nondecreasing bounded right-continuous function and \( F(x) = 0, x \leq -\pi, F(x) = R(0), x > \pi \).

From the Helly - Bray theorem for \( f(x) = e^{itx} \), where \( t \in \mathbb{Z} \),

\[
\int_{-\pi}^{\pi} e^{it\lambda} dF_{n_k}(\lambda) \longrightarrow \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda) \quad \text{as} \quad k \to \infty, \ n_k \to \infty.
\]
On the other hand,

\[
\int_{-\pi}^{\pi} e^{it\lambda} dF_{n_k}(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} \varphi_{n_k}(\lambda) d\lambda
\]

\[
= \int_{-\pi}^{\pi} e^{it\lambda} \left[ \frac{1}{2\pi n_k} \sum_{\kappa = -n_k+1}^{n_k-1} e^{-i\kappa\lambda} R(\kappa)(n_k - |\kappa|) \right] d\lambda
\]

\[
= \frac{1}{2\pi n_k} \sum_{\kappa = -n_k+1}^{n_k-1} R(\kappa)(n_k - |\kappa|) \int_{-\pi}^{\pi} e^{i(t - \kappa)\lambda} d\lambda,
\]

thus

\[
\int_{-\pi}^{\pi} e^{it\lambda} dF_{n_k}(\lambda) = \begin{cases} 
R(t) \left( 1 - \frac{|t|}{n_k} \right), & |t| < n_k \\
0 & \text{elsewhere.}
\end{cases}
\]

We get

\[
\lim_{k \to \infty} \int_{-\pi}^{\pi} e^{it\lambda} dF_{n_k}(\lambda) = \lim_{k \to \infty} R(t) \left( 1 - \frac{|t|}{n_k} \right)
\]

\[
= R(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda)
\]

Uniqueness:

Let \( R(t) = \int_{-\pi}^{\pi} e^{it\lambda} dG(\lambda) \), where \( G \) is a right-continuous non-decreasing bounded function on \([-\pi, \pi]\) and \( G(-\pi) = 0 \).

Then

\[
\int_{-\pi}^{\pi} e^{it\lambda} d\mu_F = \int_{-\pi}^{\pi} e^{it\lambda} d\mu_G,
\]

where \( \mu_F \) a \( \mu_G \) are finite measures on Borel subsets of the interval \([-\pi, \pi]\) induced by functions \( F \) a \( G \), respectively. The rest follows from Lemma 1.

Formula (2) is called the spectral decomposition of the autocovariance function of a stationary random sequence. Function \( F \) is called the spectral distribution function of a random sequence.

If there exists a function \( f(\lambda) \geq 0 \) for \( \lambda \in [-\pi, \pi] \) such that \( F(\lambda) = \int_{-\pi}^{\lambda} f(x) dx \) (\( F \) is absolutely continuous), then \( f \) is called spectral density. Obviously \( f = F' \).

In case that the spectral density exists, the spectral decomposition of the autocovariance function is of the form

\[
R(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{Z}.
\]
Theorem 20. A complex-valued function $R(t)$, $t \in \mathbb{R}$, is the autocovariance function of a centered stationary mean square continuous process if and only if

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda), \quad t \in \mathbb{R},$$

(4)

where $F$ is non-decreasing right-continuous function such that

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = R(0) < \infty.$$

Function $F$ is determined uniquely.

Proof. 1. Let $R$ be a complex-valued function on $\mathbb{R}$ that satisfies (4), where $F$ is non-decreasing right-continuous function, $F(-\infty) = 0$, $F(+\infty) = R(0) < \infty$. Then $R$ is positive semidefinite, moreover, it is continuous. According to Theorem 6 there exists a stationary centered process with the autocovariance function $R$. Since $R$ is continuous (hence, continuous at zero), this process is mean square continuous which follows from Theorem 15.

2. Suppose that $R$ is the autocovariance function of a centered stationary mean square continuous process. Then, it is positive semidefinite and continuous at zero. For the proof that $R$ satisfies (4), see, e.g., Anděl (1976), IV.1, Theorem 2.

If the spectral distribution function in (4) is absolutely continuous, its derivative $f$ is called spectral density and (4) can be written in the form

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{R},$$

(5)

Remark 5. Two different stochastic processes may have the same spectral distribution functions and thus the same autocovariance functions.

6.3 Existence and computation of spectral density

Theorem 21. Let $K$ be a complex-valued function of an integer-valued variable $t \in \mathbb{Z}$, let $\sum_{t=-\infty}^{\infty} |K(t)| < \infty$. Then

$$K(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{Z},$$
where
\[
  f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-it\lambda} K(t), \quad \lambda \in [-\pi, \pi].
\]

**Proof.** Let us consider function \( K \), such that \( \sum_{t=-\infty}^{\infty} |K(t)| < \infty \). Since the series on the right-hand side of
\[
  f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-it\lambda} K(t)
\]
is absolutely convergent uniformly for \( \lambda \in [-\pi, \pi] \), we can interchange the integration and the summation and for any \( t \in \mathbb{Z} \) we get
\[
\begin{align*}
  \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda &= \int_{-\pi}^{\pi} e^{it\lambda} \left[ \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} K(k) \right] d\lambda \\
  &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[ K(k) \int_{-\pi}^{\pi} e^{i(t-k)\lambda} d\lambda \right] \\
  &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} K(k) 2\pi \delta(t-k) = K(t).
\end{align*}
\]

**Theorem 22.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a stationary sequence such that its autocovariance function \( R \) is absolutely summable, i.e. \( \sum_{t=-\infty}^{\infty} |R(t)| < \infty \). Then the spectral density of the sequence \( \{X_t, t \in \mathbb{Z}\} \) exists and for every \( \lambda \in [-\pi, \pi] \)
\[
  f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} R(k). \tag{6}
\]

**Proof.** \( \sum_{t=-\infty}^{\infty} |R(t)| < \infty \) \( \Rightarrow \)
\[
  R(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{Z},
\]
\[
  f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-it\lambda} R(t), \quad \lambda \in [-\pi, \pi]
\]

(from the previous theorem). Due to the uniqueness of the spectral decomposition (3) it suffices to prove, that \( f(\lambda) \geq 0 \) for every \( \lambda \in [-\pi, \pi] \).

For every \( \lambda \in [-\pi, \pi] \),
\[
  \varphi_n(\lambda) = \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} e^{-in\lambda} R(\kappa)(n - |\kappa|) \geq 0
\]

29
(it follows from the proof of Theorem 19). We show that \( f(\lambda) = \lim_{n \to \infty} \varphi_n(\lambda) \).

We have

\[
|f(\lambda) - \varphi_n(\lambda)| \leq \left| \frac{1}{2\pi} \sum_{|k| \geq n} e^{-ik\lambda} R(k) \right| \\
+ \left| \frac{1}{2\pi n} \sum_{\kappa = -n+1}^{n-1} e^{-i\kappa\lambda} R(\kappa) |\kappa| \right| \\
\leq \frac{1}{2\pi} \sum_{|k| \geq n} |R(k)| + \frac{1}{2\pi n} \sum_{\kappa = -n+1}^{n-1} |R(\kappa)| |\kappa| \to 0
\]

(Kronecker lemma: \( \sum_{k=1}^\infty a_k < \infty \Rightarrow \frac{1}{n} \sum_{k=1}^{n} k a_k \to 0 \text{ as } n \to \infty \).)

Formula (6) is called inverse formula for computing the spectral density of a stationary random sequence.

**Theorem 23.** Let \( \{X_t, t \in \mathbb{R}\} \) be a centered weakly stationary mean square process. Let its autocovariance function \( R \) satisfies condition \( \int_{-\infty}^{\infty} |R(t)| dt < \infty \). Then there exists spectral density of the process and it holds

\[
f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} R(t) dt, \quad \lambda \in (-\infty, \infty).
\] (7)

The proof is quite analogous to the computation of a probability density function by means of a characteristic function (Fourier transformation).

**Example 16.** (white noise) Let \( \{X_t, t \in \mathbb{Z}\} \) be a sequence of uncorrelated random variables with zero mean and a finite positive variance \( \sigma^2 \): \( EX_t = 0 \), \( \text{var} X_t = \sigma^2 \), \( \text{cov}(X_s, X_t) = \sigma^2 \delta(s-t) = R(s-t) \).

\[\sum_{t=-\infty}^{\infty} |R(t)| = \sigma^2 < \infty \Rightarrow \text{spectral density exists}\]

According to the inverse formula

\[
f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} R(k) = \frac{1}{2\pi} R(0) = \frac{\sigma^2}{2\pi}, \quad \lambda \in [-\pi, \pi].
\]

The spectral distribution function of the white noise is

\[
F(\lambda) = \begin{cases} 
0, & \lambda \leq -\pi, \\
\frac{\sigma^2}{2\pi} (\lambda + \pi), & \lambda \in [-\pi, \pi], \\
\sigma^2, & \lambda \geq \pi.
\end{cases}
\]

Notation: \( WN(0, \sigma^2) \) (white noise)
Example 17. Consider a stationary sequence with the autocovariance function $R(t) = a^{|t|}$, $t \in \mathbb{Z}$, $|a| < 1$.

$$\sum_{t=-\infty}^{\infty} |R(t)| = \sum_{t=-\infty}^{\infty} |a|^{|t|} = 1 + 2 \sum_{t=1}^{\infty} |a|^t < \infty,$$

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} a^{|k|}$$

$$= \frac{1}{2\pi} \sum_{k=0}^{\infty} e^{-ik\lambda} a^k + \frac{1}{2\pi} \sum_{k=0}^{\infty} e^{-ik\lambda} a^{-k}$$

$$= \frac{1}{2\pi} \sum_{k=0}^{\infty} (ae^{-i\lambda})^k + \frac{1}{2\pi} \sum_{k=1}^{\infty} (ae^{i\lambda})^k$$

$$= \frac{1}{2\pi} \frac{1}{1 - ae^{-i\lambda}} + \frac{1}{2\pi} \frac{1}{1 - ae^{i\lambda}}$$

$$= \frac{1}{2\pi} \frac{1 - a^2}{|1 - ae^{-i\lambda}|^2}$$

$$= \frac{1}{2\pi} \frac{1 - a^2}{1 - 2a \cos \lambda + a^2}$$

Example 18. Centered weakly stationary process with the autocovariance function $R(t) = ce^{-\alpha|t|}$, $t \in \mathbb{R}$, $c > 0$, $\alpha > 0$.

The process is mean square continuous. It holds

$$\int_{-\infty}^{\infty} |R(t)| dt = \int_{-\infty}^{\infty} ce^{-\alpha|t|} dt < \infty,$$

thus, the spectral density exists and by formula (7)

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} R(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} ce^{-\alpha|t|} dt$$

$$= \frac{c}{2\pi} \int_{-\infty}^{\infty} (\cos \lambda t - i \sin \lambda t) e^{-\alpha|t|} dt$$

$$= \frac{c}{\pi} \int_{0}^{\infty} \cos(\lambda t) e^{-\alpha t} dt = \frac{ca}{\pi} \frac{1}{\alpha^2 + \lambda^2}$$

for every $\lambda \in \mathbb{R}$.

Example 19. A centered mean square process with the spectral distribution function

$$F(\lambda) = \begin{cases} 0, & \lambda < -1, \\ \frac{1}{2}, & -1 \leq \lambda < 1, \\ 1, & \lambda \geq 1. \end{cases}$$
Figure 3: Trajectories of a process with the autocovariance function $R(t) = a^{|t|}$, up: $a = 0.8$, down $a = -0.8$.

Spectral distribution function is not absolutely continuous; the spectral density of the process does not exist. According to (4) the autocovariance function is

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda) = \frac{1}{2} e^{-it} + \frac{1}{2} e^{it} = \cos t, \quad t \in \mathbb{R}.$$  

The process has a discrete spectrum with non-zero value at frequencies $\lambda_1 = -1, \lambda_2 = 1$.

**Example 20.** The process $\{X_t, t \in \mathbb{R}\}$ of uncorrelated random variables with zero mean and a finite positive variance does not satisfy decomposition (4), since it is not mean square continuous.

### 7 Spectral representation of stochastic processes

#### 7.1 Orthogonal increment processes

**Definition 30.** Let $\{X_t, t \in T\}$, $T$ an interval, be a (generally complex-valued) second order process on $(\Omega, \mathcal{A}, P)$. We say that $\{X_t, t \in T\}$ is orthogonal increment process, if for any $t_1, \ldots, t_4 \in T$ such that, $\emptyset \neq (t_1, t_2) \cap (t_3, t_4) = \emptyset$,

$$\mathbb{E}(X_{t_2} - X_{t_1})(\overline{X}_{t_4} - \overline{X}_{t_3}) = 0$$
Figure 4: Autocovariance function $R(t) = a^{|t|}$ (left) and the spectral density (right), $a = 0.8$

Figure 5: Autocovariance function $R(t) = a^{|t|}$ (left) and the spectral density (right), $a = -0.8$

Figure 6: Autocovariance function $R(t) = ce^{-a|t|}, t \in \mathbb{R}$ (left) and the spectral density (right), $c = 1, \alpha = 1$
In what follows we will consider only centered right-mean square continuous processes i.e., such that \( \mathbb{E}|X_t - X_{t_0}|^2 \to 0 \) as \( t \to t_0^+ \) for any \( t_0 \in T \).

**Theorem 24.** Let \( \{Z_{\lambda}, \, \lambda \in [a, b]\} \) be a centered orthogonal increment right-mean square continuous process, \( [a, b] \) a bounded interval. Then there exists a unique non-decreasing right-continuous function \( F \) such that

\[
F(\lambda) = 0, \quad \lambda \leq a,
\]
\[
= F(b), \quad \lambda \geq b,
\]
\[
F(\lambda_2) - F(\lambda_1) = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2, \quad a \leq \lambda_1 < \lambda_2 \leq b.
\]

**Proof.** Define function

\[
F(\lambda) = \mathbb{E}|Z_{\lambda} - Z_a|^2, \quad \lambda \in [a, b]
\]
\[
= 0, \quad \lambda \leq a,
\]
\[
= F(b), \quad \lambda \geq b.
\]

We show that this function is non-decreasing, right-continuous and satisfies the condition of the theorem. Obviously, it suffices to consider \( \lambda \in [a, b] \), only.

Let \( a < \lambda_1 < \lambda_2 < b \). Then

\[
F(\lambda_2) = \mathbb{E}|Z_{\lambda_2} - Z_a|^2 = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1} + Z_{\lambda_1} - Z_a|^2
\]
\[
= \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 + \mathbb{E}|Z_{\lambda_1} - Z_a|^2
\]
\[
+ \mathbb{E}(Z_{\lambda_2} - Z_{\lambda_1})(Z_{\lambda_1} - Z_a) + \mathbb{E}(Z_{\lambda_1} - Z_a)(Z_{\lambda_1} - Z_{\lambda_2})
\]
\[
= \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 + F(\lambda_1),
\]

(increments \( Z_{\lambda_2} - Z_{\lambda_1} \) and \( Z_{\lambda_1} - Z_a \) are orthogonal). From here

\[
F(\lambda_2) - F(\lambda_1) = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 \geq 0,
\]

\( \Rightarrow \) \( F \) is non-decreasing right-continuous due to the right continuity of the process \( \{Z_{\lambda}, \, \lambda \in [a, b]\} \). Condition (8) is satisfied.

Uniqueness of function \( F \): Let \( G \) be a non-decreasing right-continuous function that satisfies conditions of the theorem. Then \( G(a) = 0 = F(a) \) and for \( \lambda \in (a, b] \) it holds \( G(\lambda) = G(\lambda) - G(a) = \mathbb{E}|Z_{\lambda} - Z_a|^2 = F(\lambda) - F(a) = F(\lambda) \).

The function \( F \) is bounded, non-decreasing, right-continuous, and we call it *distribution function associated with the orthogonal increment process.*
Example 21. Wiener process on $[0, T]$ is a centered mean square continuous Gaussian process with independent and stationary increments $\Rightarrow$ with orthogonal increments, $W_0 = 0$, $W_s - W_t \sim \mathcal{N}(0, \sigma^2|s - t|), 0 \leq s, t, \leq T$. The associated distribution function on $[0, T]$ is

$$F(\lambda) = 0, \quad \lambda \leq 0$$

$$= \mathbb{E}[W_\lambda - W_0]^2 = \sigma^2 \lambda, \quad 0 \leq \lambda \leq T$$

$$= \sigma^2 T, \quad \lambda \geq T.$$

Example 22. Let $\tilde{W}_\lambda$ be a transformation of the Wiener process on the interval $[-\pi, \pi]$ given by $\tilde{W}_\lambda = W_\lambda(\lambda + \pi)/2\pi, \lambda \in [-\pi, \pi]$.

The process $\{\tilde{W}_\lambda, \lambda \in [\pi, \pi]\}$ is a Gaussian process with orthogonal increments and the associated distribution function

$$F(\lambda) = 0, \quad \lambda \leq -\pi,$$

$$= \frac{\sigma^2}{2\pi}(\lambda + \pi), \quad \lambda \in [-\pi, \pi],$$

$$= \sigma^2, \quad \lambda \geq \pi.$$

7.2 Integral with respect to an orthogonal increment process

Let $\{Z_\lambda, \lambda \in [a, b]\}$ be a centered right-mean square continuous process with orthogonal increments on $(\Omega, \mathcal{A}, P)$, $[a, b]$ a bounded interval, let $F$ be the associated distribution function of this process. Let $\mu_F$ be a measure induced by $F$.

Consider a space of complex-valued functions $L_2([a, b], \mathcal{B}, \mu_F) := L_2(F)$, i.e. space of measurable functions $f$ on $[a, b]$ such that

$$\int_a^b |f(\lambda)|^2 d\mu_F(\lambda) = \int_a^b |f(\lambda)|^2 dF(\lambda) < \infty.$$

Properties of $L_2(F)$:

- The inner product on the space of functions $L_2(F)$ (more exactly, on equivalence classes of $L_2(F)$ with respect to measure $\mu_F$) $^1$

$$\langle f, g \rangle = \int_a^b f(\lambda)\overline{g(\lambda)}dF(\lambda), \quad f, g \in L_2(F)$$

$^1f \sim g$ if $f = g \mu_F$ almost everywhere

35
• Norm in $L_2(F)$:
  \[ \|f\| = \left[ \int_a^b |f(\lambda)|^2 dF(\lambda) \right]^{\frac{1}{2}}. \]

• convergence in $L_2(F)$:
  \[ f_n \to f \text{ in } L_2(F) \text{ as } n \to \infty, \text{ if } \|f_n - f\| \to 0, \text{ or} \]
  \[ \int_a^b |f_n(\lambda) - f(\lambda)|^2 dF(\lambda) \to 0, \quad n \to \infty. \]

• space $L_2(F)$ is complete. (Rudin (2003), Theorem 3.11).

Definition of the integral
I. Let $f \in L_2(F)$ be a simple function, i.e.,
  \[ f(\lambda) = \sum_{k=1}^n c_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda), \]  
where $J_A(y) = 1$ for $y \in A$ and $J_A(y) = 0$ otherwise is the indicator function of a set $A$,
c_1, \ldots, c_n$ are complex-valued constants, $c_k \neq c_{k+1}$, $1 \leq k \leq n - 1$. We define
  \[ \int_{[a,b]} f(\lambda) dZ(\lambda) := \sum_{k=1}^n c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}). \]
which is a random variable from the space $L_2(\Omega, \mathcal{A}, P)$.

Convention: Instead of $\int_{[a,b]} f(\lambda) dZ(\lambda)$ we will write $\int_a^b f(\lambda) dZ(\lambda)$.

Notation: $\int_a^b f(\lambda) dZ(\lambda) := I(f)$.

Properties of the integral for simple functions:

Theorem 25. Let \{\(Z_\lambda, \lambda \in [a,b]\)\} be a centered right-continuous process with orthogonal
increments and the associated distribution function $F$, let $f, g$ be simple functions in $L_2(F)$, $\alpha, \beta$ complex-valued constants. Then

1. $E \int_a^b f(\lambda) dZ(\lambda) = 0$.

2. $\int_a^b [\alpha f(\lambda) + \beta g(\lambda)] dZ(\lambda) = \alpha \int_a^b f(\lambda) dZ(\lambda) + \beta \int_a^b g(\lambda) dZ(\lambda)$.

3. $E \int_a^b f(\lambda) dZ(\lambda) \int_a^b g(\lambda) dZ(\lambda) = \int_a^b f(\lambda) g(\lambda) dF(\lambda)$. 

36
Proof. 1. Let $f(\lambda) = \sum_{k=1}^{n} c_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda)$.

Then
\[
E \int_{a}^{b} f(\lambda) dZ(\lambda) = E \left[ \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) \right]
\]
\[
= \sum_{k=1}^{n} c_k E(Z_{\lambda_k} - Z_{\lambda_{k-1}}) = 0,
\]

since $\{Z_\lambda, \lambda \in [a, b]\}$ is centered.

Wlog: Let
\[
f(\lambda) = \sum_{k=1}^{n} c_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda)
\]
\[
g(\lambda) = \sum_{k=1}^{n} d_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda).
\]

Then
\[
\int_{a}^{b} [\alpha f(\lambda) + \beta g(\lambda)] dZ(\lambda) = \sum_{k=1}^{n} (\alpha c_k + \beta d_k)(Z_{\lambda_k} - Z_{\lambda_{k-1}})
\]
\[
= \alpha \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) + \beta \sum_{k=1}^{n} d_k (Z_{\lambda_k} - Z_{\lambda_{k-1}})
\]
\[
= \alpha \int_{a}^{b} f(\lambda) dZ(\lambda) + \beta \int_{a}^{b} g(\lambda) dZ(\lambda).
\]

3. Let
\[
f(\lambda) = \sum_{k=1}^{n} c_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda)
\]
\[
g(\lambda) = \sum_{k=1}^{n} d_k J_{[\lambda_{k-1}, \lambda_k]}(\lambda).
\]
Then
\[
E \int_a^b f(\lambda) d\lambda \int_a^b g(\lambda) d\lambda = E \sum_{k=1}^n c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) \sum_{k=1}^n d_k (Z_{\lambda_k} - Z_{\lambda_{k-1}})
\]
\[
= \sum_{k=1}^n c_k d_k \|Z_{\lambda_k} - Z_{\lambda_{k-1}}\|^2
\]
\[
= \sum_{k=1}^n c_k d_k \big( F(\lambda_k) - F(\lambda_{k-1}) \big) = \int_a^b f(\lambda) g(\lambda) dF(\lambda).
\]

\[\square\]

II. Let \( f \in L_2(F) \) be a measurable function. The set of simple functions is dense in \( L_2(F) \) and its closure is \( L_2(F) \) (Rudin (2003), Theorem 3.13) \( \Rightarrow \) there exists a sequence of simple functions \( f_n \in L_2(F) \) such that \( f_n \to f \) in \( L_2(F) \) as \( n \to \infty \).

Integral \( I(f_n) \) is defined for simple functions and \( I(f_n) \in L_2(\Omega, \mathcal{A}, P) \). The sequence \( \{I(f_n)\} \) is a Cauchy sequence in \( L_2(\Omega, \mathcal{A}, P) \):

\[
E|I(f_m) - I(f_n)|^2 = E(I(f_m) - I(f_n))(I(f_m) - I(f_n))
\]
\[
= E \int_a^b (f_m(\lambda) - f_n(\lambda)) d\lambda \int_a^b (f_m(\lambda) - f_n(\lambda)) d\lambda
\]
\[
= \int_a^b (f_m(\lambda) - f_n(\lambda))(f_m(\lambda) - f_n(\lambda)) dF(\lambda)
\]
\[
= \int_a^b |f_m(\lambda) - f_n(\lambda)|^2 dF(\lambda) \to 0
\]
as \( m, n \to \infty \), since \( f_n \to f \) in \( L_2(F) \).

Since \( \{I(f_n)\} \) is the Cauchy sequence in \( L_2(\Omega, \mathcal{A}, P) \), it has a mean square limit

\[
I(f) = \lim_{n \to \infty} I(f_n) := \int_a^b f(\lambda) d\lambda,
\]

which is called to be integral of the function \( f \) with respect to the process with orthogonal increments, or stochastic integral.

\( I(f) \) does not depend on the choice of the sequence \( \{f_n\} \): Let \( f \in L_2(F) \) and \( f_n \) a \( g_n \) be simple, \( f_n \to f \) and \( g_n \to f \) in \( L_2(F) \). Then \( I(f_n), I(g_n) \) have mean square limits \( I, J \), respectively.

38
Define sequence \( \{h_n\} = \{f_1, g_1, f_2, g_2, \ldots \} \) which is simple and \( h_n \to f \) in \( L_2(F) \). Then \( I(h_n) \to K \) in mean square. Since selected subsequences \( \{I(f_n)\} \) and \( \{I(g_n)\} \) have mean square limits, \( I \equiv J \equiv K \).

**Theorem 26.** Let \( \{Z_\lambda, \lambda \in [a, b]\} \) be a centered right-mean square continuous process with orthogonal increments and the associated distribution function \( F \). Then integral (11) has the following properties.

1. Let \( f \in L_2(F) \). Then \( \mathbb{E} I(f) = \mathbb{E} \int_a^b f(\lambda) dZ(\lambda) = 0 \).

2. Let \( f, g \in L_2(F) \), \( \alpha, \beta \in \mathbb{C} \) be constants. Then \( I(\alpha f + \beta g) = \alpha I(f) + \beta I(g) \).

3. Let \( f, g \in L_2(F) \). Then

\[
\mathbb{E} I(f) I(g) = \int_a^b f(\lambda) g(\lambda) dF(\lambda). \tag{12}
\]

4. Let \( \{f_n, n \in \mathbb{N}\} \) and \( f \) be functions in \( L_2(F) \), respectively. Then as \( n \to \infty \)

\[
f_n \to f \text{ in } L_2(F) \iff I(f_n) \to I(f) \text{ in } L_2(\Omega, \mathcal{A}, P). \tag{13}
\]

**Proof.**

1. Let \( f \in L_2(F) \) and \( \{f_n, n \in \mathbb{N}\} \) be a sequence of simple functions in \( L_2(F) \) such that \( f_n \to f \) in \( L_2(F) \). Then \( I(f) = 1 \) i.m. \( I(f_n) \). Since \( \mathbb{E} I(f_n) = 0 \), then also \( \mathbb{E} I(f) = 0 \) (from the properties of the mean square convergence).

2. Let \( f, g \in L_2(F) \) and \( \{f_n, n \in \mathbb{N}\} \), respectively \( \{g_n, n \in \mathbb{N}\} \) be sequences of simple functions in \( L_2(F) \) such that \( f_n \to f \) and \( g_n \to g \) in \( L_2(F) \) respectively; thus, \( I(f_n) \to I(f) \) and \( I(g_n) \to I(g) \) in \( L_2(\Omega, \mathcal{A}, P) \) (in mean square).

The sequence of simple functions \( h_n = \alpha f_n + \beta g_n \) converges to \( h = \alpha f + \beta g \) in \( L_2(F) \), since

\[
\int_a^b |\alpha f_n(\lambda) + \beta g_n(\lambda) - (\alpha f(\lambda) + \beta g(\lambda))|^2 dF(\lambda)
\]

\[
\leq 2|\alpha|^2 \int_a^b |f_n(\lambda) - f(\lambda)|^2 dF(\lambda)
\]

\[
+ 2|\beta|^2 \int_a^b |g_n(\lambda) - g(\lambda)|^2 dF(\lambda) \to 0
\]

We have:

- \( h_n = \alpha f_n + \beta g_n \) simple
- \( I(h_n) = I(\alpha f_n + \beta g_n) = \alpha I(f_n) + \beta I(g_n) \)
• $h_n \to h$ in $L_2(F) \Rightarrow I(h_n) \to I(h)$ in mean square

• $h = \alpha f + \beta g$

• $I(h_n) \to \alpha I(f) + \beta I(g)$ in mean square, since

\[
\begin{align*}
\mathbb{E}|\alpha I(f_n) + \beta I(g_n) - (\alpha I(f) + \beta I(g))|^2 \\
&= \mathbb{E}|\alpha(I(f_n) - I(f)) + \beta(I(g_n) - I(g))|^2 \\
&\leq 2|\alpha|^2\mathbb{E}|I(f_n) - I(f)|^2 + 2|\beta|^2\mathbb{E}|I(g_n) - I(g)|^2 \to 0
\end{align*}
\]

$\Rightarrow I(h) = I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$.

3. Let $f, g \in L_2(F)$, \( \{f_n, n \in \mathbb{N}\} \) and \( \{g_n, n \in \mathbb{N}\} \) be sequences of simple functions, $f_n \to f$ and $g_n \to g$ in $L_2(F)$. Thus, $I(f_n) \to I(f)$ and $I(g_n) \to I(g)$ in mean square.

From the continuity of the inner product in $L_2(\Omega, \mathcal{A}, P)$:

$$\mathbb{E}I(f_n) \overline{I(g_n)} = \langle I(f_n), I(g_n) \rangle \to \langle I(f), I(g) \rangle = \mathbb{E}I(f) \overline{I(g)}.$$  

From the continuity of the inner product in $L_2(F)$:

$$\mathbb{E}I(f_n) \overline{I(g_n)} = \int_a^b f_n(\lambda) \overline{g_n(\lambda)} dF(\lambda) = \langle f_n, g_n \rangle \to \langle f, g \rangle$$

$$= \int_a^b f(\lambda) \overline{g(\lambda)} dF(\lambda),$$

from here (12) follows.

4. Let $f_n, f \in L_2(F)$. According to 2 and 3,

$$\mathbb{E}|I(f_n) - I(f)|^2 = \mathbb{E}|I(f_n - f)|^2 = \int_a^b |f_n(\lambda) - f(\lambda)|^2 dF(\lambda),$$

from which (13) follows. 

\[\square\]

**Remark 6.** let \( \{Z_\lambda, \lambda \in \mathbb{R}\} \) be a centered right-continuous process with orthogonal increments. Function $F$ defined by

$$F(\lambda_2) - F(\lambda_1) = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2, \quad -\infty < \lambda_1 < \lambda_2 < \infty$$

is non-decreasing, right-continuous and unique (up to an additive constant). If $F$ is bounded it induces a finite measure $\mu_F$, and for $f$ such that

$$\int_{-\infty}^{\infty} |f(\lambda)|^2 d\mu_F(\lambda) = \int_{-\infty}^{\infty} |f(\lambda)|^2 dF(\lambda) < \infty,$$

$$\int_{-\infty}^{\infty} f(\lambda) d\mu_F(\lambda) := 1. i. m. \int_a^b f(\lambda) d\mu_F(\lambda)$$

as $a \to -\infty, b \to \infty$.  

40
7.3 Spectral decomposition of a stochastic process

Theorem 27. let $X_t$, $t \in \mathbb{Z}$, be random variables such that

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda),$$

where $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ is a centered right-continuous process with orthogonal increments on $[-\pi, \pi]$ and associated distribution function $F$. Then $\{X_t, t \in \mathbb{Z}\}$ is a centered weakly stationary sequence with the spectral distribution function $F$.

Proof. The associated distribution function $F$ of the process $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ is bounded, non-decreasing and right-continuous, $F(\lambda) = 0$ for $\lambda \leq -\pi$, $F(\lambda) = F(\pi) < \infty$ for $\lambda \geq \pi$. For $t \in \mathbb{Z}$ define function $e_t$:

$$e_t(\lambda) = e^{it\lambda}, -\pi \leq \lambda \leq \pi.$$ 

Then

$$\int_{-\pi}^{\pi} |e_t(\lambda)|^2 dF(\lambda) = \int_{-\pi}^{\pi} |e^{it\lambda}|^2 dF(\lambda) = F(\pi) - F(-\pi) < \infty,$$

$\Rightarrow e_t \in L_2(F)$ and $X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)$ is well defined random variable. According to Theorem 26

1. $E X_t = E \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) = 0$ for any $t \in \mathbb{Z}$

2. 

$$E|X_t|^2 = E \left| \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) \right|^2 = E \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) = \int_{-\pi}^{\pi} |e^{it\lambda}|^2 dF(\lambda) = \int_{-\pi}^{\pi} dF(\lambda) < \infty$$

3. 

$$\text{cov}(X_{t+h}, X_t) = E \int_{-\pi}^{\pi} e^{i(t+h)\lambda} dZ(\lambda) \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda) := R(h).$$

We have:
• \( \text{cov}(X_{t+h}, X_t) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda) \) depends on \( h \) only
• sequence \( \{X_t, t \in \mathbb{Z}\} \) is centered and weakly stationary
• function \( F \) has the same properties as the spectral distribution function (see spectral decomposition of the autocovariance function)
• from the uniqueness of the spectral decomposition it follows that \( F \) is the spectral distribution function of the sequence \( \{X_t, t \in \mathbb{Z}\} \).

\[ \square \]

**Example 23.** Let \( \widetilde{W}_\lambda \) be a transformation of the Wiener process to the interval \([-\pi, \pi]\) given by \( \widetilde{W}_\lambda = W_{(\lambda+\pi)/2\pi}, \lambda \in [-\pi, \pi] \). Then random variables
\[ X_t = \int_{-\pi}^{\pi} e^{it\lambda} d\tilde{W}(\lambda), \quad t \in \mathbb{Z} \]
are centered, uncorrelated with the same variance \( \sigma^2 \); they have white noise property.

**Example 24.** Consider a sequence of functions \( \{f_t, t \in \mathbb{Z}\} \) on \([-\pi, \pi]\)
\[ f_{tk}(\lambda) = \sum_{j=1}^{k} e^{it\lambda_j} J(\lambda_j - \lambda_{j-1})(\lambda), \]
where \(-\pi = \lambda_0 < \lambda_1 < \cdots < \lambda_k = \pi\). Let \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) be a centered right-continuous process with orthogonal increments on \([-\pi, \pi]\) with associated distribution function \( F \).

\( f_{tk} \) are simple functions in \( L_2(F) \), thus
\[ X_{tk} := \int_{-\pi}^{\pi} f_{tk}(\lambda)dZ(\lambda) = \sum_{j=1}^{k} e^{it\lambda_j} (Z_{\lambda_j} - Z_{\lambda_{j-1}}) = \sum_{j=1}^{k} e^{it\lambda_j} \tilde{Z}_j \]
\( \tilde{Z}_j = Z_{\lambda_j} - Z_{\lambda_{j-1}}, \ j = 1, \ldots, k, \) are uncorrelated random variables with zero mean and the variance \( E[\tilde{Z}_j]^2 = F(\lambda_j) - F(\lambda_{j-1}) := \sigma_j^2 \).

Then \( EX_{tk} = 0, E X_{(t+h),k} X_{tk} = \sum_{j=1}^{k} e^{ih\lambda_j} \sigma_j^2 := R_k(h) \)
\[ R_k(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF_{Xk}(\lambda) \]
\( F_{Xk} \) - spectral distribution function with jumps at points \( \lambda_j \)
\( F_{Xk}(\lambda_j) - F_{Xk}(\lambda_{j-1}) = \sigma_j^2 = F(\lambda_j) - F(\lambda_{j-1}). \) Since \( F(-\pi) = F_{Xk}(-\pi) = 0, F \) equals to \( F_{Xk} \) at points \( \lambda_j, j = 0, 1, \ldots, k \).

Then it is possible to consider limits as \( k \to \infty \) a max \( \max_{1 \leq j \leq k} |\lambda_j - \lambda_{j-1}| \to 0. \)
Theorem 28. Let \( \{X_t, t \in \mathbb{Z}\} \) be a centered weakly stationary sequence with spectral distribution function \( F \). Then there exists a centered orthogonal increment process \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) such that
\[
X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda), \quad t \in \mathbb{Z}
\]
and
\[
E|Z(\lambda) - Z(-\pi)|^2 = F(\lambda), \quad -\pi \leq \lambda \leq \pi.
\]
Relation (14) is called spectral decomposition of a stationary random sequence.


Meaning of Theorem 28: Any random variable of a centered stationary random sequence can be approximated (in the mean square limit) by a sum \( \sum e^{it\lambda_j} Y_j \) of uncorrelated random variables \( Y_j \), the variance of which is an increment of the spectral distribution function at points (frequencies) \( \lambda_{j-1} \) and \( \lambda_j \).

Theorem 29. Let \( \{X_t, t \in \mathbb{R}\} \) be a centered weakly stationary mean square continuous process. Then there exists an orthogonal increment process \( \{Z_\lambda, \lambda \in \mathbb{R}\} \) such that
\[
X_t = \int_{-\infty}^{\infty} e^{it\lambda} dZ(\lambda), \quad t \in \mathbb{R},
\]
and the associated distribution function of the process \( \{Z_\lambda, \lambda \in \mathbb{R}\} \) is the spectral distribution function of the process \( \{X_t, t \in \mathbb{R}\} \).

Relation (15) is said to be spectral decomposition of a stationary mean square continuous process.


Theorem 30. Let \( \{X_t, t \in \mathbb{Z}\} \) be a centered stationary sequence with a spectral distribution function \( F \). Let \( \mathcal{H}\{X_t, t \in \mathbb{Z}\} \) be the Hilbert space generated by \( \{X_t, t \in \mathbb{Z}\} \). Then \( U \in \mathcal{H}\{X_t, t \in \mathbb{Z}\} \) if and only if
\[
U = \int_{-\pi}^{\pi} \varphi(\lambda) dZ(\lambda),
\]
where \( \varphi \in L_2(F) \) and \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is the orthogonal increments process as given in the spectral decomposition of the sequence \( \{X_t, t \in \mathbb{Z}\} \).
Proof. 1. Let \( U \in \mathcal{H}\{X_t, t \in \mathbb{Z}\} \). Then either \( U \in \mathcal{M}\{X_t, t \in \mathbb{Z}\} \) (linear span), or \( U = \lim_{n \to \infty} U_n, \ U_n \in \mathcal{M}\{X_t, t \in \mathbb{Z}\} \).

a) Let \( U \in \mathcal{M}\{X_t, t \in \mathbb{Z}\} \); then \( U = \sum_{j=1}^{N} c_j X_{t_j} \), for \( c_1, \ldots, c_N \in \mathbb{C} \) and \( t_1, \ldots, t_N \in \mathbb{Z} \). From the spectral decomposition (14)

\[
U = \sum_{j=1}^{N} c_j X_{t_j} = \sum_{j=1}^{N} c_j \int_{-\pi}^{\pi} e^{it_j \lambda} d\zeta(\lambda) = \int_{-\pi}^{\pi} \left[ \sum_{j=1}^{N} c_j e^{it_j \lambda} \right] d\zeta(\lambda) = \int_{-\pi}^{\pi} \varphi(\lambda) d\zeta(\lambda),
\]

where \( \varphi(\lambda) = \sum_{j=1}^{N} c_j e^{it_j \lambda} \).

Obviously, \( \varphi \) is a finite linear combination of functions from \( L_{2}(F) \), thus \( \varphi \in L_{2}(F) \).

b) Let \( U = \lim_{n \to \infty} U_n, \ U_n \in \mathcal{M}\{X_t, t \in \mathbb{Z}\}. \) According to a)

\[
U_n = \int_{-\pi}^{\pi} \varphi_n(\lambda) d\zeta(\lambda), \ \varphi_n \in L_{2}(F).
\]

\( \{U_n\} \) is the Cauchy sequence in \( \mathcal{H}\{X_t, t \in \mathbb{Z}\} \) (it is convergent there) thus \( \{\varphi_n\} \) is the Cauchy sequence in \( L_{2}(F) \), since

\[
E|U_m - U_n|^2 = E\left| \int_{-\pi}^{\pi} \varphi_m(\lambda) d\zeta(\lambda) - \int_{-\pi}^{\pi} \varphi_n(\lambda) d\zeta(\lambda) \right|^2 = E\left| \int_{-\pi}^{\pi} [\varphi_m(\lambda) - \varphi_n(\lambda)] d\zeta(\lambda) \right|^2 = \int_{-\pi}^{\pi} |\varphi_m(\lambda) - \varphi_n(\lambda)|^2 d\zeta(\lambda).
\]

Thus, there exists \( \varphi \in L_{2}(F) \) such that \( \varphi_n \to \varphi \) \( \text{v} \ L_{2}(F) \). By (13)

\[
U_n = \int_{-\pi}^{\pi} \varphi_n(\lambda) d\zeta(\lambda) \to \int_{-\pi}^{\pi} \varphi(\lambda) d\zeta(\lambda) \ (\text{v} \ L_{2}(\Omega, A, P)),
\]

thus \( U = \int_{-\pi}^{\pi} \varphi(\lambda) d\zeta(\lambda) \).

2. Let \( \hat{U} \) be a random variable that satisfies (16). Since \( \varphi \in L_{2}(F) \), there is a sequence of trigonometric polynomials \( \varphi_n(\lambda) = \sum_{k=-n}^{n} c_k^{(n)} e^{i\lambda k} \) on \( [-\pi, \pi] \) such that \( \varphi_n \to \varphi \) \( \text{v} \ L_{2}(F) \). According to (13)

\[
\int_{-\pi}^{\pi} \varphi(\lambda) d\zeta(\lambda) = \lim_{n \to \infty} \int_{-\pi}^{\pi} \varphi_n(\lambda) d\zeta(\lambda),
\]

44
hence

\[ U = \int_{-\pi}^{\pi} \varphi(\lambda) dZ(\lambda) = 1. \text{ i. m.} \int_{-\pi}^{\pi} \left[ \sum_{k=-n}^{n} c_k^{(n)} e^{i\lambda k} \right] dZ(\lambda) \]

\[ = 1. \text{ i. m.} \sum_{k=-n}^{n} c_k^{(n)} \left[ \int_{-\pi}^{\pi} e^{i\lambda k} dZ(\lambda) \right] \]

\[ = 1. \text{ i. m.} \sum_{k=-n}^{n} c_k^{(n)} X_k^{(n)} \in \mathcal{H}\{X_t, t \in \mathbb{Z}\}. \]

\[ \square \]

8 Linear models of time series

8.1 White noise

Recall that a white noise sequence \( WN(0, \sigma^2) \) is to be defined as a sequence \( \{Y_t, t \in \mathbb{Z}\} \) of uncorrelated random variables with mean zero and variance \( 0 < \sigma^2 < \infty \) and summarize its basic properties:

- autocovariance function: \( R_Y(t) = \sigma^2 \delta(t), \quad t \in \mathbb{Z} \)
- spectral density: \( f_Y(\lambda) = \frac{\sigma^2}{2\pi}, \quad \lambda \in [-\pi, \pi] \)
- spectral decomposition:

\[ Y_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_Y(\lambda), \]

\( Z_Y = \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is a process with orthogonal increments and associated distribution function

\[ F(\lambda) = \frac{\sigma^2}{2\pi}(\lambda + \pi), \quad \lambda \in [-\pi, \pi] \]

(same as the spectral distribution function \( F_Y(\lambda) \))

8.2 Moving average sequences

**Definition 31.** A random sequence \( \{X_t, t \in \mathbb{Z}\} \) defined by

\[ X_t = b_0 Y_t + b_1 Y_{t-1} + \cdots + b_n Y_{t-n}, \quad t \in \mathbb{Z}, \quad (17) \]

where \( \{Y_t, t \in \mathbb{Z}\} \) is a white noise \( WN(0, \sigma^2) \) and \( b_0, b_1, \ldots, b_n \) are real- or complex-valued constants, \( b_0 \neq 0, b_n \neq 0 \), is called to be a **moving average sequence of order** \( n \). Notation: MA\((n)\), special case: \( b_i = \frac{1}{n+1}, i = 0, \ldots, n. \)
The mean and the autocovariance function (in time domain):
1. $EX_t = 0$ for all $t \in \mathbb{Z}$.
2. For $t \geq 0$

$$cov(X_{s+t}, X_s) = EX_{s+t}X_s = E \left( \sum_{j=0}^{n} b_j Y_{s+t-j} \sum_{k=0}^{n} b_k Y_{s-k} \right)$$

$$= \sum_{j=0}^{n} \sum_{k=0}^{n} b_j b_k E(Y_{s+t-j}Y_{s-k})$$

$$= \sigma^2 \sum_{j=0}^{n} \sum_{k=0}^{n} b_j b_k \delta(t - j + k)$$

$$= \sigma^2 \sum_{k=0}^{n-t} b_{t+k} b_k, \quad 0 \leq t \leq n$$

$$= 0, \quad t > n.$$

For $t \leq 0$ proceed analogously.

Since $cov(X_{s+t}, X_s)$ depends on $t$, only, the sequence is weakly stationary.

Computation in spectral domain:

$$X_t = \sum_{j=0}^{n} b_j Y_{t-j} = \sum_{j=0}^{n} b_j \left[ \int_{-\pi}^{\pi} e^{i(t-j)\lambda} dZ_Y(\lambda) \right]$$

$$= \int_{-\pi}^{\pi} \left[ \sum_{j=0}^{n} b_j e^{i(t-j)\lambda} \right] dZ_Y(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} \left[ \sum_{j=0}^{n} b_j e^{-ij\lambda} \right] dZ_Y(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ_Y(\lambda),$$

where $g(\lambda) = \sum_{j=0}^{n} b_j e^{-ij\lambda} \in L_2(F)$.

46
From the properties of stochastic integral: $E X_t = 0,$

$$EX_{s+t}X_s = E \int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) dZ_Y(\lambda) \int_{-\pi}^{\pi} e^{i\lambda\lambda} g(\lambda) dZ_Y(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) e^{-i\lambda\lambda} g(\lambda) dF_Y(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 f_Y(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 \frac{\sigma^2}{2\pi} d\lambda = R_X(t),$$

is again a function of $t$. Due to the uniqueness of the spectral decomposition of an autocovariance function it follows that function $|g(\lambda)|^2 \frac{\sigma^2}{2\pi}$ is the spectral density of the sequence $\{X_t, t \in \mathbb{Z}\}$ as given by (17).

We have proved the following theorem.

**Theorem 31.** The moving average sequence $\{X_t, t \in \mathbb{Z}\}$ of order $n$ defined by (17) is centered and weakly stationary, with the autocovariance function

$$R_X(t) = \sigma^2 \sum_{k=0}^{n-t} b_{k+t} \overline{b_k}, \quad 0 \leq t \leq n, \quad (18)$$

$$= R_X(-t), \quad -n \leq t \leq 0,$$

$$= 0, \quad |t| > n.$$

The spectral density $f_X$ of sequence (17) exists and it holds

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{n} b_k e^{-ik\lambda} \right|^2, \quad \lambda \in [-\pi, \pi]. \quad (19)$$

**Remark 7.** For real-valued constants $b_0, \ldots, b_n$ the autocovariance function of the sequence MA($n$) takes form

$$R_X(t) = \sigma^2 \sum_{k=0}^{n-|t|} b_k b_{k+|t|}, \quad |t| \leq n, \quad (20)$$

$$= 0, \quad |t| > n.$$
Figure 5.1: Autokorelační funkce (vlevo) a spektrální hustota (vpravo) posloupnosti MA(1): \[ X_t = Y_t + Y_{t-1}, \]
\[ Y_t \sim N(0, 1) \]

Figure 5.2: Autokorelační funkce (vlevo) a spektrální hustota (vpravo) posloupnosti MA(2): \[ X_t = Y_t + Y_{t-1} - 2Y_{t-2}; \]
\[ Y_t \sim N(0, 1) \]

8.3 Linear process

**Theorem 32.** Let \( \{Y_t, t \in \mathbb{Z}\} \) be a white noise \( WN(0, \sigma^2) \) and \( \{c_j, j \in \mathbb{N}_0\} \) be a sequence of complex-valued constants.

1. If \( \sum_{j=0}^{\infty} |c_j|^2 < \infty \), the series \( \sum_{j=0}^{\infty} c_j Y_{t-j} \) converges in mean square for every \( t \in \mathbb{Z} \), i.e., for every \( t \in \mathbb{Z} \) there exists a random variable \( X_t \) such that

\[
X_t = \text{l. i. m.} \sum_{j=0}^{n} c_j Y_{t-j}.
\]

2. If \( \sum_{j=0}^{\infty} |c_j| < \infty \), the series \( \sum_{j=0}^{\infty} c_j Y_{t-j} \) converges for every \( t \in \mathbb{Z} \) absolutely with probability one.

**Proof.**

1. We will show that \( \{\sum_{j=0}^{n} c_j Y_{t-j}\} \) is a Cauchy sequence in \( L_2(\Omega, \mathcal{A}, P) \) for every \( t \in \mathbb{Z} \).

Assume wlog \( m < n \). \( Y_k \) are uncorrelated with a constant variance \( \sigma^2 \) ⇒

\[
E \left| \sum_{j=0}^{m} c_j Y_{t-j} - \sum_{k=0}^{n} c_k Y_{t-k} \right|^2 = E \left| \sum_{j=m+1}^{n} c_j Y_{t-j} \right|^2
\]

\[
= \sum_{j=m+1}^{n} |c_j|^2 E|Y_{t-j}|^2 = \sigma^2 \sum_{j=m+1}^{n} |c_j|^2 \to 0
\]

as \( m, n \to \infty \)

⇒ there is a mean square limit of the sequence \( \{\sum_{j=0}^{n} c_j Y_{t-j}\} \), which we denote by \( \sum_{j=0}^{\infty} c_j Y_{t-j} \).
2. Since $E|Y_{t-j}| \leq (E|Y_{t-j}|^2)^{\frac{1}{2}} = \sqrt{\sigma^2} < \infty$, we can see that
\[
\sum_{j=0}^{\infty} E|c_j Y_{t-j}| = \sum_{j=0}^{\infty} |c_j| E|Y_{t-j}| \leq \sigma \sum_{j=0}^{\infty} |c_j| < \infty,
\]
and thus $\sum_{j=0}^{\infty} |c_j Y_{t-j}|$ converges almost surely (Rudin, 2003, Theorem 1.38).

**Theorem 33.** Let $\{X_t, t \in \mathbb{Z}\}$ be a weakly stationary centered random sequence with an autocovariance function $R$, let $\{c_j, j \in \mathbb{N}_0\}$ be a sequence of complex-valued constants, such that $\sum_{j=0}^{\infty} |c_j| < \infty$. Then for any $t \in \mathbb{Z}$ the series $\sum_{j=0}^{\infty} c_j X_{t-j}$ converges in mean square and also absolutely with probability one.

**Proof.** 1. For $m < n$ we have
\[
E \left| \sum_{j=m+1}^{n} c_j X_{t-j} \right|^2 \leq E \left( \sum_{j=m+1}^{n} |c_j||X_{t-j}| \right)^2
= \sum_{j=m+1}^{n} \sum_{k=m+1}^{n} |c_j| |c_k| E|X_{t-j}||X_{t-k}|.
\]
The weak stationarity and the Schwarz inequality imply
\[
E|X_{t-j}||X_{t-k}| \leq (E|X_{t-j}|^2)^{\frac{1}{2}} (E|X_{t-k}|^2)^{\frac{1}{2}} = R(0),
\]
\[
\Rightarrow E \left| \sum_{j=m+1}^{n} c_j X_{t-j} \right|^2 \leq R(0) \left( \sum_{j=m+1}^{n} |c_j| \right)^2 \to 0
\]
as $m, n \to \infty$. We have proved the mean square convergence.

2. From the weak stationarity we also get
\[
\sum_{j=0}^{\infty} E|c_j X_{t-j}| = \sum_{j=0}^{\infty} |c_j| E|X_{t-j}| \leq \sqrt{R(0)} \sum_{j=0}^{\infty} |c_j| < \infty,
\]
from which the rest of the proof follows.

**Definition 32.** Let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise WN(0, $\sigma^2$) and $\{c_j, j \in \mathbb{N}_0\}$ a sequence of constants such that, $\sum_{j=0}^{\infty} |c_j| < \infty$. A random sequence $\{X_t, t \in \mathbb{Z}\}$ defined by
\[
X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z}
\]
is called *causal linear process.*
Remark 8. A weaker condition $\sum_{j=0}^{\infty} |c_j|^2 < \infty$ implies mean square convergence in (21) only.

Theorem 34. A causal linear process $\{X_t, t \in \mathbb{Z}\}$ defined by (21), where $\{Y_t, t \in \mathbb{Z}\}$ is WN$(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |c_j| < \infty$, is a centered weakly stationary sequence, with the autocovariance function

$$R_X(t) = \sigma^2 \sum_{k=0}^{\infty} c_{k+t} \overline{c_k}, \quad t \geq 0,$$

$$= \overline{R_X(-t)}, \quad t \leq 0.$$ (22)

The spectral density $f_X$ of the sequence (21) exists and it holds

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{\infty} c_k e^{-ik\lambda} \right|^2, \quad \lambda \in [-\pi, \pi].$$ (23)

Proof. Notice that

- $X_t^{(n)} := \sum_{j=0}^{n} c_j Y_{t-j} \sim \text{MA}(n)$ for a fix $n$
- $\sum_{j=0}^{\infty} |c_j| < \infty \Rightarrow X_t^{(n)} \to X_t$ in mean square as $n \to \infty$ for every $t \in \mathbb{Z}$
- $\{X_t^{(n)}, t \in \mathbb{Z}\}$ is centered, weakly stationary, with the autocovariance function (18) $\Rightarrow \{X_t, t \in \mathbb{Z}\}$ is centered.
- According to Theorem 12, the autocovariance function of $\{X_t^{(n)}, t \in \mathbb{Z}\}$ converges to the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$.

We have proved (22) as well as the stationarity of sequence (21). Further notice that $\{X_t^{(n)}, t \in \mathbb{Z}\}$ has the spectral decomposition

$$X_t^{(n)} = \int_{-\pi}^{\pi} e^{it\lambda} g_n(\lambda) dZ_Y(\lambda), \quad g_n(\lambda) = \sum_{j=0}^{n} c_j e^{-ij\lambda} \in L_2(F_Y).$$

Denote $g(\lambda) = \sum_{j=0}^{\infty} c_j e^{-ij\lambda}$, then:

$$\int_{-\pi}^{\pi} |g_n(\lambda) - g(\lambda)|^2 dF_Y(\lambda) = \int_{-\pi}^{\pi} \left| \sum_{j=n+1}^{\infty} c_j e^{-ij\lambda} \right|^2 dF_Y(\lambda)$$

$$\leq \int_{-\pi}^{\pi} \left( \sum_{j=n+1}^{\infty} |c_j| \right)^2 f_Y(\lambda) d\lambda = \sigma^2 \left( \sum_{j=n+1}^{\infty} |c_j| \right)^2 \to 0$$
Thus \( g_n \to g \) in \( L^2(F_Y) \). According to Theorem 26,
\[
X_t^{(n)} = \int_{-\pi}^{\pi} e^{it\lambda} g_n(\lambda) dZ_Y(\lambda) \to \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ_Y(\lambda)
\]
in mean square as \( n \to \infty \) and since also \( X_t^{(n)} \to X_t \) in mean square, it means that
\[
X_t = \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ_Y(\lambda).
\]

By Theorem 26
\[
E(X_{s+t}X_s) = E \int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) dZ_Y(\lambda) \int_{-\pi}^{\pi} e^{is\lambda} g(\lambda) dZ_Y(\lambda)
\]
\[
= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 dF_Y(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 \frac{\sigma^2}{2\pi} d\lambda,
\]
(the spectral decomposition of the autocovariance function of the sequence (21)). Function \( \frac{\sigma^2}{2\pi}|g(\lambda)|^2 \) is the spectral density of the process (21).

\[ \square \]

**Example 25.** Let us consider a causal linear process (21),
\[
X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z}, \quad c_j = \varphi^j, \quad |\varphi| < 1
\]
The autocovariance function is
\[
R_X(t) = \sigma^2 \frac{\varphi^t}{1 - \varphi^2}, \quad t \geq 0,
\]
\[
= R_X(-t), \quad t \leq 0,
\]
equivalently
\[
R_X(t) = \sigma^2 \frac{\varphi^{|t|}}{1 - \varphi^2}, \quad t \in \mathbb{Z}.
\]
The spectral density is
\[
f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \varphi^j e^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\lambda}|^2}
\]
Further, we can write
\[
X_t = \sum_{j=0}^{\infty} \varphi^j Y_{t-j} = Y_t + \sum_{j=1}^{\infty} \varphi^j Y_{t-j} = Y_t + \varphi \sum_{j=1}^{\infty} \varphi^{j-1} Y_{t-j}
\]
\[
= Y_t + \varphi \sum_{k=0}^{\infty} \varphi^k Y_{t-1-k}
\]
\[
= \varphi X_{t-1} + Y_t.
\]
The sequence \( \{X_t, t \in \mathbb{Z}\} \) defined by (24) is called autoregressive sequence of order one, AR(1).

### 8.4 Autoregressive sequences

**Definition 33.** A random sequence \( \{X_t, t \in \mathbb{Z}\} \) is called autoregressive sequence of order \( n \) (notation AR\((n)\)), if it satisfies equation

\[
X_t = \varphi_1 X_{t-1} + \cdots + \varphi_n X_{t-n} + Y_t, \quad t \in \mathbb{Z},
\]

where \( \varphi_1, \ldots, \varphi_n \) are real-valued constants, \( \varphi_n \neq 0 \) and \( \{Y_t, t \in \mathbb{Z}\} \) is a white noise.

An equivalent notation:

\[
X_t + a_1 X_{t-1} + \cdots + a_n X_{t-n} = \sum_{j=0}^{n} a_j X_{t-j} = Y_t,
\]

where \( a_0 = 1 \).

We want to express an AR\((n)\) sequence as a causal linear process. First, we define a backward-shift operator by

\[
BX_t = X_{t-1}, \quad B^0 X_t = X_t, \quad B^k X_t = B^{k-1}(BX_t) = X_{t-k}, \quad k \in \mathbb{Z}.
\]

Using this operator, the relation (26) can be shortly written in the form

\[
a(B)X_t = Y_t
\]

where \( a(B) \) is a polynomial operator, formally identical with the algebraic polynomial

\[
a(z) = 1 + a_1 z + \cdots + a_n z^n.
\]

Similarly, let \( \{c_k, k \in \mathbb{Z}\} \) be a sequence of constants such that \( \sum_{k=-\infty}^{\infty} |c_k| < \infty \). The series

\[
c(z) = \sum_{k=-\infty}^{\infty} c_k z^k
\]

is absolutely convergent at least inside the unite circle and defines the operator

\[
c(B) = \sum_{k=-\infty}^{\infty} c_k B^k.
\]

This operator has usual properties of algebraic power series.
Theorem 35. Let \( \{X_t, t \in \mathbb{Z}\} \) be an autoregressive sequence of order \( n \) defined by (26). If all the roots of the polynomial \( a(z) = 1 + a_1z + \cdots + a_nz^n \) lie outside the unit circle in \( \mathbb{C} \) then \( \{X_t, t \in \mathbb{Z}\} \) is a causal linear process, i.e.,

\[
X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \; t \in \mathbb{Z},
\]

where \( c_j \) are defined by

\[
c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{a(z)}, \quad |z| \leq 1.
\]

The autocovariance function of this sequence is given by (22) and the spectral density is

\[
f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{n} a_j e^{-ij\lambda} \right|^2, \quad \lambda \in [-\pi, \pi],
\]

where \( a_0 = 1 \).

Proof. Consider an AR(n) sequence,

\[
X_t + a_1X_{t-1} + \cdots + a_nX_{t-n} = a(B)X_t = Y_t.
\]

If all the roots \( z_i, i = 1, \ldots, n, \) of the polynomial \( a(z) = 1 + a_1z + \cdots + a_nz^n \) are outside the unit circle, \( a(z) \neq 0 \) for \( |z| \leq 1 \). Since \( |z_i| \geq \min_{1 \leq i \leq n} |z_i| \geq 1 + \delta \) for a \( \delta > 0 \), \( a(z) \neq 0 \) for \( |z| \leq 1 + \delta \), and \( c(z) = \frac{1}{a(z)} \) is holomorphic in the region \( |z| < 1 + \delta \) and has a representation

\[
c(z) = \sum_{j=0}^{\infty} c_j z^j, \; |z| < 1 + \delta.
\]

The series (29) is absolutely convergent in any closed circle with the radius \( r < 1 + \delta \) which means that \( \sum_{j=0}^{\infty} |c_j| < \infty \) and \( c(z)a(z) = 1, |z| \leq 1 \). Thus,

\[
c(B)a(B)X_t = X_t = c(B)Y_t = \sum_{j=0}^{\infty} c_j Y_{t-j}.
\]

We have proved that the sequence \( \{X_t, t \in \mathbb{Z}\} \) is a causal linear process, that satisfies Theorem 34. It is centered and weakly stationary, with the autocovariance function (22) and spectral density

\[
f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{\infty} e^{-ik\lambda} c_k \right|^2 = \frac{\sigma^2}{2\pi} |c(e^{-i\lambda})|^2
\]

\[
= \frac{\sigma^2}{2\pi} \left| \frac{1}{a(e^{-i\lambda})} \right|^2 = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{n} a_j e^{-ij\lambda} \right|^2.
\]
If all the roots of \( a(z) \) are simple, we can obtain coefficients \( c_j \) in the representation (29) by using decomposition into partial fractions:

\[
c(z) = \frac{1}{a(z)} = \frac{A_1}{z_1 - z} + \frac{A_2}{z_2 - z} + \cdots + \frac{A_n}{z_n - z}
\]

For \( |z| \leq 1 \) and \( |z_j| > 1 \),

\[
\frac{A_j}{z_j - z} = \frac{A_j}{z_j \left(1 - \frac{z}{z_j}\right)} = \frac{A_j}{z_j} \sum_{k=0}^{\infty} \left(\frac{z}{z_j}\right)^k
\]

\[
c(z) = \sum_{j=1}^{n} \frac{A_j}{z_j - z} = \sum_{j=1}^{n} \frac{A_j}{z_j} \sum_{k=0}^{\infty} \left(\frac{z}{z_j}\right)^k
\]

\[
= \sum_{k=0}^{\infty} z^k \sum_{j=1}^{n} \frac{A_j}{z_j^{k+1}} = \sum_{k=0}^{\infty} c_k z^k, \quad c_k = \sum_{j=1}^{n} \frac{A_j}{z_j^{k+1}}.
\]

Since for all \( i = 1, \ldots, n \) we have \( |z_i| \geq 1 + \delta > 1 \), it holds

\[
|c_k| < \frac{1}{(1 + \delta)^{k+1}} \sum_{j=1}^{n} |A_j|
\]

from which we conclude that \( \sum_{k=0}^{\infty} |c_k| < \infty \).

If the roots of the polynomial \( a(z) \) are multiple, we proceed analogously.

Coefficients \( c_k \) can be also obtained by a solution of the system of equations

\[
c_0 = 1, \quad c_1 + a_1 c_0 = 0, \quad c_2 + a_1 c_1 + a_2 c_0 = 0, \quad \ldots \quad c_p + a_1 c_{p-1} + \cdots + a_n c_{p-n} = 0, \quad p = n, n + 1, \ldots,
\]

if we compare coefficients with the same powers of \( z \) at both sides of the relation \( a(z)c(z) = 1 \). The system of equations

\[
c_p + a_1 c_{p-1} + \cdots + a_n c_{p-n} = 0
\]

for \( p \geq n \) can be solved as a system of homogeneous difference equations of order \( n \) with constant coefficients, and initial conditions \( c_0, c_1, \ldots, c_{n-1} \).
Yule-Walker equations:

Let us consider a sequence \( \{X_t, t \in \mathbb{Z}\} \),

\[
X_t + a_1X_{t-1} + \cdots + a_nX_{t-n} = Y_t,
\]

that satisfies conditions of Theorem 35, with \( a_1, \ldots, a_n \) being real-valued coefficients and \( \{Y_t, t \in \mathbb{Z}\} \) is real white noise WN(0,\( \sigma^2 \)). The sequence \( \{X_t, t \in \mathbb{Z}\} \) is a real-valued causal linear process, \( Y_t \) are uncorrelated \( \Rightarrow \mathbb{E}X_sY_t = \langle X_s, Y_t \rangle = 0 \) for \( s < t \).

Multiplying (30) by \( Y_t \) and taking expectation we get

\[
\mathbb{E}X_tY_t + a_1\mathbb{E}X_{t-1}Y_t + \cdots + a_n\mathbb{E}X_{t-n}Y_t = \mathbb{E}Y_t^2,
\]

thus

\[
\mathbb{E}X_tY_t = \sigma^2.
\]

Multiplying (30) by \( X_{t-k} \) for \( k \geq 0 \) and taking expectation we get a system of equations

\[
\mathbb{E}X_tX_{t-k} + a_1\mathbb{E}X_{t-1}X_{t-k} + \cdots + a_n\mathbb{E}X_{t-n}X_{t-k} = \mathbb{E}Y_tX_{t-k},
\]

or, if we put \( R_X(t) = R(t) \),

\[
R(0) + a_1R(1) + \cdots + a_nR(n) = \sigma^2, \quad k = 0, \tag{31}
\]

\[
R(k) + a_1R(k-1) + \cdots + a_nR(n-k) = 0, \quad k \geq 1. \tag{32}
\]

Equations (31) and (32) are called Yule-Walker equations.

Solution: Dividing (32) for \( k \geq 1 \) by \( R(0) \) we get equations for the autocorrelation function \( r(t) = R(t)/R(0) \).

- First solve the system for \( k = 1, \ldots, n - 1 \):

\[
\begin{align*}
r(1) + a_1 + a_2r(1) + a_3r(2) + \cdots + a_nr(n-1) &= 0, \\
r(2) + a_1r(1) + a_2 + a_3r(1) + \cdots + a_nr(n-2) &= 0, \\
&\vdots \\
r(n-1) + a_1r(n-2) + \cdots + a_nr(1) &= 0.
\end{align*}
\]

- Values \( r(1), \ldots, r(n-1) \) together with \( r(0) = 1 \) serve as initial conditions to solve the system of difference equations

\[
r(k) + a_1r(k-1) + \cdots + a_nr(n-k) = 0, \quad k \geq n,
\]

with the characteristic polynomial

\[
\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = L(\lambda).
\]
In this way we get the solution $r(t)$ for $t \geq 0$. For a real sequence $r(t) = r(-t)$.

If we insert $R(k) = r(k)R(0)$ into (31) we get the equation for $R(0)$:

$$R(0)[1 + a_1r(1) + \cdots + a_nr(n)] = \sigma^2,$$

thus

$$R(0) = \frac{\sigma^2}{1 + a_1r(1) + \cdots + a_nr(n)}.$$  \hspace{1cm} (33)

Remark: If $z_i, i = 1, \ldots, n,$ are the roots of the polynomial $a(z) = 1 + a_1z + \cdots + a_nz^n$, then $\lambda_i = z_i^{-1}$, $i = 1, \ldots, n,$ are the roots of the polynomial $L(z) = z^n + a_1z^{n-1} + \cdots + a_n$.

The AR($n$) sequence is a causal linear process, if all the roots of the polynomial $L(z)$ are inside the unit circle.

**Example 26.** An AR(1) sequence:

$$X_t + aX_{t-1} = Y_t, \ Y_t \sim WN(0, \sigma^2), |a| < 1.$$ 

Polynomial $a(z) = 1 + az$, root $-\frac{1}{a}$ is outside the unit circle $\Rightarrow \{X_t, t \in \mathbb{Z}\}$ is a causal linear process, weakly stationary.

Yule - Walker equations for the autocovariance function $R_X(t) = R(t)$:

$$R(0) + aR(1) = \sigma^2,$$

$$R(k) + aR(k-1) = 0, \ k \geq 1.$$ 

A general solution to the difference equation for the autocorrelation function: $r(k) = c(-a)^k$, initial condition $r(0) = 1 = c$. Value $R(0)$ can be determined from formula (33):

$$R(0) = \frac{\sigma^2}{1 + a r(1)} = \frac{\sigma^2}{1 - a^2}.$$ 

**Example 27.** Consider AR(2) sequence

$$X_t - \frac{3}{4}X_{t-1} + \frac{1}{8}X_{t-2} = Y_t, \ Y_t \sim WN(0, \sigma^2).$$

The polynomial $a(z) = 1 - \frac{3}{4}z + \frac{1}{8}z^2$ has roots $z_1 = 2, \ z_2 = 4$, $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process, weakly stationary.

Yule-Walker equations:

$$R(0) - \frac{3}{4}R(1) + \frac{1}{8}R(2) = \sigma^2,$$

$$R(k) - \frac{3}{4}R(k-1) + \frac{1}{8}R(k-2) = 0, \ k \geq 1.$$ 

56
Equations for the autocorrelation function:

\[ r(1) - \frac{3}{4} + \frac{1}{8} r(1) = 0, \]
\[ r(k) - \frac{3}{4} r(k-1) + \frac{1}{8} r(k-2) = 0, \quad k \geq 2. \quad (34) \]

Solving the first equation we get \( r(1) = \frac{2}{3}. \)

For \( k \geq 2 \) we solve a second order difference equation with initial conditions \( r(0) = 1, r(1) = \frac{2}{3}. \)

Characteristic equation \( L(\lambda) = \lambda^2 - \frac{3}{4} \lambda + \frac{1}{8} = 0 \) has two different real roots \( \lambda_1 = \frac{1}{2}, \lambda_2 = \frac{1}{4}. \) A general solution of the difference equation (34) is

\[ r(k) = c_1 \lambda_1^k + c_2 \lambda_2^k = c_1 \left( \frac{1}{2} \right)^k + c_2 \left( \frac{1}{4} \right)^k. \]

Constants \( c_1, c_2 \) satisfy equations

\[ c_1 + c_2 = r(0), \]
\[ \lambda_1 c_1 + \lambda_2 c_2 = r(1), \]

so that \( c_1 = \frac{5}{3}, c_2 = -\frac{2}{3}, \) and

\[ r(k) = \frac{5}{3} \left( \frac{1}{2} \right)^k - \frac{2}{3} \left( \frac{1}{4} \right)^k, \quad k = 0, 1, \ldots \]
\[ r(k) = r(-k), \quad k = -1, -2, \ldots \]

8.5 ARMA sequences

**Definition 34.** A random sequence \( \{X_t, t \in \mathbb{Z}\} \) satisfies an ARMA\((m,n)\) model if

\[ X_t + a_1 X_{t-1} + \cdots + a_m X_{t-m} = Y_t + b_1 Y_{t-1} + \cdots + b_n Y_{t-n}, \quad t \in \mathbb{Z}, \quad (35) \]

where \( a_i, i = 1, \ldots, m, \ b_i, i = 1, \ldots, n, \) are real constants, \( a_m \neq 0, b_n \neq 0 \) and the sequence \( \{Y_t, t \in \mathbb{Z}\} \) is a white noise.

Equivalently:

\[ X_t = \varphi_1 X_{t-1} + \cdots + \varphi_m X_{t-m} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_n Y_{t-n}. \]

The model called ARMA\((m,n)\) is a mixed model of autoregressive and moving average sequences.

Consider polynomials \( a(z) = 1 + a_1 z + \cdots + a_m z^m \) and \( b(z) = 1 + b_1 z + \cdots + b_n z^n. \) Then we can write ARMA\((m,n)\) model in the form

\[ a(B) X_t = b(B) Y_t. \quad (36) \]
Theorem 36. Let \( \{X_t, t \in \mathbb{Z}\} \) be a random ARMA\((m,n)\) sequence given by (36). Suppose that the polynomials \( a(z) \) and \( b(z) \) have no common roots and all the roots of the polynomial \( a(z) = 1 + a_1 z + \cdots + a_m z^m \) are outside the unit circle. Then \( X_t \) is of the form

\[
X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z},
\]

where coefficients \( c_j \) satisfy

\[
c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)}, \quad |z| \leq 1.
\]

The spectral density of the sequence \( \{X_t, t \in \mathbb{Z}\} \) is

\[
f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{n} b_j e^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| c(e^{-i\lambda}) \right|^2 = \frac{\sigma^2}{2\pi} \left| \frac{b(e^{-i\lambda})}{a(e^{-i\lambda})} \right|^2, \quad \lambda \in [-\pi, \pi],
\]

where \( a_0 = 1, b_0 = 1 \).

Proof. We proceed analogously as in the Proof of Theorem 35. Since all the roots of the polynomial \( a(z) \) are lying outside the unit circle it holds for \( |z| \leq 1 \)

\[
\frac{1}{a(z)} = h(z) = \sum_{j=0}^{\infty} h_j z^j, \quad \text{where} \quad \sum_{j=0}^{\infty} |h_j| < \infty.
\]

Thus, \( h(z)a(z) = 1 \) for \( |z| \leq 1 \) and if we apply the operator \( h(B) \) to both sides of equations (36), we have

\[
h(B)a(B)X_t = X_t = h(B)b(Y_t) = c(B)Y_t,
\]

where \( c(z) = b(z)/a(z) \) and \( \sum_{j=0}^{\infty} |c_j| < \infty \).

The sequence \( \{X_t, t \in \mathbb{Z}\} \) is a causal linear process with the autocovariance function (22) and spectral density

\[
f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{m} c_j e^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| c(e^{-i\lambda}) \right|^2 = \frac{\sigma^2}{2\pi} \left| \frac{b(e^{-i\lambda})}{a(e^{-i\lambda})} \right|^2 = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{n} b_j e^{-ij\lambda} \right|^2
\]

Remark. If the polynomials \( a(z) \) a \( b(z) \) have common roots the polynomial \( c(z) = b(z)/a(z) \) defines an ARMA\((p,q)\), process with \( p < m, q < n \).
Example 28. Consider an ARMA(1, 1) model

\[ X_t + aX_{t-1} = Y_t + bY_{t-1}, \quad t \in \mathbb{Z}, \]

where \( Y_t \) is a white noise WN(0, \( \sigma^2 \)), \( a, b \neq 0, a \neq b, |a| < 1 \).

We have \( a(z) = 1 + az, \ b(z) = 1 + bz \), the roots \( z_a = -\frac{1}{a}, \ z_b = -\frac{1}{b} \), respectively, are different and \( |z_a| > 1 \). All the assumptions of the previous theorem are satisfied.

For \( |z| \leq 1 \),

\[ c(z) = \frac{1 + bz}{1 + az} = (1 + bz) \sum_{j=0}^{\infty} (-a)^j z^j = \sum_{j=0}^{\infty} c_j z^j, \]

and if we compare coefficients with the same powers of \( z \) we obtain

\[ c_0 = 1, \ c_j = (-a)^j (b - a), \ j \geq 1. \]

The autocovariance function of the sequence \( \{X_t, t \in \mathbb{Z}\} \) is

\[ R_X(k) := R(k) = \sigma^2 \sum_{j=0}^{\infty} c_j c_{j+k}, \quad k \in \mathbb{Z}. \]

Computation of \( R(0) \):

\[
R(0) = \sigma^2 \sum_{j=0}^{\infty} c_j^2 = \sigma^2 \left[ 1 + \sum_{j=1}^{\infty} \left( (a^{-1}b - a)^2 \right) \right] \\
= \sigma^2 \left( 1 + \frac{(b - a)^2}{1 - a^2} \right) = \sigma^2 \frac{1 - 2ab + b^2}{1 - a^2}.
\]

For \( k \geq 1 \)

\[
R(k) = \sigma^2 \sum_{j=0}^{\infty} c_j c_{j+k} = \sigma^2 \left[ c_0 c_k + \sum_{j=1}^{\infty} c_j c_{j+k} \right] \\
= \sigma^2 \left[ (-a)^{k-1}(b - a) + (-a)^k (b - a)^2 \sum_{j=0}^{\infty} (-a)^{2j} \right] \\
= \sigma^2 \left[ (-a)^{k-1}(b - a) + (-a)^k \frac{(b - a)^2}{1 - a^2} \right] \\
= \sigma^2 (-a)^{k-1}(b - a) \frac{1 - ab}{1 - a^2} = (-a)^{k-1} R(1).
\]

The spectral density is

\[ f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|1 + be^{-i\lambda}|^2}{|1 + ae^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \cdot \frac{1 + 2b \cos \lambda + b^2}{1 + 2a \cos \lambda + a^2}, \quad \lambda \in [-\pi, \pi]. \]
An analogy of the Yule-Walker equations: Multiplying (35) by $X_{t-k}, k \geq 0$ and taking the expectation we get

\[EX_tX_{t-k} + \sum_{j=1}^{m} a_j EX_{t-j}X_{t-k} = EY_tX_{t-k} + \sum_{j=1}^{n} b_j EY_{t-j}X_{t-k}.\]

From Theorem 36 and properties of the white noise,

\[EX_{t-j}Y_{t-k} = \begin{cases} \sigma^2 \delta_{k-j}, & k \geq j, \\ 0, & k < j \end{cases}\]

and the previous equation for $k \geq 0$ can be written in the form

\[
R(k) + \sum_{j=1}^{m} a_j R(k - j) = \sigma^2 \sum_{j=k}^{n} b_j c_{j-k}, \quad k \leq n, \quad (38)
\]

\[
R(k) + \sum_{j=1}^{m} a_j R(k - j) = 0 \quad k > n. \quad (39)
\]

For $k \geq \max(m, n + 1)$, (39) is solved as a difference equation with initial conditions that can be obtained from the system of equations for $k < \max(m, n + 1)$.

**Example 29.** Consider again the ARMA(1, 1) model

\[X_t + aX_{t-1} = Y_t + bY_{t-1}, \quad t \in \mathbb{Z},\]

where $a \neq b \neq 0, |a| < 1, Y_t \sim WN(0, \sigma^2)$.

Equations (38) a (39) are of the form

\[
R(0) + aR(1) = \sigma^2 + b(b - a)\sigma^2,
\]

\[
R(1) + aR(0) = \sigma^2 b,
\]

\[
R(k) + aR(k - 1) = 0, \quad k \geq 2.
\]

The difference equation $R(k) + aR(k - 1) = 0$ with an initial condition for $R(1)$ has the solution $R(k) = (-a)^{k-1}R(1), k \geq 1$.

The values of $R(1)$ a $R(0)$ will be computed from the first and the second equations:

\[
R(0) = \frac{1}{1 - a^2} \left[ \sigma^2 (1 - 2ab + b^2) \right],
\]

\[
R(1) = \frac{1}{1 - a^2} \left[ \sigma^2 (b - a)(1 - ab) \right].
\]
**Definition 35.** Let \( \{X_t, \ t \in \mathbb{Z}\} \) be a stationary ARMA\((m,n)\) sequence defined by \( (36) \),
\[
a(B)X_t = b(B)Y_t, \quad t \in \mathbb{Z},
\]
where \( \{Y_t, \ t \in \mathbb{Z}\} \) is a white noise WN\((0, \sigma^2)\). The sequence \( \{X_t, \ t \in \mathbb{Z}\} \) is said to be invertible, if there exists a sequence of constants \( \{d_j, \ j \in \mathbb{N}_0\} \) such that \( \sum_{j=0}^{\infty} |d_j| < \infty \) and
\[
Y_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \quad t \in \mathbb{Z}.
\]

(40)

Conditions under which an ARMA sequence is invertible:

**Theorem 37.** Let \( \{X_t, \ t \in \mathbb{Z}\} \) be a stationary ARMA\((m,n)\) random sequence defined by \( (36) \). Let the polynomials \( a(z) \) and \( b(z) \) have no common roots and the polynomial \( b(z) = 1 + b_1 z + \cdots + b_n z^n \) has all the roots outside the unit circle. Then \( \{X_t, \ t \in \mathbb{Z}\} \) is invertible and
\[
Y_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \quad t \in \mathbb{Z},
\]
where coefficients \( d_j \) are defined by
\[
d(z) = \sum_{j=0}^{\infty} d_j z^j = \frac{a(z)}{b(z)}, \quad |z| \leq 1.
\]

Proof. The theorem can be proved analogously as Theorem 36 by inverting the polynomial \( b(z) \). The correctness of all operations is guaranteed by Theorem 33 since we assume that \( \{X_t, \ t \in \mathbb{Z}\} \) is stationary.

Let us notice that the equation \( d(z)b(z) = a(z) \) with polynomials \( a(z) = 1 + a_1 z + \cdots + a_m z^m, \ b(z) = 1 + b_1 z + \cdots + b_n z^n \), respectively, implies \( d_0 = 1 \). Relation (40) can be written as
\[
X_t + \sum_{j=1}^{\infty} d_j X_{t-j} = Y_t, \quad t \in \mathbb{Z}.
\]

(41)
The invertible ARMA\((m,n)\) sequence can be thus expressed as an AR\((\infty)\) sequence.

**8.6 Linear filters**

**Definition 36.** Let \( \{Y_t, \ t \in \mathbb{Z}\} \) be a centered weakly stationary sequence. Let \( \{c_k, \ k \in \mathbb{Z}\} \) be a sequence of (complex-valued) numbers such that \( \sum_{j=-\infty}^{\infty} |c_j| < \infty \).
We say that a random sequence \(\{X_t, t \in \mathbb{Z}\}\) is obtained by filtration of a sequence \(\{Y_t, t \in \mathbb{Z}\}\), if

\[
X_t = \sum_{j=-\infty}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z}.
\]

(42)

The sequence \(\{c_j, j \in \mathbb{Z}\}\) is called \textit{time-invariant linear filter}. Provided that \(c_j = 0\) for all \(j < 0\), we say that the filter \(\{c_j, j \in \mathbb{Z}\}\) is \textit{causal}.

**Theorem 38.** Let \(\{Y_t, t \in \mathbb{Z}\}\) be a centered weakly stationary sequence with an autocovariance function \(R_Y\) and spectral density \(f_Y\) and let \(\{c_k, k \in \mathbb{Z}\}\) be a linear filtr such that \(\sum_{k=-\infty}^{\infty} |c_k| < \infty\). Then \(\{X_t, t \in \mathbb{Z}\}\), where \(X_t = \sum_{k=-\infty}^{\infty} c_k Y_{t-k}\), is a centered weakly stationary sequence with the autocovariance function

\[
R_X(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_j c_k R_Y(t-j+k), \quad t \in \mathbb{Z}
\]

and the spectral density

\[
f_X(\lambda) = |\Psi(\lambda)|^2 f_Y(\lambda), \quad \lambda \in [-\pi, \pi],
\]

(43)

where

\[
\Psi(\lambda) = \sum_{k=-\infty}^{\infty} c_k e^{-ik\lambda}
\]

for \(\lambda \in [\pi, \pi]\) is called the transfer function of the filter.

**Proof.** Let \(X_t^{(n)} = \sum_{k=-n}^{n} c_k Y_{t-k}\); obviously, for each \(t \in \mathbb{Z}\), \(X_t^{(n)} \to X_t\) in mean square as \(n \to \infty\).

For any \(t \in \mathbb{Z}\), \(Y_t\) has the spectral decomposition \(Y_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_Y(\lambda)\), where \(Z_Y\) is a process with orthogonal increments and the associated distribution function \(F_Y(\lambda)\). Thus

\[
X_t^{(n)} = \sum_{k=-n}^{n} c_k Y_{t-k} = \sum_{k=-n}^{n} c_k \int_{-\pi}^{\pi} e^{i(t-k)\lambda} dZ_Y(\lambda)
\]

\[
= \int_{-\pi}^{\pi} e^{it\lambda} \sum_{k=-n}^{n} c_k e^{-ik\lambda} dZ_Y(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} h_n(\lambda) dZ_Y(\lambda),
\]

where \(h_n(\lambda) = \sum_{k=-n}^{n} c_k e^{-ik\lambda}\). For the same reasons as in the proof of Theorem 34, \(h_n\) converges to a function \(\Psi\) in the space \(L_2(F_Y)\), where \(\Psi(\lambda) = \sum_{k=-\infty}^{\infty} c_k e^{-ik\lambda}\), and by Theorem 26

\[
X_t = \lim_{n \to \infty} X_t^{(n)} = \int_{-\pi}^{\pi} e^{it\lambda} \Psi(\lambda) dZ_Y(\lambda)
\]
for any \( t \in \mathbb{Z} \).

Since \( \{X_{t}^{(n)}, t \in \mathbb{Z}\} \) is centered, \( \{X_{t}, t \in \mathbb{Z}\} \) is also centered, and according to the theorem on convergence of processes the autocovariance functions \( \{X_{t}^{(n)}, t \in \mathbb{Z}\} \) converge to the autocovariance function of \( \{X_{t}, t \in \mathbb{Z}\} \), and so

\[
\text{E}X_{s+t}X_{s} = \lim_{n \to \infty} \text{E}X_{s+t}^{n}X_{s}^{n} = \lim_{n \to \infty} \sum_{j=-n}^{n} \sum_{k=-n}^{n} c_{j}c_{k} \text{E}(Y_{s+t-j}Y_{s-k})
= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j}c_{k} R_{Y}(t-j+k) := R_{X}(t).
\]

\( \text{E}X_{s+t}X_{s} = R_{X}(t) \) is a function of one variable, only, \( \{X_{t}, t \in \mathbb{Z}\} \) is weakly stationary. It also holds

\[
R_{X}(t) = \text{E} \int_{-\pi}^{\pi} e^{i(t+s)\lambda} \Psi(\lambda) dZ_{Y}(\lambda) \int_{-\pi}^{\pi} e^{i\lambda\lambda} \Psi(\lambda) dZ_{Y}(\lambda)
= \int_{-\pi}^{\pi} e^{it\lambda} |\Psi(\lambda)|^{2} dF_{Y}(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} |\Psi(\lambda)|^{2} f_{Y}(\lambda) d\lambda
\]

and from the spectral decomposition of the autocovariance function it follows that the function

\[
|\Psi(\lambda)|^{2} f_{Y}(\lambda) := f_{X}(\lambda)
\]

is the spectral density of the sequence \( \{X_{t}, t \in \mathbb{Z}\} \).

\[\square\]

**Example 30.** Let \( \{Y_{t}, t \in \mathbb{Z}\} \) be a white noise \( \text{WN}(0, \sigma^{2}) \) sequence, \( \{c_{k}, k \in \mathbb{Z}\} \) be a sequence of constants such that \( \sum_{k=-\infty}^{\infty} |c_{k}| < \infty \). Then a linear process defined by formula \( X_{t} = \sum_{k=-\infty}^{\infty} c_{k}Y_{t-k} \) is obtained by a linear filtration of the white noise. Similarly, a causal linear process is obtained by a filtration of the white noise by using a causal linear filter with \( c_{k} = 0, \ k < 0 \).

**Example 31.** Let \( \{X_{t}, t \in \mathbb{Z}\} \) be a random sequence defined by \( X_{t} = \varphi X_{t-1} + Y_{t} \), where \( Y_{t} \) are elements of a white noise sequence and \( |\varphi| > 1 \).

Then \( \{X_{t}, t \in \mathbb{Z}\} \) is not a causal linear process, but we can write

\[
X_{t} = -\sum_{k=1}^{\infty} \varphi^{-k}Y_{t+k}.
\]

In this case, we have the linear filter, such that

\[
c_{k} = \begin{cases} 
0, & k \geq 0 \\
-(\varphi)^{k}, & k < 0.
\end{cases}
\]
9 Selected limit theorems

9.1 Laws of large numbers

Definition 37. We say that a stationary sequence \( \{X_t, t \in \mathbb{Z}\} \) with mean \( \mu \) is \textit{mean square ergodic} or it follows the law of large numbers in \( L_2(\Omega, \mathcal{A}, P) \), if, as \( n \to \infty \),

\[
\frac{1}{n} \sum_{t=1}^{n} X_t \to \mu \quad \text{in mean square.} \tag{44}
\]

If \( \{X_t, t \in \mathbb{Z}\} \) is a sequence that is mean square ergodic then

\[
\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{P} \mu,
\]
i.e., \( \{X_t, t \in \mathbb{Z}\} \) satisfies the \textit{weak law of large numbers} for stationary sequences.

Theorem 39. A stationary random sequence \( \{X_t, t \in \mathbb{Z}\} \) with mean \( \mu \) and autocovariance function \( R \) is mean square ergodic if and only if

\[
\frac{1}{n} \sum_{t=1}^{n} R(t) \to 0 \quad \text{as} \quad n \to \infty. \tag{45}
\]

Proof. Wlog put \( \mu = 0 \) (otherwise we consider \( \tilde{X}_t := X_t - \mu \)). Consider the spectral decomposition

\[
X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda),
\]
where \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is an orthogonal increment process with the associated distribution function \( F \), which is same as the spectral distribution function of \( \{X_t, t \in \mathbb{Z}\} \).

Then

\[
\frac{1}{n} \sum_{t=1}^{n} X_t = \frac{1}{n} \sum_{t=1}^{n} \left( \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) \right) = \int_{-\pi}^{\pi} \left( \frac{1}{n} \sum_{t=1}^{n} e^{it\lambda} \right) dZ(\lambda)
\]

\[
= \int_{-\pi}^{\pi} h_n(\lambda)dZ(\lambda),
\]
where

\[
h_n(\lambda) = \frac{1}{n} \sum_{t=1}^{n} e^{it\lambda} = \begin{cases} \frac{1}{n} e^{i\lambda} \frac{1 - e^{n\lambda}}{1 - e^{i\lambda}}, & \lambda \neq 0, \\ 1, & \lambda = 0. \end{cases}
\]

Further, let us consider function

\[
h(\lambda) = \begin{cases} 0, & \lambda \neq 0, \\ 1, & \lambda = 0. \end{cases}
\]

64
and define the random variable
\[ Z_0 = \int_{-\pi}^{\pi} h(\lambda) dZ(\lambda). \]

Obviously, \( h_n(\lambda) \to h(\lambda) \) for any \( \lambda \in [-\pi, \pi] \). Moreover, \( h_n \to h \) in \( L_2(F) \), since \( |h_n(\lambda) - h(\lambda)|^2 \leq 4 \) and by the Lebesgue theorem, as \( n \to \infty \),
\[ \int_{-\pi}^{\pi} |h_n(\lambda) - h(\lambda)|^2 dF(\lambda) \to 0. \]
Hence, as \( n \to \infty \)
\[ \frac{1}{n} \sum_{t=1}^{n} X_t = \int_{-\pi}^{\pi} h_n(\lambda) dZ(\lambda) \to \int_{-\pi}^{\pi} h(\lambda) dZ(\lambda) = Z_0 \]
in mean square.

Now, it suffices to show that
\[ Z_0 = 0 \text{ a.s. } \iff \frac{1}{n} \sum_{t=1}^{n} R(t) \to 0 \text{ as } n \to \infty. \] (46)

From Theorem 26 we have \( E Z_0 = 0 \); thus \( Z_0 = 0 \) a.s. if and only if \( E|Z_0|^2 = 0 \). Further from Theorem 26,
\[ E|Z_0|^2 = E \left| \int_{-\pi}^{\pi} h(\lambda) dZ(\lambda) \right|^2 = \int_{-\pi}^{\pi} |h(\lambda)|^2 dF(\lambda). \]

From the spectral decomposition of the autocovariance function
\[ \frac{1}{n} \sum_{t=1}^{n} R(t) = \frac{1}{n} \sum_{k=1}^{n} \left[ \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda) \right] = \int_{-\pi}^{\pi} \left( \frac{1}{n} \sum_{t=1}^{n} e^{it\lambda} \right) dF(\lambda) \]
\[ = \int_{-\pi}^{\pi} h_n(\lambda) dF(\lambda) \to \int_{-\pi}^{\pi} h(\lambda) dF(\lambda) = \int_{-\pi}^{\pi} |h(\lambda)|^2 dF(\lambda) \] (47)
The rest of the proof follows from (46) and (47).

**Example 32.** Let us consider an AR(1) process
\[ X_t = \varphi X_{t-1} + Y_t, \quad Y_t \sim \text{WN}(0, \sigma^2), \quad |\varphi| < 1. \]
We know that the autocovariance function of \( \{X_t, t \in \mathbb{Z}\} \) is
\[ R_X(t) = \frac{\sigma^2}{1 - \varphi^2 |\varphi|^t}. \]
Obviously,
\[
\frac{1}{n} \sum_{t=1}^{n} R_X(t) = \frac{\sigma^2}{1 - \varphi^2} \frac{1}{n} \sum_{t=1}^{n} \varphi^t = \frac{1}{n} \frac{\sigma^2}{1 - \varphi} \varphi(1 - \varphi^n) \to 0
\]
as \(n \to \infty\), from which we conclude that \(\{X_t, t \in \mathbb{Z}\}\) is mean square ergodic.

**Example 33.** Let \(\{X_t, t \in \mathbb{Z}\}\) be a stationary mean square ergodic sequence with mean \(\mu\) and autocovariance function \(R_X\). Define a random sequence \(\{Z_t, t \in \mathbb{Z}\}\) by
\[
Z_t = X_t + Y, \quad t \in \mathbb{Z},
\]
where \(EY = 0\), \(\text{var}Y = \sigma^2 \in (0, \infty)\), and \(EX_tY = 0 \forall t \in \mathbb{Z}\).

Then \(EZ_t = EX_t + EY = \mu\) for all \(t \in \mathbb{Z}\) and
\[
E(Z_{s+t} - \mu)(Z_t - \mu) = R_X(s) + \sigma^2 := R_Z(s),
\]
from which we get that \(\{Z_t, t \in \mathbb{Z}\}\) is weakly stationary. However, it is not mean square ergodic, since, as \(n \to \infty\),
\[
\frac{1}{n} \sum_{t=1}^{n} R_Z(t) = \frac{1}{n} \sum_{t=1}^{n} R_X(t) + \sigma^2 \to \sigma^2 > 0.
\]

**Theorem 40.** Let \(\{X_t, t \in \mathbb{Z}\}\) be a real-valued stationary sequence with mean \(\mu\) and autocovariance function \(R\), such that \(\sum_{t=-\infty}^{\infty} |R(t)| < \infty\). Then, as \(n \to \infty\)
\[
\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t \to \mu \text{ in mean square,} \tag{48}
\]
\[
n \text{var} \overline{X}_n \to \sum_{k=-\infty}^{\infty} R(k). \tag{49}
\]

**Proof.** 1. \(\sum_{k=-\infty}^{\infty} |R(k)| < \infty \Rightarrow R(k) \to 0 \text{ as } k \to \infty\), thus \(\frac{1}{n} \sum_{k=1}^{n} R(k) \to 0\) as \(n \to \infty\) and assertion (48) follows from Theorem 39 39.
\[ \text{var} \overline{X}_n = \text{var} \left( \frac{1}{n} \sum_{k=1}^{n} X_k \right) \]

\[ = \frac{1}{n^2} \left[ \sum_{k=1}^{n} \text{var} X_k + \sum_{1 \leq j \neq k \leq n} \text{cov} (X_j, X_k) \right] \]

\[ = \frac{1}{n^2} \left[ nR(0) + 2 \sum_{j=1}^{n-1} (n - j)R(j) \right] \]

\[ = \frac{1}{n} \left[ R(0) + 2 \sum_{j=1}^{n-1} \left( 1 - \frac{j}{n} \right) R(j) \right] \]

\[ = \frac{1}{n} \sum_{j=-n+1}^{n-1} \left( 1 - \frac{|j|}{n} \right) R(j). \quad (50) \]

(For a real-valued sequence, \( R(k) = R(-k). \))

Thus,

\[ n \text{var} \overline{X}_n = \sum_{j=-n+1}^{n-1} R(j) - 2 \sum_{j=1}^{n-1} jR(j). \]

Assertion (49) now follows from the assumptions of the theorem and from the Kronecker lemma. \( \square \)

**Definition 38.** A stationary mean square continuous process \( \{X_t, t \in \mathbb{R}\} \) with mean \( \mu \) is **mean square ergodic** if, as \( \tau \to \infty \),

\[ \frac{1}{\tau} \int_0^\tau X_t dt \to \mu \quad \text{in mean square}. \]

**Theorem 41.** A stationary, mean square continuous process \( \{X_t, t \in \mathbb{R}\} \) is mean square ergodic if and only if its autocovariance function satisfies condition

\[ \frac{1}{\tau} \int_0^\tau R(t) dt \to 0 \quad \text{as} \quad \tau \to \infty. \]
Theorem 42. Let \( \{X_t, t \in \mathbb{R}\} \) be a real-valued stationary, mean square continuous process, with mean \( \mu \) and autocovariance function \( R \), such that \( \int_{-\infty}^{\infty} |R(t)| dt < \infty \).

Then, as \( \tau \to \infty \)
\[
X_\tau = \frac{1}{\tau} \int_0^\tau X_t dt \to \mu \text{ in mean square}, \tag{51}
\]
\[
\tau \text{ var } X_\tau \to \int_{-\infty}^{\infty} R(t) dt. \tag{52}
\]

Example 34. Let \( \{X_t, t \in \mathbb{R}\} \) be a stationary centered stochastic process with the autocovariance function
\[
R(t) = ce^{-\alpha|t|}, \quad t \in \mathbb{R}, \quad \alpha > 0, \quad c > 0.
\]

The process is mean square continuous. Moreover,
\[
\frac{1}{\tau} \int_0^\tau R(t) dt = \frac{c}{\tau} \int_0^\tau e^{-\alpha t} dt = \frac{c}{\tau} \frac{1 - e^{-\alpha \tau}}{\alpha} \to 0
\]
as \( \tau \to \infty \), the process \( \{X_t, t \in \mathbb{R}\} \) is mean square ergodic and \( \tau \text{ var } X_\tau \to \frac{2c}{\alpha} \).

9.2 Central limit theorems

Some preliminary asymptotic results

Theorem 43. (Cramér-Slutsky Theorem) Let \( \{X_n, n \in \mathbb{N}\}, \{Y_n, n \in \mathbb{N}\} \) be sequences of random variables and \( X \) be a random variable such that, as \( n \to \infty \), \( X_n \xrightarrow{D} X \), \( Y_n \xrightarrow{P} 0 \). Then \( X_n + Y_n \xrightarrow{D} X \) as \( n \to \infty \).

Theorem 44. Let \( \{\xi_n, n \in \mathbb{N}\}, \{S_{kn}, n \in \mathbb{N}, k \in \mathbb{N}\}, \{\psi_k, k \in \mathbb{N}\} \) and \( \psi \) be random variables such that

1. \( S_{kn} \xrightarrow{D} \psi_k \), \( n \to \infty \), for all \( k = 1, 2, \ldots \),
2. \( \psi_k \xrightarrow{D} \psi \), \( k \to \infty \),
3. \( \lim_{k \to \infty} \lim_{n \to \infty} P(|\xi_n - S_{kn}| > \epsilon) = 0 \) for all \( \epsilon > 0 \).

Then \( \xi_n \xrightarrow{D} \psi \) as \( n \to \infty \).

Proof. Brockwell, Davis (1991), Proposition 6.3.9. \( \square \)
Theorem 45. (Lévy-Lindeberg CLT) Let \( \{Y_t, t \in \mathbb{Z}\} \) be a sequence of independent identically distributed random variables with mean \( \mu \) and finite positive variance \( \sigma^2 \). Let 
\[
Y_n = \frac{1}{n} \sum_{j=1}^{n} Y_j.
\]
Then, as \( n \to \infty \)
\[
\sqrt{n} \frac{Y_n - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).
\]

Theorem 46. (Cramér-Wold Theorem) Let \( \mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots \), be \( k \)-dimensional random vectors. Then 
\[
\mathbf{X}_n \xrightarrow{d} \mathbf{X} \quad \text{as} \quad n \to \infty
\]
if and only if for every \( c \in \mathbb{R}_k \)
\[
c' \mathbf{X}_n \xrightarrow{d} c' \mathbf{X} \quad \text{pro} \quad n \to \infty.
\]

Central limit theorems for stationary sequences

Theorem 47. Let \( \{X_t, t \in \mathbb{Z}\} \) be a random sequence defined by
\[
X_t = \mu + \sum_{j=0}^{m} b_j Y_{t-j},
\]
where \( \mu \in \mathbb{R} \), \( \{Y_t, t \in \mathbb{Z}\} \) is a strict white noise, i.e., a sequence of independent identically distributed (iid) random variables with zero mean and finite positive variance \( \sigma^2 \). Let \( b_0 = 1 \) and \( b_1, \ldots, b_m \) be real-valued constants such that \( \sum_{j=0}^{m} b_j \neq 0 \). Then, as \( n \to \infty \),
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{d} N(0, \Delta^2),
\]
where \( \Delta^2 = \sigma^2 \left( \sum_{j=0}^{m} b_j \right)^2 \).

Proof.
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \sum_{j=0}^{m} b_j Y_{t-j} \right)
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{b_1}{\sqrt{n}} \sum_{t=1}^{n} Y_{t-1} + \ldots + \frac{b_m}{\sqrt{n}} \sum_{t=1}^{n} Y_{t-m}
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{b_1}{\sqrt{n}} \left( \sum_{t=1}^{n} Y_t + Y_0 - Y_n \right) + \ldots
+ \frac{b_m}{\sqrt{n}} \left( \sum_{t=1}^{n} Y_t + \sum_{k=-m+1}^{0} Y_k - \sum_{j=n-m+1}^{n} Y_j \right)
= \left( \sum_{j=0}^{m} b_j \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{1}{\sqrt{n}} \xi_n,
\]

69
where
\[ \xi_n = \sum_{s=1}^{m} Y_{1-s} \left( \sum_{j=s}^{m} b_j \right) - \sum_{s=0}^{m-1} Y_{n-s} \left( \sum_{j=s+1}^{m} b_j \right) \]
is a finite linear combination of iid random variables \( Y_0, Y_1, \ldots, Y_{-m+1}, Y_n, Y_{n-1}, \ldots, Y_{n-m+1} \) with zero mean and variance \( \sigma^2 \).

According to Theorem 45, \( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \overset{d}{\rightarrow} \mathcal{N}(0, \sigma^2) \) as \( n \rightarrow \infty \). From here
\[ \left( \sum_{j=0}^{m} b_j \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \overset{d}{\rightarrow} \mathcal{N}(0, \Delta^2), \]where \( \Delta^2 = \sigma^2 \left( \sum_{j=0}^{m} b_j \right)^2 \). (55)

Now, using Theorem 43 it suffices to prove that \( \frac{1}{\sqrt{n}} \xi_n \overset{p}{\rightarrow} 0 \) as \( n \rightarrow \infty \):
\[ P \left( \left| \frac{1}{\sqrt{n}} \xi_n \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \mathbb{E} \left( \frac{1}{n} \xi_n^2 \right) = \frac{1}{\epsilon^2} \frac{\sigma^2 \cdot \text{const}}{n} \rightarrow 0. \]

\[ \Box \]

**Theorem 48.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a random sequence such that
\[ X_t = \mu + \sum_{j=0}^{\infty} b_j Y_{t-j}, \]
where \( \mu \in \mathbb{R} \), \( \{Y_t, t \in \mathbb{Z}\} \) is a sequence of iid random variables with zero mean and finite positive variance \( \sigma^2 \). Let \( b_j, j \in \mathbb{N}_0 \), be real-valued constants such that \( \sum_{j=0}^{\infty} |b_j| < \infty \), \( \sum_{j=0}^{\infty} b_j \neq 0 \) and \( b_0 = 1 \). Then, as \( n \rightarrow \infty \),
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \overset{d}{\rightarrow} \mathcal{N}(0, \Delta^2), \]
where \( \Delta^2 = \sigma^2 \left( \sum_{j=0}^{\infty} b_j \right)^2 \).

**Proof.** Choose \( k \in \mathbb{N} \). Then
\[ X_t - \mu = \sum_{j=0}^{k} b_j Y_{t-j} + \sum_{j=k+1}^{\infty} b_j Y_{t-j} =: U_{kt} + V_{kt}, \]
thus
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_{kt} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_{kt}. \]
If we denote
\[
\xi_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu), \quad S_{kn} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_{kt}, \quad D_{kn} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_{kt},
\]
we have
\[
\xi_n = S_{kn} + D_{kn}.
\]
From Theorem 47 we have, as \( n \to \infty \) and every \( k \in \mathbb{N} \)
\[
S_{kn} \xrightarrow{p} \psi_k,
\]
where \( \psi_k \sim \mathcal{N}(0, \Delta^2_k) \), \( \Delta^2_k = \sigma^2 \left( \sum_{j=0}^{k} b_j \right)^2 \).
Further, from the assumptions of the theorem it follows that
\[
\Delta^2_k = \sigma^2 \left( \sum_{j=0}^{k} b_j \right)^2 \to \sigma^2 \left( \sum_{j=0}^{\infty} b_j \right)^2 = \Delta^2,
\]
as \( k \to \infty \), and thus
\[
\psi_k \xrightarrow{D} \mathcal{N}(0, \Delta^2).
\]
According to the Chebyshev inequality
\[
P(|\xi_n - S_{kn}| > \varepsilon) = P(|D_{kn}| > \varepsilon) \leq \frac{1}{\varepsilon^2 \text{var} D_{kn}}
= \frac{1}{\varepsilon^2 \text{var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_{kt} \right)}.
\]
From the assumption \( \sum_{j=0}^{\infty} |b_j| < \infty \) and Theorem 34 it follows that for any \( k \in \mathbb{N} \), \( \{V_{kt}, t \in \mathbb{Z}\} \) is the centered stationary sequence with the autocovariance function
\[
R_V(t) = \sigma^2 \sum_{j=k+1}^{\infty} b_j b_{j+t}.
\]
Using formula (50) we can write
\[ P(|\xi_n - S_{kn}| > \epsilon) \leq \frac{1}{\epsilon^2} \text{var}\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_{kt} \right) \]

\[ = \frac{1}{\epsilon^2} \sum_{j=-n+1}^{n-1} R_V(j) \left( 1 - \frac{|j|}{n} \right) \leq \frac{1}{\epsilon^2} \sum_{j=-n+1}^{n-1} |R_V(j)| \]

\[ = \frac{1}{\epsilon^2} \left[ R_V(0) + 2 \sum_{j=1}^{n-1} |R_V(j)| \right] \]

\[ = \frac{\sigma^2}{\epsilon^2} \left[ \sum_{j=k+1}^{\infty} b_j^2 + 2 \sum_{j=1}^{n-1} \sum_{\nu=k+1}^{\infty} b_{\nu} b_{\nu+j} \right] \]

\[ \leq \frac{\sigma^2}{\epsilon^2} \left[ \sum_{j=k+1}^{\infty} b_j^2 + 2 \sum_{j=1}^{n-1} \sum_{\nu=k+1}^{\infty} |b_{\nu}| |b_{\nu+j}| \right] \]

\[ \leq \frac{\sigma^2}{\epsilon^2} \left[ \sum_{j=k+1}^{\infty} |b_j|^2 + 2 \sum_{j=1}^{n-1} \sum_{\nu=k+1}^{\infty} |b_{\nu}| \sum_{j=1}^{\infty} |b_{\nu+j}| \right] = \frac{\sigma^2}{\epsilon^2} \left( \sum_{j=k+1}^{\infty} |b_j| \right)^2, \]

so that

\[ \lim_{k \to \infty} \lim_{n \to \infty} P(|\xi_n - S_{kn}| > \epsilon) \leq \lim_{k \to \infty} \frac{\sigma^2}{\epsilon^2} \left( \sum_{j=k+1}^{\infty} |b_j| \right)^2 = 0 \quad (58) \]

for any \( \epsilon > 0. \)

Combining this result with (56) and (57) we can see that the assumptions of Theorem 44 are met and thus, as \( n \to \infty, \)

\[ \xi_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{d} \mathcal{N}(0, \Delta^2). \]

\[ \square \]

**Example 35.** Let us consider a sequence \( \{X_t, t \in \mathbb{Z}\}, \) defined by

\[ X_t = \mu + Z_t, \quad Z_t = aZ_{t-1} + Y_t, \]

where \( \mu \in \mathbb{R}, |a| < 1 \) and \( \{Y_t, t \in \mathbb{Z}\} \) is a strict white noise with finite variance \( \sigma^2 > 0. \) The assumption \( |a| < 1 \) implies that \( \sum_{j=0}^{\infty} |a|^j < \infty, \) thus

\[ X_t = \mu + \sum_{j=0}^{\infty} a^j Y_{t-j}, \quad t \in \mathbb{Z}. \]
Since \( \sum_{j=0}^{\infty} a^j \neq 0 \), it holds, as \( n \to \infty \)
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{d} \mathcal{N}(0, \Delta^2), \quad \Delta^2 = \sigma^2 \frac{1}{(1-a)^2}.
\]

For large \( n \), \( \overline{X}_n \approx \mathcal{N} \left( \mu, \frac{\sigma^2}{m(1-a)^2} \right) \).

**Definition 39.** We say that a strictly stationary sequence \( \{X_t, t \in \mathbb{Z}\} \) is \( m \)-dependent, where \( m \in \mathbb{N}_0 \) is a given number, if for every \( t \in \mathbb{Z} \), the sets of random variables \((\ldots, X_{t-1}, X_t)\) and \((X_{t+m+1}, X_{t+m+2}, \ldots)\) are independent.

**Remark 9.** A sequence of iid random variables is \( m \)-dependent with \( m = 0 \).

**Example 36.** An MA\((m)\) sequence generated from a strict white noise is the sequence of \( m \)-dependent random variables.

**Example 37.** Let \( \{Y_t, t \in \mathbb{Z}\} \) be a strict white noise. Define \( \{X_t, t \in \mathbb{Z}\} \) by
\[
X_t = Y_t Y_{t+m}, \quad t \in \mathbb{Z},
\]
for some \( m \in \mathbb{N} \). Then
\[
\begin{align*}
\text{EX}_t &= E(Y_t Y_{t+m}) = 0, \\
\text{EX}_s X_t &= E(Y_t Y_{t+m} Y_s Y_{s+m}) = 0 \quad \text{pro} \ t \neq s.
\end{align*}
\]
In this case, \( X_t \) are mutually uncorrelated but not independent. They are \( m \)-dependent.

**Theorem 49.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a real-valued strictly stationary centered \( m \)-dependent random sequence with finite second-order moments and autocovariance function \( R \), such that
\[
\Delta_m^2 = \sum_{k=-m}^{m} R(k) \neq 0.
\]

Then, as \( n \to \infty \),
\[
\begin{align*}
\text{var} \overline{X}_n &\to \Delta_m^2, \\
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t &\xrightarrow{d} \mathcal{N}(0, \Delta_m^2).
\end{align*}
\]
Proof. 1. Since the sequence \( \{X_t, t \in \mathbb{Z}\} \) is strictly stationary with finite second-order moments, it is weakly stationary. From \( m \)-dependence it follows that \( R(k) = 0 \) for \( |k| > m \). According to Theorem 39 we have

\[
\lim_{n \to \infty} n \text{var} X_n = \sum_{k=-\infty}^{\infty} R(k) = \sum_{k=-m}^{m} R(k) = \Delta_m^2.
\]

2. Let \( k > 2m \) and \( n = k \cdot r \), where \( k \in \mathbb{N}, r \in \mathbb{N} \). Then

\[
(X_1, \ldots, X_n) = (U_1, V_1, U_2, V_2, \ldots, U_r, V_r),
\]

\[
U_j = (X_{(j-1)k+1}, \ldots, X_{jk-m}), \quad j = 1, \ldots, r,
\]

\[
V_j = (X_{jk-m+1}, \ldots, X_{jk}), \quad j = 1, \ldots, r.
\]

\( U_1, \ldots, U_r \) are mutually independent (it follows from \( m \)-dependence and the assumption \( k > 2m \)) and identically distributed (from strict stationarity). Similarly, \( V_1, \ldots, V_r \) are iid. Thus,

\[
\sum_{t=1}^{n} X_i = \sum_{j=1}^{r} S_j + \sum_{j=1}^{r} T_j,
\]

\( S_j, j = 1, \ldots, r, \) are iid (\( S_j \) is the sum of elements of the vector \( U_j \))

\( T_j, j = 1, \ldots, r, \) are iid (the sum of elements of the vectors \( V_j \)).

For \( k > 2m \) it holds \( \mathbb{E}S_1 = 0, \ \mathbb{E}T_1 = 0 \) and

\[
\text{var} S_1 = \text{var} (X_1 + \cdots + X_{k-m}) = \sum_{\nu=-m}^{m} (k-m-|\nu|)R(\nu) = \Delta_{mk}^2.
\]

Similarly, utilizing the strict stationarity

\[
\text{var} T_1 = \text{var} (X_{k-m+1} + \cdots + X_k) = \text{var} (X_1 + \cdots + X_m)
\]

\[
= \sum_{\nu=-m+1}^{m-1} (m-|\nu|)R(\nu) = \delta_m^2.
\]

Now, we can write

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_i := \xi_n = S_{kn} + D_{kn}, \quad (61)
\]

where

\[
S_{kn} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n/k} S_j = \frac{1}{\sqrt{k}} \frac{1}{\sqrt{r}} \sum_{j=1}^{r} S_j, \quad (62)
\]

\[
D_{kn} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n/k} T_j = \frac{1}{\sqrt{k}} \frac{1}{\sqrt{r}} \sum_{j=1}^{r} T_j. \quad (63)
\]
From the Lévy-Lindeberg theorem, as \( r \to \infty \),
\[
\frac{1}{\sqrt{r}} \sum_{j=1}^{r} S_j \xrightarrow{d} \mathcal{N}(0, \Delta_{mk}^2).
\]

For a fixed \( k \) and \( r \to \infty \) also \( n \to \infty \), so that
\[
S_{kn} \xrightarrow{d} \psi_k,
\]
where \( \psi_k \) has the normal distribution \( \mathcal{N}\left(0, \frac{\Delta_{mk}^2}{k}\right) \). As \( k \to \infty \),
\[
\frac{\Delta_{mk}^2}{k} \to \sum_{j=-m}^{m} R(j) = \Delta_m^2,
\]
\[
\psi_k \xrightarrow{d} \mathcal{N}(0, \Delta_m^2).
\]

From the Chebyshev inequality
\[
P(|\xi_n - S_{kn}| > \epsilon) = P(|D_{kn}| > \epsilon) \leq \frac{1}{\epsilon^2} \cdot \frac{1}{n} \text{var}\left(\sum_{j=1}^{r} T_j\right)
= \frac{1}{\epsilon^2} \cdot \frac{1}{k} \text{var}T_1 = \frac{1}{\epsilon^2 k} \delta_m^2.
\]
Thus,
\[
\lim_{k \to \infty} \lim_{n \to \infty} P(|\xi_n - S_{kn}| > \epsilon) = 0
\]
and the proof follows from (64), (65), (66) and Theorem 44. \( \square \)

**Example 38.** Consider
\[
X_t = \mu + Y_t + a_1 Y_{t-1} + a_2 Y_{t-2}, \quad t \in \mathbb{Z},
\]
where \( Y_t \) are iid, \( EY_t = 0, \text{var}Y_t = \sigma^2 > 0 \).

The sequence \( \{X_t, t \in \mathbb{Z}\} \) is strictly stationary and \( m \)-dependent, \( m = 2 \). The autocovariance function of \( \{X_t, t \in \mathbb{Z}\} \) takes values
\[
R(0) = \sigma^2(1 + a_1^2 + a_2^2),
R(1) = \sigma^2(a_1 + a_1 a_2) = R(-1),
R(2) = \sigma^2 a_2 = R(-2),
R(k) = 0, \quad |k| > 2.
\]

Therefore
\[
\Delta_m^2 = \sum_{k=-m}^{m} R(k) = R(0) + 2R(1) + 2R(2) = \sigma^2(1 + a_1 + a_2)^2.
\]

From the previous theorem, \( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{d} \mathcal{N}(0, \Delta_m^2) \), provided \( \Delta_m^2 \neq 0 \).
Example 39. Let \( \{Y_t, t \in \mathbb{Z}\} \) be a sequence of iid random variables, \( EY_t = 0, \) \( var Y_t = \sigma^2, \) \( EY_t^4 < \infty. \) Prove that for every \( k > 0 \) as \( n \to \infty \) it holds

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t^2 - \sigma^2) \xrightarrow{D} \mathcal{N}(0, \tau^2),
\]

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t Y_{t+k} \xrightarrow{D} \mathcal{N}(0, \sigma^4),
\]

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} Y_t Y_{t+k} \xrightarrow{D} \mathcal{N}(0, \sigma^4),
\]

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{D} \mathcal{N}_k(0, \sigma^4 I),
\]

where \( \tau^2 = var Y_t^2, \) \( X_t = (Y_t Y_{t+1}, \ldots, Y_t Y_{t+k})' \) and \( I \) is the identity matrix of order \( k. \)

Solution.

1. \( Y_t^2 \) are iid, \( EY_t^2 = \sigma^2, \) \( var Y_t^2 = \tau^2. \)

The Central limit theorem (Theorem 45) \( \Rightarrow \)

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t^2 - \sigma^2) \xrightarrow{D} \mathcal{N}(0, \tau^2).
\]

2. Denote \( X_t := Y_t Y_{t+k} \) for \( k > 0. \)

The sequence \( \{X_t, t \in \mathbb{Z}\} \) is strictly stationary, \( EX_t = 0, \) \( EX_t^2 = \sigma^4, \) \( X_t \) are mutually uncorrelated but \( k \)-dependent. By Theorem 49,

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{D} \mathcal{N}(0, \Delta^2_k),
\]

where \( \Delta^2_k = \sum_{j=-k}^{k} RX(j) = \sigma^4. \)

3.

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} Y_t Y_{t+k} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t Y_{t+k} - \frac{1}{\sqrt{n}} \sum_{t=n-k+1}^{n} Y_t Y_{t+k}.
\]

From step 2, as \( n \to \infty, \)

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t Y_{t+k} \xrightarrow{D} \mathcal{N}(0, \sigma^4).
\]
Form the Chebyshev inequality, as \( n \to \infty \),

\[
P\left( \left| \frac{1}{\sqrt{n}} \sum_{t=n-k+1}^{n} Y_{t} Y_{t+k} \right| > \epsilon \right) = P\left( \left| \frac{1}{\sqrt{n}} \sum_{t=n-k+1}^{n} X_{t} \right| > \epsilon \right)
\leq \frac{1}{\epsilon^2 n} \mathbb{E} \left( \sum_{t=n-k+1}^{n} X_{t} \right)^2 = \frac{1}{\epsilon^2 n} \sum_{t=n-k+1}^{n} \mathbb{E} X_{t}^2 = \frac{1}{\epsilon^2} \frac{k \sigma^4}{n} \to 0
\]

since \( k \) is fixed.

4. Define \( Z_t := c'X_t, t \in \mathbb{Z}, c \in \mathbb{R}_k \).

- Random vectors \( X_t \) have zero mean and the variance matrix \( \sigma^4 I \) and are mutually uncorrelated.
- Random variables \( Z_t \) are centered, with the variance \( \sigma^4 c'Ic \), uncorrelated and \( k \)-dependent
- \( \{Z_t, t \in \mathbb{Z}\} \) is strictly stationary.

By Theorem 49,

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t \xrightarrow{n} N(0, \Delta_k^2),
\]

where \( \Delta_k^2 = \sum_{j=-k}^{k} R_Z(j) = \sigma^4 c'Ic \). From here the final result follows when we apply Theorem 46 and properties of normal distribution.

10 Prediction in time domain

10.1 Projection in Hilbert space

**Definition 40.** Let \( H \) be a Hilbert space with the inner product \( \langle \cdot, \cdot \rangle \) and the norm \( ||.|| \). We say that two elements \( x, y \in H \) are orthogonal (perpendicular) if \( \langle x, y \rangle = 0 \). We write \( x \perp y \).

Let \( M \subset H \) be a subset of \( H \). We say that an element \( x \in H \) is orthogonal to \( M \), if it is orthogonal to every element of \( M \), i.e., \( \langle x, y \rangle = 0 \) for every \( y \in M \). We write \( x \perp M \).

The set \( M^\perp = \{ y \in H : y \perp M \} \) is called to be the orthogonal complement of the set \( M \).

**Theorem 50.** Let \( H \) be a Hilbert space, \( M \subset H \) any subset. Then \( M^\perp \) is a closed subspace of \( H \).

**Proof.** The null element \( 0 \in M^\perp \), since \( \langle 0, x \rangle = 0 \) for every \( x \in M \).

The linearity of the inner product implies that any linear combination of two elements of \( M^\perp \) is an element of \( M^\perp \).

Continuity of the inner product implies that any limit of a sequence of elements of \( M^\perp \) is an element of \( M^\perp \).

\[ \square \]
**Theorem 51.** (Projection Theorem) Let $M$ be a closed subspace of a Hilbert space $H$. Then for every element $x \in H$ there exists a unique decomposition $x = \hat{x} + (x - \hat{x})$, such that $\hat{x} \in M$ and $x - \hat{x} \in M^\perp$. Further

$$||x - \hat{x}|| = \min_{y \in M} ||x - y||$$  \hspace{1cm} (67)

and

$$||x||^2 = ||\hat{x}||^2 + ||x - \hat{x}||^2.$$  \hspace{1cm} (68)

**Proof.** Rudin (2003), Theorem 4.11., or Brockwell and Davis (1992), Theorem 2.3.1.  \hspace{1cm} $\square$

The element $\hat{x} \in M$ with property (67) is called to be the *orthogonal projection* of $x$ onto the subspace $M$. The mapping $P_M : H \rightarrow M$ such that $P_Mx \in M$ and $(I - P_M)x \in M^\perp$ where $I$ is the identity mapping, is called the *projection mapping*. Obviously, for any $x \in H$

$$x = P_Mx + (x - P_Mx) = P_Mx + (I - P_M)x.$$  \hspace{1cm} (69)

**Theorem 52.** Let $H$ be a Hilbert space, $P_M$ the projection mapping of $H$ onto a closed subspace $M$. It holds:

1. For every $x, y \in H$ and any $\alpha, \beta \in \mathbb{C}$, $P_M(\alpha x + \beta y) = \alpha P_Mx + \beta P_My$.
2. If $x \in M$, then $P_Mx = x$.
3. If $x \in M^\perp$, then $P_Mx = 0$.
4. If $M_1, M_2$ are closed subspaces of $H$ such that $M_1 \subseteq M_2$, then $P_{M_1}x = P_{M_1}(P_{M_2}x)$ for every $x \in H$.
5. If $x_n, x$ are elements of $H$ such that $||x_n - x|| \rightarrow 0$ as $n \rightarrow \infty$, then $||P_Mx_n - P_Mx|| \rightarrow 0$.

**Proof.**

1. By using (69) we get

$$\alpha x + \beta y = \alpha(P_Mx + (x - P_Mx)) + \beta(P_My + (y - P_My))$$

$$= \alpha P_Mx + \beta P_My + \alpha(x - P_Mx) + \beta(y - P_My).$$

Obviously,

$$\alpha P_Mx + \beta P_My \in M, \quad \alpha(x - P_Mx) + \beta(y - P_My) \in M^\perp$$

since $M$ and $M^\perp$ are linear subspaces, and thus, $\alpha P_Mx + \beta P_My = P_M(\alpha x + \beta y)$. 78
2. The uniqueness of decomposition (69) implies assertion 2.

3. The uniqueness of decomposition (69) implies assertion 3.

4. Since $x = P_{M_2}x + (x - P_{M_2}x)$, $P_{M_2}x \in M_2$, $x - P_{M_2}x \in M_2^\perp$, we have $P_{M_1}x = P_{M_1}(P_{M_2}x) + P_{M_1}(x - P_{M_2}x)$. Thus,

$$P_{M_1}(P_{M_2}x) \in M_1, \text{ and } M_2^\perp \subseteq M_1^\perp \Rightarrow P_{M_1}(x - P_{M_2}x) = 0.$$ 

5. From the linearity of the projection mapping and equation (68)

$$||P_{M_2}x_n - P_{M_2}x||^2 = ||P_{M_2}(x_n - x)||^2 \leq ||x_n - x||^2.$$ 

\[ \square \]

### 10.2 Prediction based on finite history

Problem: Let us consider random variables $X_1, \ldots, X_n$ with zero mean and finite second moments. Utilizing observations $X_1, \ldots, X_n$ we want to forecast $X_{n+h}$, where $h > 0$. We want to approximate $X_{n+h}$ by a measurable function $g(X_1, \ldots, X_n)$ (prediction) of observations $X_1, \ldots, X_n$ that minimizes

$$E|X_{n+h} - g(X_1, \ldots, X_n)|^2.$$ 

It is known that the best approximation is given by the conditional mean value

$$g(X_1, \ldots, X_n) = E(X_{n+h}|X_1, \ldots, X_n).$$

Indeed, if we denote $(X_1, \ldots, X_n)' = X_n$, we can write

$$E(X_{n+h} - g(X_n))^2$$

$$= E(X_{n+h} - E(X_{n+h}|X_n) + E(X_{n+h}|X_n) - g(X_n))^2$$

$$= E(X_{n+h} - E(X_{n+h}|X_n))^2 + E(E(X_{n+h}|X_n) - g(X_n))^2$$

$$+ 2E[(X_{n+h} - E(X_{n+h}|X_n)) (E(X_{n+h}|X_n) - g(X_n))] ,$$

where the last summand is

$$E[(X_{n+h} - E(X_{n+h}|X_n)) (E(X_{n+h}|X_n) - g(X_n))]$$

$$= E[E(X_{n+h} - E(X_{n+h}|X_n)) (E(X_{n+h}|X_n) - g(X_n)) | X_n]$$

$$= E[(E(X_{n+h}|X_n) - E(X_{n+h}|X_n)) (E(X_{n+h}|X_n) - g(X_n))] = 0.$$ 

79
Thus,
\[
E(X_{n+h} - g(X_n))^2
= E (X_{n+h} - E(X_{n+h}|X_n))^2 + E (E(X_{n+h}|X_n) - g(X_n))^2
\geq E (X_{n+h} - E(X_{n+h}|X_n))^2
\]
with equality for \( g(X_n) = E(X_{n+h}|X_n) \).

In the next, we will confine ourselves to linear functions of \( X_1, \ldots, X_n \). Then the problem to find the best linear approximation of \( X_{n+h} \) can be solved by using the projection method in a Hilbert space. The best linear prediction of \( X_{n+h} \) from \( X_1, \ldots, X_n \) will be denoted by \( \hat{X}_{n+h}(n) \).

**Direct method**

Let \( H := \mathcal{H}\{X_1, \ldots, X_n, \ldots, X_{n+h}\} \) be the Hilbert space generated by centered random variables \( X_1, \ldots, X_{n+h} \) and \( H^n_1 := \mathcal{H}\{X_1, \ldots, X_n\} \) be the Hilbert subspace generated by random variables \( X_1, \ldots, X_n \).

The best linear prediction of \( X_{n+h} \) is the random variable
\[
\hat{X}_{n+h}(n) = \sum_{j=1}^n c_j X_j \in H^n_1, \tag{70}
\]
such that
\[
E|X_{n+h} - \hat{X}_{n+h}(n)|^2 = \|X_{n+h} - \hat{X}_{n+h}(n)\|^2
\]
takes minimum with respect to all linear combinations of \( X_1, \ldots, X_n \).

It means that
\[
\hat{X}_{n+h}(n) = P_{H^n_1}(X_{n+h}) \in H^n_1, \quad X_{n+h} - \hat{X}_{n+h}(n) \perp H^n_1 \tag{71}
\]
and the element \( \hat{X}_{n+h}(n) \) is determined uniquely due to the projection theorem.

Since the space \( H^n_1 \) is a linear span generated by \( X_1, \ldots, X_n \), condition (71) is satisfied if and only if
\[
X_{n+h} - \hat{X}_{n+h}(n) \perp X_j, \quad j = 1, \ldots, n,
\]
i.e., if and only if
\[
E(X_{n+h} - \hat{X}_{n+h}(n))X_j = 0, \quad j = 1, \ldots, n.
\]

Constants \( c_1, \ldots, c_n \) can be therefore obtained from the equations
\[
E(X_{n+h} - \sum_{k=1}^n c_k X_k)X_j = 0, \quad j = 1, \ldots, n. \tag{72}
\]
For \( X_1, \ldots, X_{n+h} \) supposed to be elements of a real-valued centered stationary sequence with the autocovariance function \( R \), system (72) is of the form
\[
\sum_{k=1}^{n} c_k R(k - j) = R(n + h - j), \quad j = 1, \ldots, n, \tag{73}
\]
or
\[
c_1 R(0) + c_2 R(1) + \cdots + c_n R(n - 1) = R(n + h - 1),
\]
\[
c_1 R(1) + c_2 R(0) + \cdots + c_n R(n - 2) = R(n + h - 2),
\]
\[
\vdots
\]
\[
c_1 R(n - 1) + c_2 R(n - 2) + \cdots + c_n R(0) = R(h).
\]
Equivalently, system (73) can be written in the form
\[
\Gamma_n c_n = \gamma_{nh}
\]
where \( c_n := (c_1, \ldots, c_n)' \), \( \gamma_{nh} := (R(n + h - 1), \ldots, R(h))' \) and
\[
\Gamma_n := \begin{pmatrix}
R(0) & R(1) & \cdots & R(n - 1) \\
R(1) & R(0) & \cdots & R(n - 2) \\
\vdots & \vdots & \ddots & \vdots \\
R(n - 1) & R(n - 2) & \cdots & R(0)
\end{pmatrix}.
\]
Provided that \( \Gamma_n^{-1} \) exists we get \( c_n = \Gamma_n^{-1} \gamma_{nh} \), thus
\[
\hat{X}_{n+h}(n) = \sum_{j=1}^{n} c_j X_j = c_n' X_n = \gamma_{nh} \Gamma_n^{-1} X_n. \tag{74}
\]
It is obvious that \( \Gamma_n = \text{var} (X_1, \ldots, X_n) = \text{var} X_n = E (X_n X_n') \).

The prediction error is
\[
\delta^2_h := E |X_{n+h} - \hat{X}_{n+h}(n)|^2 = \|X_{n+h} - \hat{X}_{n+h}(n)\|^2.
\]
By (68)
\[
\|X_{n+h}\|^2 = \|\hat{X}_{n+h}(n)\|^2 + \|X_{n+h} - \hat{X}_{n+h}(n)\|^2,
\]
so that
\[
\delta^2_h = \|X_{n+h}\|^2 - \|\hat{X}_{n+h}(n)\|^2. \tag{75}
\]
For a real-valued centered stationary sequence such that \( \Gamma_n \) is regular,
\[
\delta^2_h = \|X_{n+h}\|^2 - \|\hat{X}_{n+h}(n)\|^2 = E |X_{n+h}|^2 - E |\hat{X}_{n+h}(n)|^2
\]
\[
= R(0) - E (c_n' X_n)^2 = R(0) - c_n' E (X_n X_n') c_n
\]
\[
= R(0) - c_n' \Gamma_n c_n = R(0) - \gamma_{nh} \Gamma_n^{-1} \gamma_{nh}
\]
\[
= R(0) - \gamma_{nh} \Gamma_n^{-1} \gamma_{nh}. \tag{76}
\]
Theorem 53. Let \( \{X_t, t \in \mathbb{Z}\} \) be a real-valued centered stationary sequence with auto-covariance function \( R \), such that \( R(0) > 0 \) and \( R(k) \to 0 \) as \( k \to \infty \). Then the matrix \( \Gamma_n = \text{var}(X_1, \ldots, X_n) \) is regular for every \( n \in \mathbb{N} \).

Proof. We prove the theorem by contradiction: suppose that \( \Gamma_n \) is singular for an \( n \in \mathbb{N} \); then there is a nonzero vector \( c = (c_1, \ldots, c_n)' \) such that \( c'\Gamma_n c = 0 \) and for \( X_n = (X_1, \ldots, X_n)' \), \( c'X_n = 0 \) a.s. holds true, since \( E c'X_n = 0 \) and \( \text{var}(c'X_n) = c'\Gamma_n c = 0 \).

Thus there exists a positive integer \( 1 \leq r < n \) and constants \( a_1, \ldots, a_r \) such that \( \Gamma_r \) is regular and

\[
X_{r+1} = \sum_{j=1}^r a_j X_j.
\]

By stationarity of \( \{X_t, t \in \mathbb{Z}\} \),

\[
\text{var}(X_1, \ldots, X_r) = \cdots = \text{var}(X_h, \ldots, X_{h+r-1}) = \Gamma_r.
\]

From here, for any \( h \geq 1 \)

\[
X_{r+h} = \sum_{j=1}^r a_j X_{j+h-1}.
\]

For every \( n \geq r + 1 \) there are constants \( a_1^{(n)}, \ldots, a_r^{(n)} \) such that \( X_n = \sum_{j=1}^r a_j^{(n)} X_j = a^{(n)} X_r \), where \( a^{(n)} = (a_1^{(n)}, \ldots, a_r^{(n)})' \) and \( X_r = (X_1, \ldots, X_r)' \),

\[
\text{var} X_n = a^{(n)}' \text{var} X_r a^{(n)} = a^{(n)}' \Gamma_r a^{(n)} = R(0) > 0.
\]

The matrix \( \Gamma_r \) is positive definite, therefore there exists a decomposition \( \Gamma_r = P \Lambda P' \), where \( \Lambda \) is a diagonal matrix with the eigenvalues of the matrix \( \Gamma_r \), on the diagonal and \( PP' = I \) is the identity matrix. Since \( \Gamma_r \) is positive definite, all its eigenvalues are positive; wlog assume that \( 0 < \lambda_1 \leq \cdots \leq \lambda_r \). Then

\[
R(0) = a^{(n)}' P \Lambda P' a^{(n)} \geq \lambda_1 a^{(n)}' PP' a^{(n)} = \lambda_1 \sum_{j=1}^r \left( a_j^{(n)} \right)^2,
\]

from which for every \( j = 1, \ldots, r \) it follows that \( \left( a_j^{(n)} \right)^2 \leq R(0)/\lambda_1 \), hence, \( |a_j^{(n)}| \leq C \) independently of \( n \), where \( C \) is a positive constant.

We also have

\[
0 < R(0) = E(X_n)^2 = E \left( X_n \sum_{j=1}^r a_j^{(n)} X_j \right) = \sum_{j=1}^r a_j^{(n)} E X_n X_j
\]

\[
= \sum_{j=1}^r a_j^{(n)} R(n-j) \leq \sum_{j=1}^r |a_j^{(n)}||R(n-j)|
\]

\[
\leq C \sum_{j=1}^r |R(n-j)|.
\]
The last expression converges to zero as $n \to \infty$ due to the assumption $R(n) \to 0$ as $n \to \infty$, but this contradicts to the assumption $R(0) > 0$. Thus, we conclude that the matrix $\Gamma_n$ is regular for every $n \in \mathbb{N}$.

**Recursive methods**

Let us introduce the following notation.

- Denote by $H^k_1 = \mathcal{H}\{X_1, \ldots, X_k\}$ the Hilbert space generated by $X_1, \ldots, X_k$.
- Denote by $\hat{X}_{k+1}, k \geq 1$, the one-step prediction of $X_{k+1}$,

$$\hat{X}_{k+1} := \hat{X}_{k+1}(k) = P_{H^k_1}(X_{k+1}), \quad \hat{X}_1 := 0.$$

Then

$$H^n_1 = \mathcal{H}\{X_1, \ldots, X_n\} = \mathcal{H}\{X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n\}.$$

**Lemma 4.** $X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n$ are orthogonal random variables.

**Proof.** Let $i < j$. Then $X_i \in H^i_1 \subseteq H^{j-1}_1$ and $\hat{X}_i \in H^{j-1}_1 \subset H^{i-1}_1$, so $X_i - \hat{X}_i \in H^{i-1}_1$. Further: $\hat{X}_j = P_{H^{j-1}_1}(X_j)$, therefore $X_j - \hat{X}_j \perp H^{j-1}_1$, and also

$$X_i - \hat{X}_i \perp X_j - \hat{X}_j.$$

The one step best linear prediction of $X_{k+1}$ computed from $X_1, \ldots, X_k$ thus can be written in the form

$$\hat{X}_{k+1} = \sum_{j=1}^{k} \theta_{k,j} \left(X_{k+1-j} - \hat{X}_{k+1-j}\right).$$

Error of the one-step prediction of $X_{k+1}$:

$$v_k = E|X_{k+1} - \hat{X}_{k+1}|^2 = ||X_{k+1} - \hat{X}_{k+1}||^2, \quad k \geq 0.$$

**Theorem 54 (Innovation algorithm).** Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued centered random sequence with autocovariance function $R(i, j)$, where matrix $(R(i, j))_{i,j=1}^n$ is regular for every $n$. Then the best linear prediction of $X_{n+1}$ computed from $X_1, \ldots, X_n$ is

$$\hat{X}_1 = 0,$$

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{n,j} \left(X_{n+1-j} - \hat{X}_{n+1-j}\right), \quad n \geq 1,$$

(77)
where for $k = 0, \ldots, n - 1$,

$$ v_0 = R(1, 1), \quad (78) $$

$$ \theta_{n,n-k} = \frac{1}{v_k} \left( R(n + 1, k + 1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right), \quad (79) $$

$$ v_n = R(n + 1, n + 1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \quad (80) $$

**Proof.** Define $\hat{X}_1 := 0$, then

$$ v_0 = E|X_1 - \hat{X}_1|^2 = E|X_1|^2 = R(1, 1). $$

Since $\hat{X}_{n+1} = P_{H_1^n} X_{n+1}$, it must be of the form (77). Multiply both sides of (77) by $X_k + 1 - \hat{X}_k + 1$ and take the mean value:

$$ E\hat{X}_{n+1}(X_k + 1 - \hat{X}_k + 1) = \sum_{j=1}^n \theta_{nj} E(X_{n+1-j} - \hat{X}_{n+1-j}) (X_k + 1 - \hat{X}_k + 1) = \theta_{n,n-k} E(X_k + 1 - \hat{X}_k + 1)^2 = \theta_{n,n-k} v_k. \quad (81) $$

From here

$$ \theta_{n,n-k} = \frac{1}{v_k} E\hat{X}_{n+1}(X_k + 1 - \hat{X}_k + 1) $$

$$ = \frac{1}{v_k} \left( R(n + 1, k + 1) - E\hat{X}_{n+1} \hat{X}_{k+1} \right). $$

Further, applying formula (77) to $\hat{X}_{k+1}$ and replacing $k$ by $k - j$ in formula (81) we get

$$ E\hat{X}_{n+1} \hat{X}_{k+1} = E\hat{X}_{n+1} \sum_{j=1}^k \theta_{kj} (X_{k+1-j} - \hat{X}_{k+1-j}) $$

$$ = \sum_{j=1}^k \theta_{kj} E\hat{X}_{n+1} (X_{k+1-j} - \hat{X}_{k+1-j}) $$

$$ = \sum_{j=1}^k \theta_{kj} \theta_{n,n-(k-j)} v_{k-j}. $$
Altogether we get
\[ \theta_{n,n-k} = \frac{1}{v_k} (R(n+1, k+1) - \sum_{j=1}^{k} \theta_{kj} \theta_{n,n-(k-j)v_{k-j}}) \]
\[ = \frac{1}{v_k} (R(n+1, k+1) - \sum_{\nu=0}^{k-1} \theta_{n,n-\nu} \theta_{k,k-\nu,v_{\nu}}). \]

Computation of \( v_n \):
\[ v_n = E|X_{n+1} - \hat{X}_{n+1}|^2 = E|X_{n+1}|^2 - E|\hat{X}_{n+1}|^2 \]
\[ = R(n+1, n+1) - E\left| \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) \right|^2 \]
\[ = R(n+1, n+1) - \sum_{j=1}^{n} \theta^2_{nj} E(X_{n+1-j} - \hat{X}_{n+1-j})^2 \]
\[ = R(n+1, n+1) - \sum_{j=1}^{n} \theta^2_{nj} v_{n-j} \]
\[ = R(n+1, n+1) - \sum_{\nu=0}^{n-1} \theta^2_{n,n-\nu} v_{\nu}. \]

Computational scheme of the innovation algorithm:
\[
\begin{array}{cccc}
\hat{X}_1 & v_0 \\
\theta_{11} & \hat{X}_2 & v_1 \\
\theta_{22} & \theta_{21} & \hat{X}_3 & v_2 \\
\theta_{33} & \theta_{32} & \theta_{31} & \hat{X}_4 & v_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\end{array}
\]

**Example 40.** We have observations \( X_1, \ldots, X_n \), of an MA(1) random sequence that is generated by
\[ X_t = Y_t + bY_{t-1}, \ Y_t \sim WN(0, \sigma^2), t \in \mathbb{Z} \]
Find $\hat{X}_{n+1}$ by using the innovation algorithm. We get

\[ \hat{X}_1 = 0, \]
\[ v_0 = R(0) = \sigma^2(1 + b^2), \]
\[ \theta_{11} = \frac{1}{v_0} R(1) = \frac{b}{1 + b^2}, \]
\[ \hat{X}_2 = \theta_{11}(X_1 - \hat{X}_1) = \theta_{11}X_1, \]
\[ v_1 = R(0) - \theta_{11}^2 v_0, \]
\[ \theta_{22} = \frac{1}{v_0} R(2) = 0, \]
\[ \theta_{21} = \frac{1}{v_1} (R(1) - \theta_{22} \theta_{11} v_0) = \frac{R(1)}{v_1}, \]
\[ \hat{X}_3 = \theta_{21}(X_2 - \hat{X}_2), \]
\[ v_2 = R(0) - \theta_{21}^2 v_1, \]

generally

\[ \theta_{nk} = 0, \quad k = 2, \ldots, n, \]
\[ \theta_{n1} = \frac{R(1)}{v_{n-1}}, \]
\[ \hat{X}_{n+1} = \theta_{n1}(X_n - \hat{X}_n), \]
\[ v_n = R(0) - \theta_{n1}^2 v_{n-1}. \]

**Example 41.** Consider an MA$(q)$ sequence. Then $R(k) = 0$ for $|k| > q$. By using the recursive computations we get

\[ \hat{X}_{n+1} = \sum_{j=1}^{\min(q,n)} \theta_{nj} \left( X_{n+1-j} - \hat{X}_{n+1-j} \right), \quad n \geq 1. \]

The $h$-step prediction, $h > 1$:

\[ \hat{X}_{n+h}(n) = P_{H_1^n}(X_{n+h}), \] where $H_1^n = \mathcal{H}(X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n)$. Since $H_1^n \subset H_1^{n+1} \subset H_1^{n+2}$. 

86
\[ \cdots \subset H_{n+h-1}^n, \text{ it follows from the properties of the projection mapping and (77), that} \]
\[ \hat{X}_{n+h}(n) = P_{H_1^n}(X_{n+h}) = P_{H_1^n}(P_{H_{n+h-1}^n}(X_{n+h})) = P_{H_1^n}(\hat{X}_{n+h}) \]
\[ = P_{H_1^n}\left( \sum_{j=1}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}) \right) \]
\[ = \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} P_{H_1^n}(X_{n+h-j} - \hat{X}_{n+h-j}) \]
\[ = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right), \quad (82) \]

since \( X_{n+h-j} - \hat{X}_{n+h-j} \perp X_1^n \text{ pro } j < h. \)

The \( h \)-step prediction error is
\[
\delta_h^2 = E|X_{n+h} - \hat{X}_{n+h}(n)|^2 = E|X_{n+h}|^2 - E|\hat{X}_{n+h}(n)|^2
\]
\[ = R(n+h, n+h) - E\left| \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right) \right|^2 \]
\[ = R(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 \mu_{n+h-j-1}. \]

**Example 42.** Let us again consider the MA(1) model.

We have shown that
\[ \hat{X}_{n+1} = P_{H_1^n}(X_{n+1}) = \theta_1(X_n - \hat{X}_n). \]

For \( h > 1 \) we have
\[ \hat{X}_{n+h}(n) = P_{H_1^n}(X_{n+h}) = P_{H_1^n}(\hat{X}_{n+h}) \]
\[ = P_{H_1^n}(\theta_{n+h-1,1}(X_{n+h-1} - \hat{X}_{n+h-1})) = 0, \]

since \( (X_{n+h-1} - \hat{X}_{n+h-1}) \perp H_1^n \text{ pro } h > 1. \)

**Innovation algorithm for an ARMA process**

Consider a causal ARMA\((p, q)\) process
\[ X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \ t \in \mathbb{Z}, \]
\[ Y_t \sim \text{WN}(0, \sigma^2). \] We want to find \( \hat{X}_{n+1} = P_{H_1^n}(X_{n+1}). \)
First, consider the following transformation:

\[
W_t = \begin{cases} \frac{1}{\sigma} X_t, & t = 1, 2, \ldots, m, \\ \frac{1}{\sigma} (X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}), & t > m, \end{cases}
\]  

(83)

\(m = \max(p, q)\).

Put \(\hat{X}_1 = 0, \hat{W}_1 = 0, \hat{W}_k = P_{H_{t-1}^k}(W_k)\). Then

\[
H^n_1 = \mathcal{H}(X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n) = \mathcal{H}(X_1, \ldots, X_n)
= \mathcal{H}(W_1, \ldots, W_n) = \mathcal{H}(W_1 - \hat{W}_1, \ldots, W_n - \hat{W}_n).
\]

- Application of the innovation algorithm to the sequence \(\{W_1, \ldots, W_n\}\):

\[
\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \leq n < m, \\ \sum_{j=1}^{q} \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}) + \sum_{j=1}^{p} \varphi_j X_{n-j+1}, & n \geq m, \end{cases}
\]

(84)

(for \(t > m, W_t \sim MA(q)\)).

- Apply projection mapping onto \(H_{t-1}^1\) to both sides of (83):

\[
\hat{W}_t = \begin{cases} \frac{1}{\sigma} \hat{X}_t, & t \leq m, \\ \frac{1}{\sigma} (\hat{X}_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}), & t > m. \end{cases}
\]

(85)

We can see that for \(t \geq 1\)

\[
W_t - \hat{W}_t = \frac{1}{\sigma} (X_t - \hat{X}_t), \quad \mathbb{E}|W_t - \hat{W}_t|^2 = \frac{1}{\sigma^2} v_{t-1} := w_{t-1}.
\]

Therefore it holds

\[
\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n < m \\ \sum_{j=1}^{q} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) + \sum_{j=1}^{p} \varphi_j X_{n-j+1}, & n \geq m, \end{cases}
\]

(86)

where coefficients \(\theta_{nj}\) a \(w_n\) are computed by applying the innovation algorithm to the sequence (83).

For this we need to compute the values of the autocovariance function of \(\{W_t\}\).

We know that \(\mathbb{E} X_t = 0\), thus \(\mathbb{E} W_t = 0\).
For the covariances $R_W(s, t) = EW_s W_t$ we get

\[
R_W(s, t) = \begin{cases} \frac{1}{\sigma^2} R_X(s - t), & 1 \leq s, t \leq m, \\ \frac{1}{\sigma^2} \left[ R_X(s - t) - \sum_{j=1}^{p} \varphi_j R_X(|s - t| - j) \right], & \min(s, t) \leq m, \quad m < \max(s, t) \leq 2m, \\ \frac{1}{\sigma^2} \sum_{j=0}^{q-|s-t|} \theta_j \theta_j + |s - t|, & s, t > m, \quad |s - t| \leq q, \\ 0, & \text{elsewhere} \end{cases}
\]

(87)

(we have put $\theta_0 = 1$).

**Innovation algorithm for an AR sequence**

Let us consider a causal AR($p$) process, i.e.,

\[ X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t, \quad t \in \mathbb{Z}, \quad Y_t \sim WN(0, \sigma^2) \]

- Transformation:
  \[
  W_t = \begin{cases} \frac{1}{\sigma} X_t, & 1 \leq t \leq p, \\ \frac{1}{\sigma} (X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}) = \frac{1}{\sigma} Y_t, & t > p. \end{cases}
  \]
  \[
  (88)
  \]

- Innovation algorithm applied to $W_1, \ldots, W_n$:
  \[
  \hat{W}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & n < p, \\ 0, & n \geq p. \end{cases}
  \]
  \[
  (89)
  \]

Again, $W_t - \hat{W}_t = \frac{1}{\sigma} (X_t - \hat{X}_t)$ for $t \geq 1$ and from here

\[
\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n < p, \\ \varphi_1 X_n + \varphi_2 X_{n-1} + \cdots + \varphi_p X_{n-p+1}, & n \geq p. \end{cases}
\]

(90)

The autocovariance function needed for the calculation of the coefficients $\theta_{nj}$ is

\[
R_W(s, t) = \begin{cases} \frac{1}{\sigma^2} R_X(s - t), & 1 \leq s, t \leq p, \\ 1, & t = s > p, \\ 0, & \text{elsewhere}. \end{cases}
\]

(91)

The one-step prediction error for $n \geq p$ is

\[
v_n = E|X_{n+1} - \hat{X}_{n+1}|^2 = EY_{n+1}^2 = \sigma^2.
\]

89
10.3 Prediction from infinite history

Problem: we know all the history $X_n, X_{n-1}, \ldots$, we want to forecast $X_{n+1}, X_{n+2}, \ldots$.

Solution: by projection in Hilbert spaces.

Consider Hilbert spaces $H = \mathcal{H}\{X_t, t \in \mathbb{Z}\}$ and $H_{-\infty}^n = \mathcal{H}\{X_n, X_{n-1}, \ldots\}$. Then the best linear prediction $\hat{X}_{n+h}(n)$ of $X_{n+h}$ from the infinite history $X_n, X_{n-1}, \ldots$ is the projection of $X_{n+h} \in H$ onto $H_{-\infty}^n$, we write $\hat{X}_{n+h}(n) = P_{H_{-\infty}^n}(X_{n+h})$ The one-step prediction is again $\hat{X}_{n+1} := \hat{X}_{n+1}$.

**Prediction in a causal AR($p$) process**

Consider model

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t, \quad t \in \mathbb{Z},$$  \hspace{1cm} (92)

where $\{Y_t, t \in \mathbb{Z}\}$ is WN$(0, \sigma^2)$, and assume that all the roots of the polynomial $\lambda^p - \varphi_1 \lambda^{p-1} - \cdots - \varphi_p$ are inside the unit circle. It means that $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process and $Y_t \perp X_s$ for all $t > s$.

The one-step prediction:

- $X_{n+1} = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + Y_{n+1}$
- $\varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} \in H_{-\infty}^n$;
- $Y_{n+1} \perp X_n, X_{n-1}, \cdots \Rightarrow Y_{n+1} \perp H_{-\infty}^n$, (from the linearity and the continuity of the inner product)

It means that

$$\hat{X}_{n+1} = P_{H_{-\infty}^n}(X_{n+1}) = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p}.$$  

The prediction error is

$$\mathbb{E}|X_{n+1} - \hat{X}_{n+1}|^2 = \mathbb{E}|Y_{n+1}|^2 = \sigma^2.$$  

The $h$-step prediction, $h > 1$:

$$\hat{X}_{n+h}(n) = P_{H_{-\infty}^n}(X_{n+h}) = P_{H_{-\infty}^n} \left( P_{H_{-\infty}^{n+h-1}}(X_{n+h}) \right)$$

$$= \hat{X}_{n+h}(n)$$

$$= P_{H_{-\infty}^n} \left( \varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p} \right)$$

$$= \varphi_1 [X_{n+h-1}] + \varphi_2 [X_{n+h-2}] + \cdots + \varphi_p [X_{n+h-p}],$$

where

$$[X_{n+j}] = \begin{cases} 
X_{n+j}, & j \leq 0 \\
\hat{X}_{n+j}(n), & j > 0.
\end{cases}$$
Example 43. Consider an AR(1) process generated by $X_t = \varphi X_{t-1} + Y_t$, where $|\varphi| < 1$ and $Y_t \sim \text{WN}(0, \sigma^2)$. If we know the whole history $X_n, X_{n-1}, \ldots$, we have $\hat{X}_{n+1} = \varphi X_n$.

For $h > 1$

$$\hat{X}_{n+h}(n) = \varphi [X_{n+h-1}] = \varphi \hat{X}_{n+h-1}(n) = \varphi^2 \hat{X}_{n+h-2}(n) = \ldots = \varphi^h X_n.$$ 

The prediction error is

$$E|X_{n+h} - \hat{X}_{n+h}|^2 = E|X_{n+h}|^2 - E|\hat{X}_{n+h}(n)|^2$$

$$= R_X(0) - E|\varphi^h X_n|^2 = R_X(0) (1 - \varphi^{2h})$$

$$= \sigma^2 \frac{1 - \varphi^{2h}}{1 - \varphi^2}.$$ 

Prediction in a causal and invertible ARMA$(p, q)$ process

Consider process

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \quad t \in \mathbb{Z},$$

(93)

where $Y_t \sim \text{WN}(0, \sigma^2)$.

- Causality: for any $t \in \mathbb{Z}$, $X_t = \sum_{j=0}^\infty c_j Y_{t-j}$, where $\sum_{j=0}^\infty |c_j| < \infty$ from which it follows that $Y_t \perp X_s$ for every $s < t$.

- Invertibility: for any $t \in \mathbb{Z}$, $Y_t = \sum_{j=0}^\infty d_j X_{t-j}$, where $\sum_{j=0}^\infty |d_j| < \infty$, or,

$$X_t = -\sum_{j=1}^\infty d_j X_{t-j} + Y_t$$

(94)

It holds

$$-\sum_{j=1}^\infty d_j X_{t-j} = \text{i.m.} \quad N \to \infty \left(-\sum_{j=1}^N d_j X_{t-j}\right) \in H_{-\infty}^{l-1},$$

$Y_t \perp H_{-\infty}^{l-1}$. From decomposition (94) it follows that the best linear prediction of $X_{n+1}$ based on the whole history $X_n, X_{n-1} \ldots$ is

$$\hat{X}_{n+1} = -\sum_{j=1}^\infty d_j X_{n+1-j}.$$ 

(95)

The prediction error is

$$E|X_{n+1} - \hat{X}_{n+1}|^2 = E|Y_{n+1}|^2 = \sigma^2.$$
From the uniqueness of the decomposition $X_{n+1} = \hat{X}_{n+1} + Y_{n+1}$ and from formula (93) we can see that
\[
\hat{X}_{n+1} = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + \theta_1 Y_n + \cdots + \theta_q Y_{n+1-q},
\]
thus, if we use the relation $Y_t = X_t - \hat{X}_t$,
\[
\hat{X}_{n+1} = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + \theta_1 (X_n - \hat{X}_n) + \cdots + \theta_q (X_{n+1-q} - \hat{X}_{n+1-q}).
\]
The $h > 1$-step prediction is
\[
\hat{X}_{n+h}(n) = P_{H_n^{-\infty}}(X_{n+h}) = P_{H_n^{-\infty}}(P_{H_n^{-\infty}}(X_{n+h}))
\]
\[
= P_{H_n^{-\infty}}(\hat{X}_{n+h})
\]
\[
= P_{H_n^{-\infty}}(\varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p} + \theta_1 Y_{n+h-1} + \cdots + \theta_q Y_{n+h-q})
\]
\[
= \varphi_1 [X_{n+h-1}] + \cdots + \varphi_p [X_{n+h-p}] + \theta_1 [Y_{n+h-1}] + \cdots + \theta_q [Y_{n+h-q}],
\]
where
\[
[X_{n+j}] = \begin{cases} 
X_{n+j}, & j \leq 0, \\
\hat{X}_{n+j}(n), & j > 0 
\end{cases}
\]
and
\[
[Y_{n+j}] = \begin{cases} 
X_{n+j} - \hat{X}_{n+j}, & j \leq 0, \\
0, & j > 0 
\end{cases}
\]
From the causality we get
\[
\hat{X}_{n+h}(n) = P_{H_n^{-\infty}}(X_{n+h}) = P_{H_n^{-\infty}}(\sum_{j=0}^{\infty} c_j Y_{n+h-j}).
\]
From properties of the projection mapping
\[
P_{H_n^{-\infty}}(\sum_{j=0}^{\infty} c_j Y_{n+h-j}) = \sum_{j=0}^{\infty} c_j P_{H_n^{-\infty}}(Y_{n+h-j}),
\]
and thus
\[
\hat{X}_{n+h}(n) = \sum_{j=h}^{\infty} c_j Y_{n+h-j}.
\]
The prediction error is
\[
E |X_{n+h} - \hat{X}_{n+h}(n)|^2 = E \left| \sum_{j=0}^{h-1} c_j Y_{n+h-j} \right|^2 = \sigma^2 \sum_{j=0}^{h-1} |c_j|^2.
\]
Example 44. Consider an MA(1) model

\[ X_t = Y_t + \theta Y_{t-1}, \quad t \in \mathbb{Z}, \quad Y_t \sim \text{WN}(0, \sigma^2), \quad |\theta| < 1. \]

\{X_t, t \in \mathbb{Z}\} is invertible, \( Y_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j} \), and the best linear prediction is

\[ \hat{X}_{n+1}(n) = \hat{X}_{n+1} = -\sum_{j=1}^{\infty} (-\theta)^j X_{n+1-j} = \theta Y_n = \theta(X_n - \hat{X}_n). \]

The prediction error is

\[ \mathbb{E}|X_{n+1} - \hat{X}_{n+1}|^2 = \mathbb{E}Y_{n+1}^2 = \sigma^2. \]

The \( h > 1 \) step prediction

\[ \hat{X}_{n+h}(n) = P_{H_{-\infty}}(P_{H_{-\infty}}(X_{n+h})) = P_{H_{-\infty}}(\hat{X}_{n+h}) = \theta P_{H_{-\infty}}(Y_{n+h-1}) = 0, \]

since for \( h \geq 2, Y_{n+h-1} \perp H_{-\infty}^{n}. \)

The \( h \)-step prediction error is

\[ \mathbb{E}|X_{n+h} - \hat{X}_{n+h}(n)|^2 = \mathbb{E}|X_{n+h}|^2 = R_X(0) = \sigma^2(1 + \theta^2). \]

11 Prediction in the spectral domain

- Let \( \{X_t, t \in \mathbb{Z}\} \) be a centered stationary sequence with a spectral distribution function \( F \) and spectral density \( f \).

- We know the whole past of the sequence \( \{X_t, t \in \mathbb{Z}\} \) up to time \( n - 1 \), and want to forecast \( X_{n+h}, h = 0, 1, \ldots \), i.e., want to find \( \hat{X}_{n+h}(n-1) = P_{H_{-\infty}}(X_{n+h}) \), in other words, we want to find \( \hat{X}_{n+h}(n-1) \in H_{-\infty}^{n-1} \subset \mathcal{H}\{X_t, t \in \mathbb{Z}\} \), such that \( X_{n+h} - \hat{X}_{n+h}(n-1) \perp H_{-\infty}^{n-1}. \)

- Recall spectral decomposition: \( X_t = \int_{-\pi}^{\pi} e^{it\lambda}dZ(\lambda) \), where \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is an orthogonal increment process with the associated distribution function \( F \) (Theorem 28).

- Recall that all the elements of the Hilbert space \( \mathcal{H}\{X_t, t \in \mathbb{Z}\} \) are of the form

\[ \int_{-\pi}^{\pi} \varphi(\lambda)dZ(\lambda), \]

where \( \varphi \in L_2(F) \) (Theorem 30).
Element $\hat{X}_{n+h}(n - 1)$ should be of the form

$$\hat{X}_{n+h}(n - 1) = \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) d\lambda,$$

(96)

where $\Phi_h(\lambda) \in L_2(F)$. Condition $X_{n+h} - \hat{X}_{n+h}(n - 1) \perp H_{n-1}^*$ will be met if

$$X_{n+h} - \hat{X}_{n+h}(n - 1) \perp X_{n-j}, \quad j = 1, 2, \ldots$$

thus for $j = 1, 2, \ldots$,

$$E(X_{n+h} - \hat{X}_{n+h}(n - 1)) X_{n-j} = 0,$$

or

$$E \left( X_{n+h} - \hat{X}_{n+h}(n - 1) \right) X_{n-j} = R(h + j) - E \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) d\lambda \int_{-\pi}^{\pi} e^{i(n-j)\lambda} d\lambda = R(h + j) - \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) e^{-i(n-j)\lambda} dF(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{i(h+j)\lambda} dF(\lambda) - \int_{-\pi}^{\pi} e^{ij\lambda} \Phi_h(\lambda) f(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} e^{i(h+j)\lambda} f(\lambda) d\lambda - \int_{-\pi}^{\pi} e^{ij\lambda} \Phi_h(\lambda) f(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} e^{ij\lambda} \left( e^{ih\lambda} - \Phi_h(\lambda) \right) f(\lambda) d\lambda = 0.$$

(97)

Denote

$$\Psi_h(\lambda) := (e^{ih\lambda} - \Phi_h(\lambda)) f(\lambda).$$

Then (97) can be written as

$$\int_{-\pi}^{\pi} e^{ij\lambda} \Psi_h(\lambda) d\lambda = 0, \quad j = 1, 2, \ldots$$

(98)

It follows from condition (98) that the Fourier expansion of the function $\Psi_h$ has only terms with nonnegative powers of $e^{i\lambda}$,

$$\Psi_h(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda}, \quad \sum_{k=0}^{\infty} |b_k| < \infty.$$
Provided that
\[ \Phi_h(\lambda) = \sum_{k=1}^{\infty} a_k e^{-ik\lambda}, \quad \sum_{k=1}^{\infty} |a_k| < \infty, \]
which is a function convergent in \( L_2(F) \),

\[ \hat{X}_{n+h}(n-1) = \int_{-\pi}^{\pi} e^{in\lambda} \left[ \sum_{k=1}^{\infty} a_k e^{-ik\lambda} \right] dZ(\lambda) \]
\[ = \lim_{N \to \infty} \int_{-\pi}^{\pi} e^{in\lambda} \left[ \sum_{k=1}^{N} a_k e^{-ik\lambda} \right] dZ(\lambda) \]
\[ = \lim_{N \to \infty} \sum_{k=1}^{N} a_k \left[ \int_{-\pi}^{\pi} e^{i(n-k)\lambda} dZ(\lambda) \right] \]
\[ = \lim_{N \to \infty} \sum_{k=1}^{N} a_k X_{n-k} = \sum_{k=1}^{\infty} a_k X_{n-k}. \]

**Theorem 55.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a real-valued centered stationary random sequence with an autocovariance function \( R \) and a spectral density \( f(\lambda) = f^*(e^{i\lambda}) \), where \( f^* \) is a rational function of a complex-valued variable.

Let \( \Phi_h^* \) be a function of a complex-valued variable \( z \) holomorphic for \(|z| \geq 1\) and such that \( \Phi_h^*(\infty) = 0 \).

Let

\[ \Psi_h^*(z) = (z^h - \Phi_h^*(z)) f^*(z), \quad z \in \mathbb{C}, \]

be a function holomorphic for \(|z| \leq 1\). Then the best linear prediction of \( X_{n+h} \) from \( X_{n-1}, X_{n-2}, \ldots \) is

\[ \hat{X}_{n+h}(n-1) = \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) dZ(\lambda), \]

where \( \Phi_h(\lambda) = \Phi_h^*(e^{i\lambda}) \) and \( \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is the orthogonal increment process from the spectral decomposition of \( \{X_t, t \in \mathbb{Z}\} \). The prediction error is

\[ \delta_h^2 = \mathbb{E}|X_{n+h} - \hat{X}_{n+h}(n-1)|^2 \]
\[ = R(0) - \int_{-\pi}^{\pi} |\Phi_h(\lambda)|^2 f(\lambda) d\lambda \]
\[ = R(0) - \int_{-\pi}^{\pi} e^{-ih\lambda} \Phi_h(\lambda) f(\lambda) d\lambda. \]

**Proof.** Anděl (1976), Chap. X, Theorem 8. \( \square \)
The function $\Phi_h$ is called to be \textit{spectral characteristic of prediction} of $X_{n+h}$ from $X_{n-1}, X_{n-2}, \ldots$

\textbf{Example 45.} 

$$X_t = \varphi X_{t-1} + Y_t, \ |\varphi| < 1, \ \varphi \neq 0, Y_t \sim \text{WN}(0, \sigma^2).$$

We want to determine prediction $\hat{X}_{n+h}(n-1)$ on the basis of $X_{n-1}, X_{n-2}, \ldots, h \geq 0$ in the spectral domain.

The spectral density of the sequence $\{X_t, t \in \mathbb{Z}\}$ is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \frac{1}{(1 - \varphi e^{-i\lambda})(1 - \varphi e^{i\lambda})} = f^*(e^{i\lambda}),$$

where

$$f^*(z) = \frac{\sigma^2}{2\pi} \frac{1}{(1 - \varphi z^{-1})(1 - \varphi z)} = \frac{\sigma^2}{2\pi} \frac{z}{(1 - \varphi z)(z - \varphi)}$$

which is a rational function of the complex-valued variable $z$. The function

$$\Psi_h^*(z) = (z^h - \Phi_h^*(z)) f^*(z) = \frac{\sigma^2}{2\pi} \frac{z(z^h - \Phi_h^*(z))}{(1 - \varphi z)(z - \varphi)}, \ z \in \mathbb{C},$$

should be holomorphic for $|z| \geq 1$. Since $|\varphi| < 1$, it must be $z^h - \Phi_h^*(z) = 0$ for $z = \varphi$, otherwise $\Psi_h^*$ has a pole at $z = \varphi$.

Thus

$$\Phi_h^*(\varphi) = \varphi^h.$$

Put $\Phi_h^*(z) := \frac{z}{\varphi^h}$, where $\gamma$ is a constant. Then $\Phi_h^*$ is holomorphic for $|z| \geq 1$ and $\Phi_h^*(\infty) = 0$. The function $\Psi_h^*$ is holomorphic for $|z| \leq 1$.

The value of constant $\gamma$ follows from (101): $\varphi^h = \Phi_h^*(\varphi) = \frac{z}{\varphi}$, thus $\gamma = \varphi^{h+1}$. Functions

$$\Phi_h^*(z) = \frac{\varphi^{h+1}}{z} = \varphi^{h+1}z^{-1}, \ z \in \mathbb{C},$$

$$\Psi_h^*(z) = \frac{\sigma^2}{2\pi} \frac{z^{h+1} - \varphi^{h+1}}{(z - \varphi)(1 - \varphi z)}, \ z \in \mathbb{C},$$

satisfy the conditions of Theorem 55. The spectral characteristic of prediction is $\Phi_h(\lambda) = \varphi^{h+1}e^{-i\lambda}$ and the best linear forecast is

$$\hat{X}_{n+h}(n-1) = \int_{-\pi}^{\pi} e^{in\lambda}\Phi_h(\lambda)dZ(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{in\lambda}\varphi^{h+1}e^{-i\lambda}dZ(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{i(n-1)\lambda}dZ(\lambda)\varphi^{h+1} = \varphi^{h+1}X_{n-1}.$$
which is the same result as obtained in the time domain.

For the prediction error, from (99) we get

\[
\delta^2_h = \mathbb{E}|X_{t+h} - \hat{X}_{t+h}(t-1)|^2 = \mathbb{E}||X_{t+h}||^2 - ||\hat{X}_{t+h}(t-1)||^2
\]

\[
= R(0) - \mathbb{E}|\hat{X}_{t+h}(t-1)|^2 = R(0) - \mathbb{E}\int_{-\pi}^{\pi} e^{it\lambda}\Phi_h(\lambda)dZ(\lambda)
\]

\[
= R(0) - \int_{-\pi}^{\pi} |e^{it\lambda}\Phi_h(\lambda)|^2 f(\lambda)d\lambda = R(0) - |\varphi|^{2(h+1)}\int_{-\pi}^{\pi} f(\lambda)d\lambda = R(0) \left(1 - \varphi^{2(h+1)}\right),
\]

which is again in accordance with the result in the time domain.

12 Filtration of signal and noise

Let us consider a sequence \(\{X_t, t \in \mathbb{Z}\}\), called to be a signal, and a sequence \(\{Y_t, t \in \mathbb{Z}\}\), a noise. Further consider the sequence \(\{V_t, t \in \mathbb{Z}\}\), where

\[V_t = X_t + Y_t, \quad t \in \mathbb{Z},\]

i.e., \(\{V_t, t \in \mathbb{Z}\}\), is a mixture of the signal and the noise. Our aim is to separate the signal from this mixture.

12.1 Filtration in finite stationary sequences

Let \(\{X_t, t \in \mathbb{Z}\}\), \(\{Y_t, t \in \mathbb{Z}\}\), be real-valued centered stationary sequences, mutually uncorrelated, with autocovariance functions \(R_X, R_Y\), respectively. Let \(V_t = X_t + Y_t, \quad t \in \mathbb{Z}\). Then \(\{V_t, t \in \mathbb{Z}\}\) is also the real-valued centered and stationary sequence with the autocovariance function \(R_V = R_X + R_Y\). Suppose \(V_1, \ldots, V_n\) to be known observations.

On the basis of \(V_1, \ldots, V_n\), we want to find the best linear approximation of \(X_s\) in the form \(\hat{X}_s = \sum_{j=1}^{n} c_j V_j\), with coefficients \(c_1, \ldots, c_n\) that minimize the mean square error \(\mathbb{E}|X_s - \hat{X}_s|^2\).

Denote \(H^n_1 = \mathcal{H}\{V_1, \ldots, V_n\} \subset L_2(\Omega, \mathcal{A}, P)\). Then the best linear approximation \(\hat{X}_s\) of \(X_s\) is the projection of \(X_s \in L_2(\Omega, \mathcal{A}, P)\) onto \(H^n_1\), i.e., \(\hat{X}_s \in H^n_1\) and \(X_s - \hat{X}_s \perp H^n_1\). Since \(H^n_1 = \mathcal{H}\{V_1, \ldots, V_n\} = \mathcal{M}\{V_1, \ldots, V_n\}\), it suffices to find constants \(c_1, \ldots, c_n\) such that

\[
\hat{X}_s = \sum_{j=1}^{n} c_j V_j
\]

and

\[
X_s - \hat{X}_s \perp V_t, \quad t = 1, \ldots, n,
\]
or
\[ \mathbb{E}(X_s - \hat{X}_s)V_t = 0, \ t = 1, \ldots, n. \] (102)

Since \( V_t = X_t + Y_t \) for all \( t \) and \( X_t, Y_t \) are uncorrelated, we can see that \( \mathbb{E}X_sV_t = \mathbb{E}X_sX_t = R_X(s-t) \), and equations (102) can be written in the form

\[ R_X(s-t) - \sum_{j=1}^{n} c_j R_V(j-t) = 0, \ t = 1, \ldots, n. \] (103)

The variable \( \hat{X}_s \) is the best linear filtration of the signal \( X_s \) at time \( s \) from the mixture \( V_1, \ldots, V_n \).

The filtration error is
\[
\delta^2 = \mathbb{E}|X_s - \hat{X}_s|^2 = \|X_s - \hat{X}_s\|^2 = \|X_s\|^2 - \|\hat{X}_s\|^2
\]
\[
= R_X(0) - \mathbb{E}\left[\sum_{j=1}^{n} c_j V_j\right]^2 = R_X(0) - \sum_{j=1}^{n} \sum_{k=1}^{n} c_j c_k R_V(j-k).
\]

The system of equations (103) can be written in the obvious matrix form. For the regularity of the matrix of elements \( R_V(j-t), j, t = 1, \ldots, n \) see Theorem 53.

12.2 Filtration in an infinite stationary sequence

Consider a signal \( \{X_t, t \in \mathbb{Z}\} \), a noise \( \{Y_t, t \in \mathbb{Z}\} \) and the mixture \( \{V_t, t \in \mathbb{Z}\} \), where \( V_t = X_t + Y_t \) for any \( t \in \mathbb{Z} \). Our aim is to find the best linear filtration of \( X_s \) from the sequence of observations \( \{V_t, t \in \mathbb{Z}\} \).

**Theorem 56.** Let \( \{X_t, t \in \mathbb{Z}\} \) and \( \{Y_t, t \in \mathbb{Z}\} \) be centered stationary sequences, mutually uncorrelated, with spectral densities \( f_X \) a \( f_Y \), respectively, that are continuous and \( f_X(\lambda) + f_Y(\lambda) > 0 \) for all \( \lambda \in [-\pi, \pi] \). Let \( \{V_t, t \in \mathbb{Z}\} \) be a random sequence such that \( V_t = X_t + Y_t \) for all \( t \in \mathbb{Z} \). The the best linear filtration of \( X_s \) from \( \{V_t, t \in \mathbb{Z}\} \) is

\[
\hat{X}_s = \int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda)dZ_V(\lambda),
\]

where
\[
\Phi(\lambda) = \frac{f_X(\lambda)}{f_V(\lambda)}, \quad \lambda \in [-\pi, \pi],
\] (104)

\( f_V = f_X + f_Y \) is the spectral density of \( \{V_t, t \in \mathbb{Z}\} \) and \( Z_V = \{Z_\lambda, \lambda \in [-\pi, \pi]\} \) is an orthogonal increment process from the spectral decomposition of the sequence \( \{V_t, t \in \mathbb{Z}\} \).

The filtration error is
\[
\delta^2 = \int_{-\pi}^{\pi} \frac{f_X(\lambda)f_Y(\lambda)}{f_X(\lambda) + f_Y(\lambda)}d\lambda = \int_{-\pi}^{\pi} \Phi(\lambda)f_Y(\lambda)d\lambda.
\]
The function $\Phi$ is called \textit{spectral characteristic of filtration}.

\textbf{Remark 10.} Notice that if $\Phi(\lambda) = \sum_{k=-\infty}^{\infty} a_k e^{ik\lambda}$, where $\sum_{k=-\infty}^{\infty} |a_k| < \infty$, then $\hat{X}_s = \sum_{k=-\infty}^{\infty} a_k V_{s-k}$.

\textit{Proof.} The sequences $\{X_t, t \in \mathbb{Z}\}$ and $\{Y_t, t \in \mathbb{Z}\}$ are centered, stationary and mutually uncorrelated with spectral densities. It follows that the sequence $\{V_t, t \in \mathbb{Z}\}$ is centered and stationary with the autocovariance function $R_V = R_X + R_Y$. Then the spectral density of $\{V_t, t \in \mathbb{Z}\}$ exists and is equal to $f_V = f_X + f_Y$. (Please check!)

The best linear filtration of $X_s$ from $\{V_t, t \in \mathbb{Z}\}$ is the projection of $X_s$ onto the Hilbert space $H = \mathcal{H}\{V_t, t \in \mathbb{Z}\}$, i.e., we are interested in $\hat{X}_s = P_H(X_s)$.

Let $\Phi$ be the function defined in (104). First, we will show that $\hat{X}_s := \int_{-\pi}^{\pi} e^{ix\lambda} \Phi(\lambda) dZ_V(\lambda) \in H$. According to Theorem 30 it suffices to show that $\Phi \in L_2(F_V)$, where $F_V$ is the spectral distribution function of the sequence $\{V_t, t \in \mathbb{Z}\}$. According to the assumption, $f_X$ and $f_V$ are continuous functions and $f_V$ takes in $[-\pi, \pi]$ positive values, only. Thus,

$$\int_{-\pi}^{\pi} |\Phi(\lambda)|^2 dF_V(\lambda) = \int_{-\pi}^{\pi} \left| \frac{f_X(\lambda)}{f_V(\lambda)} \right|^2 f_V(\lambda) d\lambda = \int_{-\pi}^{\pi} \left| \frac{f_X(\lambda)}{f_V(\lambda)} \right|^2 f_V(\lambda) d\lambda < \infty$$

and $\hat{X}_s \in H$.

Further, $\hat{X}_s$ will be the projection of $X_s$ onto $H$ if $(X_s - \hat{X}_s) \perp H$, i.e., if $(X_s - \hat{X}_s) \perp V_t$ for all $t \in \mathbb{Z}$. For any $t \in \mathbb{Z}$ we have

$$E\left( (X_s - \hat{X}_s) V_t \right) = E X_s V_t - E \hat{X}_s V_t$$

$$= E X_s (X_t + Y_t) - E \left( \int_{-\pi}^{\pi} e^{ix\lambda} \Phi(\lambda) dZ_V(\lambda) \int_{-\pi}^{\pi} e^{it\lambda} dZ_V(\lambda) \right)$$

$$= EX_s X_t - \int_{-\pi}^{\pi} e^{ix\lambda} \Phi(\lambda) dF_V(\lambda)$$

$$= R_X(s-t) - \int_{-\pi}^{\pi} e^{i(s-t)\lambda} \Phi(\lambda) f_V(\lambda) d\lambda$$

$$= R_X(s-t) - \int_{-\pi}^{\pi} e^{i(s-t)\lambda} f_X(\lambda) d\lambda$$

$$= R_X(s-t) - R_X(s-t) = 0.$$

We have proved that $\hat{X}_s$ is the best linear filtration.
Let us determine the filtration error:

\[ \delta^2 = \|X_s\|^2 - \|\hat{X}_s\|^2 = R_X(0) - \mathbb{E}[\hat{X}_s]^2 \]

\[ = \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \mathbb{E}\left[ \left| \int_{-\pi}^{\pi} e^{i\lambda\Phi(\lambda)}dV(\lambda) \right|^2 \right] \]

\[ = \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \int_{-\pi}^{\pi} \left| \Phi(\lambda) \right|^2 dF_V(\lambda) \]

\[ = \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \int_{-\pi}^{\pi} \left| f_X(\lambda) \right|^2 f_V(\lambda)d\lambda \]

\[ = \int_{-\pi}^{\pi} f_X(\lambda)f_Y(\lambda)d\lambda = \int_{-\pi}^{\pi} \Phi(\lambda)f_Y(\lambda)d\lambda. \]

Example 46. Let the signal \( \{X_t, t \in \mathbb{Z}\} \) and the noise \( \{Y_t, t \in \mathbb{Z}\} \) be mutually independent sequences such that

\[ X_t = \varphi X_{t-1} + W_t, \quad t \in \mathbb{Z}, \]

where \( |\varphi| < 1, \varphi \neq 0 \) and \( \{W_t, t \in \mathbb{Z}\} \) is a white noise with zero mean and variance \( \sigma_W^2 \), and \( \{Y_t, t \in \mathbb{Z}\} \) is another white noise sequence with zero mean and variance \( \sigma_Y^2 \). We observe \( V_t = X_t + Y_t, t \in \mathbb{Z}. \)

Obviously, \( \{X_t, t \in \mathbb{Z}\} \) and \( \{Y_t, t \in \mathbb{Z}\} \) are centered stationary sequences with the spectral densities

\[ f_X(\lambda) = \frac{\sigma_W^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\lambda}|^2}, \quad f_Y(\lambda) = \frac{\sigma_Y^2}{2\pi}, \lambda \in [-\pi, \pi] \]

that satisfy conditions of Theorem 56.

The sequence \( \{V_t, t \in \mathbb{Z}\} \) has the spectral density \( f_V = f_X + f_Y \) and it can be shown that

\[ f_V(\lambda) = \frac{\sigma_W^2}{2\pi} \frac{1 - \theta e^{-i\lambda}}{|1 - \varphi e^{-i\lambda}|^2}, \quad \lambda \in [-\pi, \pi], \] (105)

where \( \sigma^2 = \varphi^2 \sigma_Y^2 \), \( \theta \) is the root of the equation \( \theta^2 - c\theta + 1 = 0 \), the absolute value of which is less than one and has the same sign as the coefficient \( \varphi \), and

\[ c = \frac{\sigma_W^2 + \sigma_Y^2(1 + \varphi^2)}{\varphi \sigma_Y^2}. \]

(See Prášková, 2006, Problem 8.1 for some hints.) Then

\[ \Phi(\lambda) = \frac{f_X(\lambda)}{f_V(\lambda)} = \frac{\sigma_W^2}{\sigma^2} \frac{1}{|1 - \theta e^{-i\lambda}|^2} = \frac{\sigma_W^2}{\sigma^2} \left| \sum_{k=0}^{\infty} \theta^k e^{-ik\lambda} \right|^2 \]

\[ = \frac{\sigma_W^2}{\sigma^2} \frac{1}{1 - \theta^2} \sum_{k=-\infty}^{\infty} \theta^{|k|} e^{-ik\lambda} \]
for all $\lambda \in [-\pi, \pi]$.

The best linear filtration of $X_s$ from $\{V_t, t \in \mathbb{Z}\}$ is

$$
\hat{X}_s = \frac{\sigma^2_w}{\sigma^2} \frac{1}{1-\theta^2} \sum_{k=-\infty}^{\infty} \theta^{|k|} V_{s-k}.
$$

(106)

The filtration error is

$$
\delta^2 = E|X_s - \hat{X}_s|^2 = \int_{-\pi}^{\pi} \phi(\lambda) f_Y(\lambda) d\lambda
$$

$$
= \frac{\sigma^2_Y \sigma^2_w}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|1-\theta e^{-i\lambda}|^2} d\lambda = \frac{\sigma^2_Y \sigma^2_w}{\sigma^2} \frac{1}{1-\theta^2}.
$$

Remark 11. It follows from (105) that $f_Y$ has the same form as the spectral density of an ARMA$(1,1)$ sequence. The mixture of an autoregressive sequence of the first order and a white noise $\{V_t, t \in \mathbb{Z}\}$ has the same covariance structure as the stationary sequence $\{Z_t, t \in \mathbb{Z}\}$ that is modeled as

$$
Z_t - \varphi Z_{t-1} = U_t - \theta U_{t-1}, \quad t \in \mathbb{Z},
$$

where $\varphi \neq 0$, $|\varphi| < 1$, $|\theta| < 1$ and $\{U_t, t \in \mathbb{Z}\}$ is a white noise with the variance $\sigma^2 = \frac{\sigma^2_Y}{\varphi^2}$. Parameter $\theta$ can be determined as given above.

Remark 12. Function $\Phi$ is the transfer function of the linear filter $\{\frac{\sigma^2_w}{\sigma^2} \frac{1}{1-\theta^2} \theta^{|k|}, k \in \mathbb{Z}\}$.

13 Partial autocorrelation function

Definition 41. Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued centered stationary sequence. The partial autocorrelation function of $\{X_t, t \in \mathbb{Z}\}$ is defined to be

$$
\alpha(k) = \begin{cases} 
\text{corr}(X_1, X_{k+1}) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}X_1} \sqrt{\text{var}X_2}} & k = 1, \\
\text{corr} \left( X_1 - \tilde{X}_1, X_{k+1} - \tilde{X}_{k+1} \right) & k > 1,
\end{cases}
$$

where $\tilde{X}_1$ is the linear projection of $X_1$ onto Hilbert space $H^k_2 = \mathcal{H}\{X_2, \ldots, X_k\}$ and $\tilde{X}_{k+1}$ is the linear projection of $X_{k+1}$ onto $H^k_2$.

From the properties of the projection mapping it follows that $\tilde{X}_1 = c_2 X_2 + \cdots + c_k X_k$ where constants $c_2, \ldots, c_k$ are determined by conditions $E(X_1 - \tilde{X}_1)X_j = 0$, $j = 2, \ldots, k$. The same holds for $\tilde{X}_{k+1}$. We can see that $\alpha(k)$ represents the correlation coefficient of
residuals $X_1 - \tilde{X}_1$ and $X_{k+1} - \tilde{X}_{k+1}$ after the best linear approximation of the variables $X_1$ and $X_{k+1}$ by random variables $X_2, \ldots, X_k$.

The stationarity of the sequence $\{X_t, t \in \mathbb{Z}\}$ implies, that for $h \in \mathbb{N}$, $\text{corr}(X_1 - \tilde{X}_1, X_{k+1} - \tilde{X}_{k+1}) = \text{corr}(X_h - \tilde{X}_h, X_{k+h} - \tilde{X}_{k+h})$, where $\tilde{X}_h, \tilde{X}_{k+h}$ are linear projections of random variables $X_h, X_{k+h}$ onto the Hilbert space $\mathcal{H}\{X_{h+1}, \ldots, X_{h+k-1}\}$. Therefore, $\alpha(k)$ is also the correlation coefficient of $X_h$ and $X_{h+k}$ after the linear dependence $X_{h+1}, \ldots, X_{h+k-1}$ was eliminated.

**Example 47.** Consider a causal AR(1) process

$$X_t = \varphi X_{t-1} + Y_t,$$

where $|\varphi| < 1$ and $Y_t \sim \text{WN}(0, \sigma^2)$.

According to the definition, $\alpha(1) = \text{corr}(X_1, X_2) = r_X(1) = \varphi$. For $k > 1$,

$$\alpha(k) = \frac{\mathbb{E}(X_1 - \tilde{X}_1)(X_{k+1} - \tilde{X}_{k+1})}{\sqrt{\mathbb{E}(X_1 - \tilde{X}_1)^2} \sqrt{\mathbb{E}(X_{k+1} - \tilde{X}_{k+1})^2}}.$$

Due to causality, $\tilde{X}_{k+1} = \mathcal{P}_{\mathcal{H}_2}(X_{k+1}) = \varphi X_k$ and $X_{k+1} - \tilde{X}_{k+1} = Y_{k+1} \perp \mathcal{H}_2^\perp$. Further, it follows from causality that $Y_{k+1} \perp X_1$, thus $\mathbb{E}(X_1 - \tilde{X}_1)(X_{k+1} - \tilde{X}_{k+1}) = \mathbb{E}(X_1 - \tilde{X}_1)Y_{k+1} = 0$, from which we conclude that $\alpha(k) = 0$ for $k > 1$.

**Remark 13.** In the same manner, for a causal AR($p$) sequence we could prove that the partial autocorrelation $\alpha(k) = 0$ for $k > p$.

**Example 48.** Consider an MA(1) process

$$X_t = Y_t + bY_{t-1},$$

where $|b| < 1$ and $Y_t \sim \text{WN}(0, \sigma^2)$. We know that in this case $R_X(0) = (1 + b^2)\sigma^2$, $R_X(1) = b\sigma^2 = R_X(-1)$ and $R_X(k) = 0$ for $|k| > 1$.

We compute the partial autocorrelations.

First, $\alpha(1) = r_X(1) = \frac{b}{1+b^2}$. Further, $\alpha(2) = \text{corr}(X_1 - \tilde{X}_1, X_3 - \tilde{X}_3)$. To determine $\tilde{X}_1$, notice that $\tilde{X}_1 = \mathcal{P}_{\mathcal{H}_2}X_1 = cX_2$ and $\left(X_1 - \tilde{X}_1\right) \perp \tilde{X}_1$. Thus $\mathbb{E}(X_1 - cX_2)X_2 = 0$, and $c = \frac{R_X(1)}{R_X(0)} = \frac{b}{1+b^2}$. We have $\tilde{X}_1 = \frac{b}{1+b^2}X_2$. Quite analogously we get $\tilde{X}_3 = \frac{b}{1+b^2}X_2$, i.e., $\tilde{X}_1 = \tilde{X}_3$. We have

$$\alpha(2) = \text{corr}\left(X_1 - \frac{b}{1+b^2}X_2, X_3 - \frac{b}{1+b^2}X_2\right).$$

102
Obviously,
\[
E\left(X_1 - \frac{b}{1 + b^2}X_2 \right) \left(X_3 - \frac{b}{1 + b^2}X_2 \right) = R_X(2) - \frac{2b}{1 + b^2}R_X(1) + \frac{b^2}{(1 + b^2)^2}R_X(0) = -\frac{\sigma^2b^2}{1 + b^2},
\]
similarly,
\[
E\left(X_1 - \frac{b}{1 + b^2}X_2 \right)^2 = R_X(0) + \frac{b^2}{(1 + b^2)^2}R_X(0) - \frac{2b}{1 + b^2}R_X(1) = \frac{\sigma^2(1 + b^2 + b^4)}{1 + b^2},
\]
and combining these results we conclude that
\[
\alpha(2) = -\frac{b^2}{1 + b^2 + b^4}.
\]

Generally, it can be shown that
\[
\alpha(k) = -\frac{(-b)^k(1 - b^2)}{1 - b^{2(k+1)}}, \quad k \geq 1.
\]

**Definition 42** (An alternative definition of the partial correlation function). Let \(\{X_t, t \in \mathbb{Z}\}\) be a a centered stationary sequence, let \(P_{H_k^1}(X_{k+1})\) be the best linear prediction of \(X_{k+1}\) on the basis of \(X_1, \ldots, X_k\). If \(H_k^1 = \mathcal{H}\{X_1, \ldots, X_k\}\), and \(P_{H_k^1}(X_{k+1}) = \varphi_1X_k + \cdots + \varphi_kX_1\), then the partial autocorrelation function at lag \(k\) is defined to be \(\alpha(k) = \varphi_k\).

**Theorem 57.** Let \(\{X_t, t \in \mathbb{Z}\}\) be a real-valued sequence with the autocovariance function \(R\), such that \(R(0) > 0, R(t) \to 0\), as \(t \to \infty\). Then the both definitions of the partial autocorrelation function are equivalent and it holds
\[
\begin{bmatrix}
1 & r(1) & \ldots & r(k-2) & r(k-1) \\
r(1) & 1 & \ldots & r(k-3) & r(k-2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r(k-1) & r(k-2) & \ldots & r(1) & r(k) \\
1 & r(1) & \ldots & r(k-1) & r(k-1)
\end{bmatrix} \quad 107
\]
\[
\alpha(k) = \begin{bmatrix}
1 & r(1) & \ldots & r(k-2) & r(k-1) \\
r(1) & 1 & \ldots & r(k-3) & r(k-2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r(k-1) & r(k-2) & \ldots & 1
\end{bmatrix}
\]
where \(r\) is the autocorrelation function of the sequence \(\{X_t, t \in \mathbb{Z}\}\).
Proof. Denote $H^k_1 = \mathcal{H}\{X_1, \ldots, X_k\}$, $H^k_2 = \mathcal{H}\{X_2, \ldots, X_k\}$, $\hat{X}_{k+1} = P_{H^k_1}(X_{k+1})$, $\tilde{X}_1 = P_{H^k_2}(X_1)$, $\tilde{X}_{k+1} = P_{H^k_2}(X_{k+1})$.

Since $X_1 = \tilde{X}_1 + (X_1 - \tilde{X}_1)$, where $\tilde{X}_1 \in H^k_2$, $X_1 - \tilde{X}_1 \perp H^k_2$, and $\tilde{X}_{k+1} \in H^k_2$, we get

$$
\mathbb{E}\tilde{X}_{k+1}(X_1 - \tilde{X}_1) = 0.
$$

Consider

$$
\hat{X}_{k+1} = \varphi_1 X_k + \cdots + \varphi_k X_1.
$$

Then

$$
\hat{X}_{k+1} = [\varphi_1 X_k + \cdots + \varphi_{k-1} X_2 + \varphi_k \tilde{X}_1] + [\varphi_k (X_1 - \tilde{X}_1)],
$$

and the random variables in the brackets are mutually orthogonal. Then $[\varphi_k (X_1 - \tilde{X}_1)]$ can be considered to be the projection of $\hat{X}_{k+1}$ onto the Hilbert space $\tilde{H} = \mathcal{H}\{X_1 - \tilde{X}_1\} \subset H^k_1$. It is also the projection of $X_{k+1}$ onto the space $\tilde{H}$ and

$$
\mathbb{E}(X_{k+1} - \varphi_k (X_1 - \tilde{X}_1))(X_1 - \tilde{X}_1) = 0 = \mathbb{E}X_{k+1}(X_1 - \tilde{X}_1) - \varphi_k \mathbb{E}(X_1 - \tilde{X}_1)^2.
$$

From here and from (108) we get

$$
\varphi_k = \frac{\mathbb{E}X_{k+1}(X_1 - \tilde{X}_1)}{\mathbb{E}(X_1 - \tilde{X}_1)^2} = \frac{\mathbb{E}(X_{k+1} - \tilde{X}_{k+1})(X_1 - \tilde{X}_1)}{\mathbb{E}(X_1 - \tilde{X}_1)^2}.
$$

Since $\mathbb{E}(X_1 - \tilde{X}_1)^2 = \mathbb{E}(X_{k+1} - \tilde{X}_{k+1})^2$, which holds from the fact that for a stationary sequence, $\text{var}(X_2, \ldots, X_k) = \text{var}(X_k, \ldots, X_2)$, we get from (109) that

$$
\varphi_k = \text{corr}(X_1 - \tilde{X}_1, X_{k+1} - \tilde{X}_{k+1}) = \alpha(k).
$$

Now we will verify (107). We know that $\hat{X}_{k+1} = \varphi_1 X_k + \varphi_2 X_{k-1} + \cdots + \varphi_k X_1 \in H^k_1$, $X_{k+1} - \hat{X}_{k+1} \perp H^k_1$ and therefore

$$
\mathbb{E}(X_{k+1} - (\varphi_1 X_k + \cdots + \varphi_k X_1))X_{k+1-j} = 0, \quad j = 1, 2, \ldots, k,
$$

which is a system of equations

$$
R(1) - \varphi_1 R(0) - \cdots - \varphi_k R(k-1) = 0
$$

$$
R(2) - \varphi_1 R(1) - \cdots - \varphi_k R(k-2) = 0
$$

$$
\vdots
$$

$$
R(k) - \varphi_1 R(k-1) - \cdots - \varphi_k R(0) = 0.
$$

104
Dividing each equation by \( R(0) \), we get the system of equations:

\[
\varphi_1 + \varphi_2 r(1) + \cdots + \varphi_k r(k-1) = r(1) \\
\varphi_1 r(1) + \varphi_2 + \cdots + \varphi_k r(k-2) = r(2) \\
\vdots \\
\varphi_1 r(k-1) + \varphi_2 r(k-2) + \cdots + \varphi_k r(k) = r(k),
\]

or, in the matrix form,

\[
\begin{pmatrix}
1 & r(1) & \cdots & r(k-1) \\
r(1) & 1 & \cdots & r(k-2) \\
\vdots & \vdots & \ddots & \vdots \\
r(k-1) & r(k-2) & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
\varphi_1 \\
\varphi_2 \\
\vdots \\
\varphi_k
\end{pmatrix} =
\begin{pmatrix}
r(1) \\
r(2) \\
\vdots \\
r(k)
\end{pmatrix},
\]

The ratio of determinants (107) gives the solution for \( \varphi_k \).

**Example 49.** Consider again a causal AR(1) process

\[ X_t = \varphi X_{t-1} + Y_t, \]

where \( |\varphi| < 1 \) and \( Y_t \sim WN(0, \sigma^2) \). Let us compute the partial autocorrelation function according to formula (107). We get

\[
\alpha(k) = \frac{1}{k > 1, (110)}
\]

We can see that the last column of the determinant in the numerator of (110) is obtained by multiplication of the first column, thus, this determinant equals zero.

### 14 Estimators of the mean and the autocorrelation function

#### 14.1 Estimation of the mean

Let \( \{X_t, t \in \mathbb{Z}\} \) be a stationary sequence with expected value \( \mathbb{E}X_t = \mu \) and autocovariance function \( R(s,t) = R(s-t) \).
Figure 8: Autocorrelation (left) and partial autocorrelation function (right) of the AR(1) sequence $X_t = -0.8X_{t-1} + Y_t$

Figure 9: Autocorrelation function (left) and partial autocorrelation function (right) of the MA(1) sequence $X_t = Y_t + 0.8Y_{t-1}$
A common estimator of the mean value is the sample mean defined by

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$  

We know that $\bar{X}_n$ is an unbiased estimator of the expected value, since $E\bar{X}_n = \frac{1}{n} \sum_{t=1}^{n} E X_t = \mu$. We also know that if the sequence $\{X_t, t \in \mathbb{Z}\}$ is mean square ergodic, then $\bar{X}_n \to \mu$ in mean square and also in probability. It guarantees the (weak) consistency of the estimator. Recall that a sufficient condition for mean square ergodicity is $R(t) \to 0$ as $t \to \infty$ (compare Theorems 39 and 40).

The variance of the sample mean in a stationary sequence is

$$\text{var} \bar{X}_n = \frac{1}{n} \sum_{k=-n+1}^{n-1} R(k) \left( 1 - \frac{|k|}{n} \right),$$

and if $\sum_{k=-\infty}^{\infty} |R(k)| < \infty$, then $n \text{var} \bar{X}_n \to \sum_{k=-\infty}^{\infty} R(k) = 2\pi f(0)$, where $f(\lambda)$ is the spectral density of the sequence $\{X_t, t \in \mathbb{Z}\}$ (Theorem 40). We have also proved some central limit theorems for selected strictly stationary sequences, saying that $\bar{X}_n$ has asymptotically distribution $N(\mu, \Delta^2/n)$, where $\Delta^2$ is an asymptotic variance (Theorems 47, 48, 49).

However, the sample mean $\bar{X}_n$ is not the best linear estimator of the expected value of a stationary sequence $\{X_t, t \in \mathbb{Z}\}$. Such estimator can be constructed as follows.

Consider a linear model

$$X_t = \mu + \tilde{X}_t, \quad t = 1, \ldots, n,$$

(111)

where $\tilde{X}_t$, $t = 1, \ldots, n$, is a centered stationary sequence with the autocovariance function $R$, such that $R(0) > 0$, $R(t) \to 0$ as $t \to \infty$. Then from the theory of general linear model (e.g., Anděl, 2002, Theorem 9.2) it holds that the best linear unbiased estimator of the parameter $\mu$ is statistic

$$\hat{\mu}_n = (1_n' \Gamma_n^{-1} 1_n)^{-1} 1_n' \Gamma_n^{-1} X_n,$$

(112)

where

$$\Gamma_n = \text{var} X_n = \begin{pmatrix}
R(0) & R(1) & \cdots & R(n-1) \\
R(1) & R(0) & \cdots & R(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
R(n-1) & R(n-2) & \cdots & R(0)
\end{pmatrix},$$

is a regular matrix according to Theorem 53, $1_n = (1, \ldots, 1)'$ and $X_n = (X_1, \ldots, X_n)'$.

The variance of this estimator is

$$\text{var} \hat{\mu}_n = (1_n' \Gamma_n^{-1} 1_n)^{-1}.$$

(113)
14.2 Estimators of the autocovariance and the autocorrelation function

The best linear estimator (112) assumes knowledge of the autocovariance function \( R \). Similarly, knowledge of the autocovariance function is assumed in prediction. For estimators we usually work with the sample autocovariance function

\[
\hat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n), \quad k = 0, 1, \ldots, n-1
\]

and \( \hat{R}(k) = \hat{R}(-k) \) pro \( k < 0 \). Let us remark that the sample autocovariance is not an unbiased estimator of the autocovariance function, i.e., \( \mathbb{E} \hat{R}(k) \neq R(k) \).

The matrix

\[
\hat{\Gamma}_n = \begin{pmatrix}
\hat{R}(0) & \hat{R}(1) & \cdots & \hat{R}(n-1) \\
\hat{R}(1) & \hat{R}(0) & \cdots & \hat{R}(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{R}(n-1) & \hat{R}(n-2) & \cdots & \hat{R}(0)
\end{pmatrix}
\]

will be regular, if \( \hat{R}(0) > 0 \). For given \( X_1, \ldots, X_n \), function

\[
\hat{R}(k) = \begin{cases}
\frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n), & |k| < n, \\
0, & |k| \geq n
\end{cases}
\]

(115)
can be viewed to be the autocovariance function of an MA\((n-1)\) sequence and the regularity of matrix \( \hat{\Gamma}_n \) is induced by Theorem 53.

If we dispose only \( n \) observations \( X_1, \ldots, X_n \), we can estimate \( R(k) \), \( k = 0, \ldots, n-1 \). From the practical point of view, it is recommended to choose \( n \geq 50 \) and \( k \leq \frac{n}{4} \).

Further, let us consider the autocorrelation function \( r(k) = \frac{R(k)}{R(0)} \).

We define the sample autocorrelation function to be

\[
\hat{r}(k) = \frac{\hat{R}(k)}{\hat{R}(0)} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\sum_{t=1}^{n} (X_t - \bar{X}_n)^2},
\]

if \( \hat{R}(0) = \frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2 > 0 \).

Asymptotic behaviour of the sample autocorrelations is described in the following theorem.

**Theorem 58.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a random sequence

\[
X_t - \mu = \sum_{j=-\infty}^{\infty} \alpha_j Y_{t-j},
\]

108
where \( Y_t, t \in \mathbb{Z}, \) are independent identically distributed random variables with zero mean and finite positive variance \( \sigma^2, \) and let \( \mathbb{E}|Y_i|^4 < \infty \) a \( \sum_{j=-\infty}^{\infty} |a_j| < \infty. \)

Let \( r(k), k \in \mathbb{Z}, \) be the autocorrelation function of the sequence \( \{X_t, t \in \mathbb{Z}\} \) and \( \hat{r}(k) \) be the sample autocorrelation at lag \( k, \) based on \( X_1, \ldots, X_n. \)

Then for each \( h = 1, 2, \ldots, \) as \( n \to \infty, \) the random vector \( \sqrt{n}(\hat{r}(h) - r(h)) \) converges in distribution to a random vector with normal distribution \( \mathcal{N}_h(0, W), \) where

\[
\hat{r}(h) = (\hat{r}(1), \ldots, \hat{r}(h))', \quad r(h) = (r(1), \ldots, r(h))',
\]

and \( W \) is an \( h \times h \) matrix elements of which are

\[
w_{ij} = \sum_{k=1}^{\infty} \left[ r(k+i) + r(k-i) - 2r(i)r(k) \right] \left[ r(k+j) + r(k-j) - 2r(j)r(k) \right]
\]

\[i, j = 1, \ldots, h.\]

**Proof.** Brockwell, Davis (1991), Theorem 7.2.1.

**Remark 14.** Formula (116) is called the Bartlett formula. From the assertion of the theorem we especially get for any \( i \)

\[
\sqrt{n}(\hat{r}(i) - r(i)) \xrightarrow{d} \mathcal{N}(0, w_{ii}), \quad n \to \infty,
\]

i.e., for large \( n, \)

\[
\hat{r}(i) \sim \mathcal{N} \left( r(i), \frac{w_{ii}}{n} \right).
\]

Applications of Theorem 58:

**Example 50.** Consider an AR(1) sequence

\[X_t = \varphi X_{t-1} + Y_t, \quad t \in \mathbb{Z},\]

where \( |\varphi| < 1 \) and \( Y_t, t \in \mathbb{Z} \) are iid with zero mean, finite non-zero variance \( \sigma^2 \) and with finite moments \( \mathbb{E}|Y_i|^4. \) Then \( r(k) = \varphi^{|k|}, \) thus \( r(1) = \varphi \) and according to Theorem 58,

\[
\sqrt{n}(\hat{r}(1) - \varphi) \xrightarrow{d} \mathcal{N}(0, w_{11}), \quad n \to \infty,
\]

where

\[
w_{11} = \sum_{k=1}^{\infty} \left[ r(k+1) + r(k-1) - 2r(1)r(k) \right]^2 = \sum_{k=1}^{\infty} (\varphi^{k-1} - \varphi^{k+1})^2
\]

\[= (1 - \varphi^2)^2 \sum_{k=1}^{\infty} \varphi^{2(k-1)} = 1 - \varphi^2.\]
If we denote $\hat{r}(1) := \hat{\varphi}$, we can write

$$\sqrt{n}(\hat{\varphi} - \varphi) \xrightarrow{D} \mathcal{N}(0,1 - \varphi^2), \ n \to \infty$$

or

$$\sqrt{n} \frac{\hat{\varphi} - \varphi}{\sqrt{1 - \varphi^2}} \xrightarrow{D} \mathcal{N}(0,1), \ n \to \infty.$$  

From here it follows that $\hat{\varphi} \xrightarrow{p} \varphi$ (see, e.g. Brockwell a Davis, 1991, Chap. 6) and also

$$\sqrt{n} \frac{\hat{\varphi} - \varphi}{\sqrt{1 - \hat{\varphi}^2}} \xrightarrow{D} \mathcal{N}(0,1), \ n \to \infty.$$  

An asymptotic 95% confidence interval for $\varphi$ is

$$\left( \hat{\varphi} - 1.96\sqrt{\frac{1 - \hat{\varphi}^2}{n}}, \hat{\varphi} + 1.96\sqrt{\frac{1 - \hat{\varphi}^2}{n}} \right).$$

**Example 51.** Let us suppose that the sequence $\{X_t, t \in \mathbb{Z}\}$ is a strict white noise. Then $r(0) = 1$ and $r(t) = 0$ for $t \neq 0$. The elements of $W$ are

$$w_{ii} = \sum_{k=1}^{\infty} r(k - i)^2 = 1,$$

$$w_{ij} = \sum_{k=1}^{\infty} r(k - i)r(k - j) = 0, \ i \neq j,$$

i.e., $W = I$ is the identity matrix. It means that for large $n$, the vector $\hat{\mathbf{r}}(h) = (\hat{r}(1), \ldots, \hat{r}(h))^\prime$ has approximately normal distribution $\mathcal{N}_h(0, \frac{1}{n}I)$. For large $n$, therefore the random variables $\hat{r}(1), \ldots, \hat{r}(h)$ are approximately independent and identically distributed with zero mean and variance $\frac{1}{n}$. In a plot of sample autocorrelations $\hat{r}(k)$ for $k = 1, \ldots$, approximately 95% of them should be in the interval $(-1.96\sqrt{\frac{1}{n}}, 1.96\sqrt{\frac{1}{n}})$.

The sample partial autocorrelation function is defined to be $\hat{\alpha}(k) = \hat{\varphi}_k$, where $\hat{\varphi}_k$ can be obtained e.g. from (107), where we insert the sample autocorrelation coefficients. The determinant in the denominator of (107) will be non-zero if $\frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2 > 0$.

**Example 52.** In figure 12 the plot of the Wolf index of annual number of sunspots (1700-1987)$^2$ is displayed. In figures 13 and 14 we can see the sample autocovariance

---

$^2$Source: WDC-SILSO, Royal Observatory of Belgium, Brussels
function and the sample partial autocorrelation function, respectively. The data was identified (after centering) with the autoregressive AR(9) process $a(B)X_t = Y_t$ where

$$a(z) = 1 - 1.182z + 0.4248z^2 + 0.1619z^3 - 0.1687z^4$$
$$+ 0.1156z^5 - 0.02689z^6 - 0.005769z^7$$
$$+ 0.02251z^8 - 0.2062z^9$$

$$Y_t \sim WN(0, \sigma^2), \sigma^2 = 219.58$$
Figure 12: Number of Sunspots, the Wolf index

Figure 13: Wolf index, estimated autocorrelation function
15 Estimation of parameters of ARMA models

15.1 Estimation in AR sequences

Let us consider a real-valued stationary causal AR($p$) sequence of a known order $p$

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t, \ t \in \mathbb{Z},$$

(117)

where $\{Y_t, t \in \mathbb{Z}\}$ denotes a white noise process WN$(0, \sigma^2)$, and $\varphi_1, \ldots, \varphi_p, \sigma^2$ are unknowns parameters to be estimated on the basis of $X_1, \ldots, X_n$.

Moment methods

The method utilizes Yule - Walker equations for the autocovariance function $R_X := R$

of the sequence $\{X_t, t \in \mathbb{Z}\}$ in the form

$$R(0) = \varphi_1 R(1) + \cdots + \varphi_p R(p) + \sigma^2,$$

(118)

$$R(k) = \varphi_1 R(k-1) + \cdots + \varphi_p R(k-p), \quad k \geq 1.$$  

(119)

The system of equations for $k = 1, \ldots, p$ can be written in the matrix form

$$\Gamma \varphi = \gamma,$$

(120)

where

$$\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}, \quad \Gamma = \begin{pmatrix} R(0) & \cdots & R(p-1) \\ \vdots & \ddots & \vdots \\ R(p-1) & \cdots & R(0) \end{pmatrix}, \quad \gamma = \begin{pmatrix} R(1) \\ \vdots \\ R(p) \end{pmatrix}. $$

Figure 14: Wolf index, estimated partial autocorrelation function
If we replace the values of \( R(k) \) in \( \Gamma \) and \( \gamma \) by their sample values

\[
\hat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n) (X_{t+k} - \bar{X}_n),
\]

we get the matrix \( \hat{\Gamma} \) and the vector \( \hat{\gamma} \). If we plug these estimators into equation (120), we obtain moment estimators of \( \varphi_1, \ldots, \varphi_p \) by solving

\[
\hat{\varphi} = \left( \hat{\varphi}_1, \ldots, \hat{\varphi}_p \right)' = \hat{\Gamma}^{-1} \hat{\gamma}, \tag{121}
\]

provided \( \hat{\Gamma}^{-1} \) exists. From the previous subsections we know that a sufficient condition for it is

\[
\hat{R}(0) = \frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2 > 0.
\]

An estimate of \( \sigma^2 \) is obtained from (118) as

\[
\hat{\sigma}^2 = \hat{R}(0) - \hat{\varphi}_1 \hat{R}(1) - \cdots - \hat{\varphi}_p \hat{R}(p) = \hat{R}(0) - \hat{\varphi}' \hat{\gamma}.
\]

**Remark 15.** The moment estimators based on Yule-Walker equations are sometimes called Yule-Walker estimators.

**Example 53.** Consider an AR(1) sequence in the form \( X_t = \varphi X_{t-1} + Y_t, \ t \in \mathbb{Z} \), where \( |\varphi| < 1 \) and \( Y_t \) is from WN(0, \( \sigma^2 \)). Moment estimators of parameters \( \varphi \) and \( \sigma^2 \) are

\[
\hat{\varphi} = \frac{\hat{R}(1)}{\hat{R}(0)} = \hat{r}(1), \quad \hat{\sigma}^2 = \hat{R}(0) - \hat{\varphi} \hat{R}(1) = \hat{R}(0) \left( 1 - \hat{\varphi}^2 \right).
\]

Moment estimator of the parameter \( \varphi \) is in this case the same as the sample autocorrelation coefficient \( \hat{r}(1) \) (compare with Example 50.)

Asymptotic properties of moment estimators are described in the following theorem.

**Theorem 59.** Let \( \{X_t, t \in \mathbb{Z}\} \) be an AR(p) sequence generated by \( X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t \) for \( t \in \mathbb{Z} \), where \( \{Y_t, t \in \mathbb{Z}\} \) is a sequence of iid random variables with zero mean and finite non-zero variance \( \sigma^2 \). Suppose that all the roots of the characteristic polynomial \( \lambda^p - \varphi_1 \lambda^{p-1} - \cdots - \varphi_p \) are inside the unit circle and let \( \hat{\varphi} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_p)' \) and \( \hat{\sigma}^2 \) be moment estimators of \( \varphi = (\varphi_1, \ldots, \varphi_p)' \) a \( \sigma^2 \) computed from \( X_1, \ldots, X_n \).

Then

\[
\sqrt{n} (\hat{\varphi} - \varphi) \xrightarrow{d} N_p \left( 0, \sigma^2 \Gamma^{-1} \right), \quad n \to \infty,
\]

where \( \Gamma \) is a matrix with elements \( \Gamma_{ij} = R(i-j), 1 \leq i, j \leq p \), \( R \) is the autocovariance function of \( \{X_t, t \in \mathbb{Z}\} \).

Further, it holds

\[
\hat{\sigma}^2 \xrightarrow{p} \sigma^2, \quad n \to \infty.
\]

114

Least squares method

Consider again sequence (117) and suppose $X_1, \ldots, X_n$ to be known. The least square estimators of parameters $\varphi_1, \ldots, \varphi_p$ are obtained by minimizing the sum of squares

$$\min_{\varphi_1, \ldots, \varphi_p} \sum_{t=p+1}^{n} (X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p})^2.$$ 

The problem leads to solution of the system of equations

$$\sum_{t=p+1}^{n} (X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}) X_{t-j} = 0, \quad j = 1, \ldots, p,$$

i.e., to the system

$$\varphi_1 \sum_{t=p+1}^{n} X_{t-1}^2 + \cdots + \varphi_p \sum_{t=p+1}^{n} X_{t-1} X_{t-p} = \sum_{t=p+1}^{n} X_t X_{t-1},$$

$$\vdots$$

$$\varphi_1 \sum_{t=p+1}^{n} X_{t-1} X_{t-p} + \cdots + \varphi_p \sum_{t=p+1}^{n} X_{t-p}^2 = \sum_{t=p+1}^{n} X_t X_{t-p}.$$ 

If we write (117) in commonly used form

$$X_t = \varphi'X_{t-1} + Y_t,$$

where $X_{t-1} = (X_{t-1}, \ldots, X_{t-p})'$, then the solution is of the form

$$\tilde{\varphi} = (\tilde{\varphi}_1, \ldots, \tilde{\varphi}_p)' = \left( \sum_{t=p+1}^{n} X_{t-1} X_{t-1}' \right)^{-1} \sum_{t=p+1}^{n} X_{t-1} X_t.$$ 

(123)

The least squares estimator of $\sigma^2$ is

$$\tilde{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} (X_t - \tilde{\varphi}' X_{t-1})^2.$$ 

(124)

It can be shown that estimators $\tilde{\varphi}$ and $\tilde{\sigma}^2$ have the same asymptotic properties as the moment estimators. In particular, as $n \to \infty$

$$\sqrt{n} (\tilde{\varphi} - \varphi) \xrightarrow{d} N_p \left( 0, \sigma^2 \Gamma^{-1} \right)$$

115
and
\[ \tilde{\sigma}^2 \xrightarrow{p} \sigma^2, \]
where \( \Gamma \) is the same matrix as in Theorem 59 (Brockwell, Davis (1991), Chap. 8).

**Maximum likelihood estimators**

The maximum likelihood method assumes that the distribution of random variables from which we are intended to construct estimators of parameters under consideration is known.

Consider first a sequence \( \{X_t, t \in \mathbb{Z}\} \), that satisfies model \( X_t = \varphi X_{t-1} + Y_t \), where \( Y_t \) are iid random variables with distribution \( \mathcal{N}(0, \sigma^2) \). We assume causality, i.e., \( |\varphi| < 1 \).

Let us have observations \( X_1, \ldots, X_n \). From the causality and independence assumption it follows that random variables \( X_1 \) and \( (Y_2, \ldots, Y_n) \) are jointly independent with the density
\[ f(x_1, y_2, \ldots, y_n) = f_1(x_1) f_2(y_2, \ldots, y_n) = f_1(x_1) (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^{n} y_t^2 \right\}. \]

From the causality it also follows that random variable \( X_1 \) has the distribution \( \mathcal{N}(0, \tau^2) \), where \( \tau^2 = \frac{\sigma^2}{1-\varphi^2} \). By transformation density theorem we easily obtain that the joint density of \( X_1, \ldots, X_n \) is
\[ f(x_1, \ldots, x_n) = (2\pi \sigma^2)^{-n/2} \sqrt{1-\varphi^2} \exp \left\{ -\frac{1}{2\sigma^2} \left( (1-\varphi^2)x_1^2 + \sum_{t=2}^{n} (x_t - \varphi x_{t-1})^2 \right) \right\}. \quad (125) \]

The likelihood function \( L(\varphi, \sigma^2) \) is of the same form as (125). Maximum likelihood estimates are then \( \hat{\varphi}, \hat{\sigma}^2 \), that maximize \( L(\varphi, \sigma^2) \) on a given parametric space.

These are the unconditional maximum likelihood estimators and even in this simple model the task to maximize the likelihood function leads to a non-linear optimization problem.

More simple solution is provided by using a conditional maximum likelihood method.

We can easily realize that the conditional density of \( X_2, \ldots, X_n \) given fixed \( X_1 = x_1 \) in our AR(1) model is
\[ f(x_2, \ldots, x_n | x_1) = (2\pi \sigma^2)^{-(n-1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^{n} (x_t - \varphi x_{t-1})^2 \right\}. \quad (126) \]

The conditional maximum likelihood estimators are obtained by maximizing function (126) with respect to \( \varphi \) and \( \sigma^2 \).

Similarly, if we consider a general causal AR(\( p \)) sequence (117), where \( Y_t \) are iid with distribution \( \mathcal{N}(0, \sigma^2) \), we can prove that the conditional density of \( (X_{p+1}, \ldots, X_n)' \) given \( X_1 = x_1, \ldots, X_p = x_p \) is
\[ f(x_{p+1}, \ldots, x_n | x_1, \ldots, x_p) = (2\pi \sigma^2)^{-(n-p)/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} (x_t - \varphi' x_{t-1})^2 \right\}, \]

116
where \( \mathbf{x}_{t-1} = (x_{t-1}, \ldots, x_{t-p})' \), and \( \varphi = (\varphi_1, \ldots, \varphi_p)' \).

By maximization of this function with respect to \( \varphi_1, \ldots, \varphi_p, \sigma^2 \) we get the conditional maximum likelihood estimators. It can be easily shown that under normality, these estimators are numerically equivalent to the least squares estimators.

### 15.2 Estimation of parameters in MA and ARMA models

In the previous paragraph we have seen that in AR models, moment estimators, as well as least squares estimators and conditional maximum likelihood estimators are computationally very simple since we are dealing with linear regression functions. In MA and generally in ARMA models the problem is more complicated since the estimation equations are generally non-linear. We will mention only a few basic methods.

**Moment method in MA(\( q \))**

Consider an MA(\( q \)) sequence defined by

\[
X_t = Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \quad t \in \mathbb{Z},
\]

where \( \{Y_t, t \in \mathbb{Z}\} \) is WN(0, \( \sigma^2 \)). Suppose that \( \theta_1, \ldots, \theta_q, \sigma^2 \) are unknown real-valued parameters to be estimated from \( X_1, \ldots, X_n \).

Recall that the autocovariance function of the MA(\( q \)) sequence under consideration is

\[
R_X(k) = \begin{cases} 
\sigma^2 \sum_{j=0}^{q-|k|} \theta_j \theta_{j+|k|}, & |k| \leq q, \\
0, & \text{elsewhere} 
\end{cases}
\]  
(127)

(we put \( \theta_0 = 1 \).)

Moment estimators of \( \theta_1, \ldots, \theta_q, \sigma^2 \) can be obtained by solving the system of equations (127) for \( k = 0, 1, \ldots, q \), where we replace the values of \( R_X(k) \) by the sample values \( \hat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \overline{X}_n) (X_{t+k} - \overline{X}_n) \). We get the system of \( q+1 \) equations

\[
\hat{R}(0) = \sigma^2 \left( 1 + \theta_1^2 + \cdots + \theta_q^2 \right),
\]
\[
\hat{R}(1) = \sigma^2 \left( \theta_1 + \theta_1 \theta_2 + \cdots + \theta_q \theta_{q-1} \right),
\]
\[
\vdots
\]
\[
\hat{R}(q) = \sigma^2 \theta_q.
\]

This system however need not have the unique solution.

**Example 54.** Consider MA(\( 1 \)) model \( X_t = Y_t + \theta Y_{t-1} \), where \( Y_t \) is WN(0, \( \sigma^2 \)) and \( \theta \neq 0 \). Obviously, \( R_X(0) = \sigma^2 (1 + \theta^2) \), \( R_X(1) = \sigma^2 \theta \), thus

\[
r(1) = \frac{R_X(1)}{R_X(0)} = \frac{\theta}{1 + \theta^2}.
\]
It can be shown that in this case, $|r(1)| \leq \frac{1}{2}$ for all real values of $\theta$. Consequently, solving the last equation with respect to $\theta$, we get either the twofold root $\theta = \frac{1}{2r(1)}$ or two real valued roots

$$\theta_{1,2} = \frac{1 \pm \sqrt{1 - 4r^2(1)}}{2r(1)}.$$  

The root with the positive sign is in absolute value larger than 1, while the other one is in absolute value less than 1, which corresponds to an invertible process.

The moment estimators of $\theta$ a $\sigma^2$ now can be obtained from equations (128) that can be rewritten into the form

$$\hat{R}(0) = \sigma^2 (1 + \theta^2),$$

$$\hat{r}(1) = \frac{\theta}{1 + \theta^2}.$$  

For $\theta$ we have two solutions

$$\hat{\theta}_{1,2} = \frac{1 \pm \sqrt{1 - 4\hat{r}^2(1)}}{2\hat{r}(1)},$$

that take real values if $|\hat{r}(1)| \leq \frac{1}{2}$.

Provided that the process is invertible and $|\hat{r}(1)| < \frac{1}{2}$, then the moment estimators are

$$\hat{\theta} = \frac{1 - \sqrt{1 - 4\hat{r}^2(1)}}{2\hat{r}(1)},$$

$$\hat{\sigma}^2 = \frac{\hat{R}(0)}{1 + \hat{\theta}^2}.$$  

If $|\hat{r}(1)| = \frac{1}{2}$, we take

$$\hat{\theta} = \frac{1}{2\hat{r}(1)} = \frac{\hat{r}(1)}{|\hat{r}(1)|},$$

$$\hat{\sigma}^2 = \frac{1}{2} \hat{R}(0).$$  

For $|\hat{r}(1)| > \frac{1}{2}$ the real-valued solution of (128) does not exist. In such a case we use the same estimates as given for $|\hat{r}(1)| = \frac{1}{2}$.

Similarly we can proceed to obtain moment estimators in ARMA models. For a causal and invertible ARMA($p,q$) process

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q},$$
where $Y_t \sim WN(0, \sigma^2)$, $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2$ are unknown parameters and $X_1, \ldots, X_n$ are given observations, we can proceed as follows:

First we use an analogy of the Yule-Walker equations for the autocovariances $R_X(k)$, $k = q + 1, \ldots, q + p$. We get equations

$$ R_X(k) = \varphi_1 R_X(k - 1) + \cdots + \varphi_p R_X(k - p) $$

for unknown parameters $\varphi_1, \ldots, \varphi_p$. If we replace the theoretical values $R_X$ by their estimates $\hat{R}_X(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n) (X_{t+k} - \bar{X}_n)$, we obtain estimates of parameters $\hat{\varphi}_1, \ldots, \hat{\varphi}_p$.

Further we put $Z_t = X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}$ and want to estimate $\theta_1, \ldots, \theta_q$ and $\sigma^2$ in the MA($q$) model

$$ Z_t = Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}. $$

Compute the autocovariance function of the sequence $\{Z_t, t \in \mathbb{Z}\}$. Since

$$ Z_t = \sum_{j=0}^{p} \beta_j X_{t-j}, $$

where $\beta_0 = 1, \beta_j = -\varphi_j, j = 1, \ldots, p$, we have

$$ R_Z(k) = \sum_{j=0}^{p} \sum_{l=0}^{p} \beta_j \beta_l R_X(k + j - l), \ k \in \mathbb{Z}. $$

Estimates of $\theta_1, \ldots, \theta_q$ a $\sigma^2$ are obtained from (128 replacing $\hat{R}(k)$ by estimates

$$ \hat{R}_Z(k) = \sum_{j=0}^{p} \sum_{l=0}^{p} \hat{\beta}_j \hat{\beta}_l \hat{R}_X(k + j - l), $$

where $\hat{\beta}_j = -\hat{\varphi}_j$ and $\hat{R}_X(k)$ are sample autocovariances computed from $X_1, \ldots, X_n$.

The moment estimators are under some assumptions consistent and asymptotically normal, but they are not too stable and and must be handled carefully. Nevertheless they can serve as preliminary estimates in more advanced estimation procedures.

**Two-step least squares estimators in MA and ARMA models**

Consider a causal and invertible ARMA($p, q$) process

$$ X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, $$

where $Y_t \sim WN(0, \sigma^2)$, $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2$ are unknown parameters and $X_1, \ldots, X_n$ are given observations.
Under the invertibility assumptions, the process has an AR(∞) representation

\[ Y_t = \sum_{j=0}^{\infty} d_j X_{t-j} = X_t + \sum_{j=1}^{\infty} d_j X_{t-j} \]

(see Theorem 37.) This can be used to obtain parameters \( \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2 \) as follows.

- Approximate \( X_t \) by an autoregressive process of a sufficiently large order \( m \), where \( m \geq p \), i.e., consider model

\[ X_t = \alpha_1 X_{t-1} + \cdots + \alpha_m X_{t-m} + \tilde{Y}_t, \ t = m+1, \ldots, n \]

and using \( X_1, \ldots, X_n \) estimate \( \alpha_1, \ldots, \alpha_m \) by the least squares method. Obtained estimates are \( \hat{\alpha}_1, \ldots, \hat{\alpha}_m \).

- Estimate residuals \( \hat{\tilde{Y}}_t, t = m+1, \ldots, n \) and use them as known regressors in the regression model

\[ X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \theta_1 \hat{\tilde{Y}}_{t-1} + \cdots + \theta_q \hat{\tilde{Y}}_{t-q} + Y_t, \ t = \text{max}(p, q, m)+1, \ldots, n \]

and estimate \( \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2 \) from this regression model with regressors \( X_{t-1}, \ldots, X_{t-p}, \hat{\tilde{Y}}_{t-1}, \hat{\tilde{Y}}_{t-q} \).

For other estimating methods see, e.g. Prášková, 2006.

16 Periodogram

**Definition 43.** Let \( X_1, \ldots, X_n \) be observations of a random sequence \( \{X_t, t \in \mathbb{Z}\} \). The *periodogram* of \( X_1, \ldots, X_n \) is defined by

\[
I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-i\lambda} \right|^2, \quad \lambda \in [-\pi, \pi]. \tag{129}
\]

To compute the values of the periodogram, it is more convenient to consider it in the form

\[
I_n(\lambda) = \frac{1}{4\pi} \left[ A^2(\lambda) + B^2(\lambda) \right], \tag{130}
\]

where

\[
A(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \cos t\lambda, \quad B(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \sin t\lambda. \tag{131}
\]

120
For a real-valued sequence, the periodogram can be also expressed as

\[ I_n(\lambda) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} X_t X_s e^{-i(t-s)\lambda} = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \sum_{s=\max(1,1-k)}^{\min(n,n-k)} X_s X_{s+k} e^{-ik\lambda} \]

\[ = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} e^{-ik\lambda} C_k, \]  

(132)

where

\[ C_k = \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}, \quad k \geq 0 \]

\[ = C_{-k}, \quad k < 0 \]  

(133)

**Distribution of values of the periodogram**

**Theorem 60.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a centered weakly stationary real-valued sequence with the autocovariance function \( R \), such that \( \sum_{k=-\infty}^{\infty} |R(k)| < \infty \). Then

\[ E I_n(\lambda) \to f(\lambda), \quad \lambda \in [-\pi, \pi], \]  

(134)

where \( f \) denotes the spectral density of the sequence \( \{X_t, t \in \mathbb{Z}\} \).

**Proof.** From formula (132), using the stationarity and the centrality we get

\[ E I_n(\lambda) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} e^{i\lambda R(k)(n - |k|)}. \]

Under the assumptions of the theorem and according to Theorem 22, the spectral density of the sequence \( \{X_t, t \in \mathbb{Z}\} \) exists and is given by

\[ f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} R(k). \]

Thus, using the same arguments as in Theorem 22, we have, as \( n \to \infty \),

\[ |f(\lambda) - E I_n(\lambda)| \leq \frac{1}{2\pi} \sum_{|k| \geq n} |R(k)| + \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} |R(k)||k| \to 0. \]  

(135)

\[ \square \]

Usually, the periodogram is computed at points \( \lambda_j = \frac{2\pi j}{n}, \lambda_j \in [-\pi, \pi] \) (Fourier frequencies).
Theorem 61. Let \( \{X_t, t \in \mathbb{Z}\} \) be a Gaussian random sequence of iid random variables with zero mean and variance \( \sigma^2 \), \( 0 < \sigma^2 < \infty \). Put \( n = 2m + 1 \) and consider the periodogram \( I_n \) computed from \( X_1, \ldots, X_n \) at frequencies \( \lambda_j = \frac{2\pi j}{n}, j = 1, \ldots, m \). Then random variables \( I_n(\lambda_1), \ldots, I_n(\lambda_m) \) are independent and identically distributed as \( \frac{\sigma^2}{4\pi} \chi^2(2) \), where \( \chi^2(2) \) denotes the \( \chi^2 \) distribution with two degrees of freedom.

Proof. Consider random vector \( J = (A(\lambda_1), \ldots, A(\lambda_m), B(\lambda_1), \ldots, B(\lambda_m))' \), where the variables \( A(\lambda_j), B(\lambda_j) \) are defined in (131). This vector has jointly normal distribution since it is a linear transformation of random vector \( (X_1, \ldots, X_n)' \). Further we prove that all the components of the vector \( J \) are mutually uncorrelated (and thus, independent), and identically distributed with zero mean and variance \( \sigma^2 \). For this we use the following identities for trigonometric functions

\[
\sum_{t=1}^{n} \cos^2(t\lambda_r) = \frac{n}{2} \quad r = 1, \ldots, m \\
\sum_{t=1}^{n} \sin^2(t\lambda_r) = \frac{n}{2} \quad r = 1, \ldots, m \\
\sum_{t=1}^{n} \sin(t\lambda_r) \cos(t\lambda_s) = 0 \quad r, s = 1, \ldots, m \\
\sum_{t=1}^{n} \sin(t\lambda_r) \sin(t\lambda_s) = 0 \quad r, s = 1, \ldots, m, r \neq s \\
\sum_{t=1}^{n} \cos(t\lambda_r) \cos(t\lambda_s) = 0 \quad r, s = 1, \ldots, m, r \neq s
\]

from that the result follows using simple computations. Particularly, we get for any \( r = 1, \ldots, m \) that \( A(\lambda_r) \sim \mathcal{N}(0, \sigma^2), B(\lambda_r) \sim \mathcal{N}(0, \sigma^2) \), thus

\[
\frac{A(\lambda_r) + B^2(\lambda_r)}{\sigma^2} = \frac{4\pi}{\sigma^2} I_n(\lambda_r) \sim \chi^2(2).
\]

Remark 16. From the assumption that \( \{X_t, t \in \mathbb{Z}\} \) is a Gaussian random sequence of iid random variables with zero mean and variance \( \sigma^2 \) we can easily conclude that the spectral density of this sequence is

\[
f(\lambda) = \frac{\sigma^2}{2\pi}, \quad \lambda \in [-\pi, \pi]
\]

since \( \{X_t, t \in \mathbb{Z}\} \) is the white noise. From Theorem 61 and properties of the \( \chi^2 \) distribution we have for \( r = 1, \ldots, m \)

\[
\mathbb{E}I_n(\lambda_r) = \frac{2\sigma^2}{4\pi} = \frac{\sigma^2}{2\pi} = f(\lambda_r)
\]
\[ \text{var } I_n(\lambda_r) = 4 \frac{\sigma^4}{16\pi^2} = f^2(\lambda_r). \]

We can see that the variance of the periodogram in this case does not depend on \( n \). More generally, it can be proved that for any Gaussian stationary centered sequence with a continuous spectral density \( f \) it holds that

\[
\lim_{n \to \infty} \text{var } I_n(\lambda) = f^2(\lambda), \quad \lambda \neq 0, \lambda \in (-\pi, \pi)
\]

\[
= 2f^2(\lambda), \quad \lambda = 0, \lambda = \pm \pi
\]

(Anděl, 1976, p. 103, Theorem 10). We see that the variance of the periodogram does not converge to zero with increasing \( n \). It means that the periodogram is not consistent estimator of the spectral density.

Periodogram was originally proposed to detect hidden periodic components in a time series. To demonstrate it, let us consider a sequence \( \{X_t, t \in \mathbb{Z}\} \) such that

\[ X_t = \alpha e^{it\lambda_0} + Y_t, \; Y_t \sim \text{WN}(0, \sigma^2) \]

where \( \alpha \) is a nonzero constant and \( \lambda_0 \in [-\pi, \pi] \).

Then

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t e^{-it\lambda} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t e^{-it\lambda} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e^{-it(\lambda - \lambda_0)}
\]

(136)

and from here we can see that if \( \lambda = \lambda_0 \), the nonrandom part of the periodogram represented by the second sum on the right-hand side of (136) tends to \( \infty \) as \( n \to \infty \) while for \( \lambda \neq \lambda_0 \) it is negligible. It means that if there is a single periodic component at frequency \( \lambda_0 \) the periodogram takes in it the largest value. Since usually the frequency \( \lambda_0 \) is unknown, it is reasonable to consider maximum of the values of the periodogram at the Fourier frequencies.

**Theorem 62.** Let \( \{X_t, t \in \mathbb{Z}\} \) be a Gaussian random sequence of iid random variables with zero mean and variance \( \sigma^2 \). Let \( n = 2m + 1 \) and \( I_n(\lambda_r) \) be the periodogram computed from \( X_1, \ldots, X_n \) at the frequencies \( \lambda_r = \frac{2\pi r}{n}, r = 1, \ldots, m \). Then the distribution of the statistic

\[ W = \frac{\max_{1 \leq r \leq m} I_n(\lambda_r)}{I_n(\lambda_1) + \cdots + I_n(\lambda_m)} \]

(137)

has density

\[ g(x) = m(m-1) \sum_{j=1}^{[1/x]} (-1)^{j-1} \binom{m-1}{j-1} (1-jx)^{m-2}, \quad 0 < x < 1 \]

and

\[ P(W > x) = 1 - \sum_{k=0}^{[1/x]} (-1)^k \binom{m}{k} (1-kx)^{m-1}, \quad 0 < x < 1. \]

(138)
Figure 15: Periodogram of the Sunspots, Wolf index. The maximum corresponds to the cycle with period 11.0769 years.

*Proof.* Anděl (1976), pp. 79–82.

**Fisher test of periodicity** We want to test the null hypothesis of no periodic component $H_0 : \{X_1, \ldots, X_n\}$ are iid with distribution $N(0, \sigma^2)$ against the alternative that the null hypothesis is violated. Test statistic is based on Theorem 62 and reject the null hypothesis at level $\alpha$ if $W > c_\alpha$ where $c_\alpha$ is a critical value that can be computed from (138).

**Estimators of spectral density**

We have seen in Theorem 60 that the periodogram is an asymptotically unbiased estimator of the spectral density, but it is not consistent since its variance does not converge to zero neither is the simplest case of the Gaussian white noise. It can be however shown that under some smoothing assumptions,

$$\int_{-\pi}^{\pi} I_n(\lambda)K(\lambda)d\lambda$$

where $K$ is a kernel function with properties

$$K(\lambda) \geq 0, \quad K(\lambda) = K(-\lambda), \quad \int_{-\pi}^{\pi} K(\lambda)d\lambda = 1, \quad \int_{-\pi}^{\pi} K^2(\lambda)d\lambda < \infty,$$
is an asymptotically unbiased and consistent estimator of
\[ \int_0^{\pi} f(\lambda)K(\lambda)d\lambda \]
and a consistent estimator of \( f(\lambda) \) is considered to be
\[ \hat{f}(\lambda) = w_0C_0 + 2\sum_{k=1}^{n-1} C_k w_k \cos(k\lambda) \]
where \( C_k \) are given by (133) and \( w_k \) are the Fourier coefficients of the function \( K \). One of the commonly used kernel function is so called Parzen window, which is usually presented by coefficients
\[
w_k = \begin{cases} 
1 - 6\left(\frac{k}{M}\right)^2 - 6\left(\frac{|k|}{M}\right)^3, & |k| < \frac{M}{2} \\
2\left(1 - \frac{|k|}{M}\right)^3, & \frac{M}{2} < |k| \leq M \\
0, & |k| > M \end{cases}
\]
where \( M \) is so-called truncation point that depends on \( n \left( \frac{n}{\pi} < M < \frac{n}{\pi^2} \right) \). For more information on the choice of \( K \), respectively of \( w_k \), see, e.g., Anděl, 1976, or Brockwell and Davis, Chap. 10.