FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Department of Probability and Mathematical Statistics

Stochastic Processes 2

Lecture Notes

Zuzana Prášková

2016

Contents

1 Definitions and basic characteristics

1.1 Definition of a stochastic process

Definition 1. Let $(\Omega, \mathcal{A}, \mathsf{P})$ 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 $(\Omega, \mathcal{A}, \mathsf{P})$ with values in S is called *a stochastic (random) process.*
If $S = \mathbb{R}$, $\{X_t,\, t \in T\}$ is called *the real-valued stochastic process.*
If $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ or $T \subset \mathbb{Z}$, $\{X_t,\, t \in T\}$ is called *the discrete time stochastic process*, time series.
If $T = [a, b]$, $-\infty \leq a < b \leq \infty$, $\{X_t,\, t \in T\}$ is *the continuous time stochastic 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. The pair $(S, \mathcal{E})$, where S is the set of values of random variables X_t and $\mathcal{E}$ is the $\sigma-$algebra of subsets of S, is called *the state space* of the process $\{X_t,\, t \in T\}$.

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

Finite-dimensional distributions of a stochastic process:

Let $\{X_t,\, t \in T\}$ be a stochastic process. Then, $\forall n \in \mathbb{N}$ and any finite subset $\{t_1, \dots, t_n\} \subset T$ there is a system of random variables $X_{t_1}, \dots, X_{t_n}$ with the joint distribution function

$$\mathsf{P}\left[X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n\right] = F_{t_1,\dots,t_n}(x_1, \dots, x_n)$$

for all real valued $x_1, \dots, x_n$.

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

1. $F_{t_{i_1},\dots,t_{i_n}}(x_{i_1}, \dots, x_{i_n}) = F_{t_1,\dots,t_n}(x_1, \dots, x_n)$ for any permutation $(i_1, \dots, i_n)$ of indices $1, \dots, n$ (symmetry)
2. $\lim_{x_n \to \infty} F_{t_1,\dots,t_n}(x_1, \dots, x_n) = F_{t_1,\dots,t_{n-1}}(x_1, \dots, x_{n-1})$ (consistency)

The characteristic function of a random vector $\mathbf{X} = (X_1, \dots, X_n)$ is

$$\varphi_{\mathbf{X}}(\mathbf{u}) := \varphi(u_1, \dots, u_n) = \mathsf{E}e^{\mathsf{i}\mathbf{u}^\top \mathbf{X}} = \mathsf{E}e^{\mathsf{i}\sum_{j=1}^{n} u_j X_j} \quad \mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}_n.$$

A system of characteristic functions is said to be consistent if

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

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

1.2 Daniell-Kolmogorov theorem

For any stochastic process there exists a consistent system of distribution functions. On the other hand, the following theorem holds.

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 $x_1, \ldots, x_n$*

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

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

1.3 Autocovariance and autocorrelation function

Definition 4. *A complex-valued random variable* X is defined by $X = Y + \mathrm{i}Z$, where Y and Z are real random variables, $\mathrm{i} = \sqrt{-1}$.
The mean value of a complex-valued random variable $X = Y + \mathrm{i}Z$ is defined by $\mathsf{E}X = \mathsf{E}Y + \mathrm{i}\mathsf{E}Z$ provided the mean values $\mathsf{E}Y$ and $\mathsf{E}Z$ exist.
The variance of a complex-valued random variable $X = Y + \mathrm{i}Z$ is defined by $\operatorname{var} X := \mathsf{E}\left[(X - \mathsf{E}X)(\overline{X} - \overline{\mathsf{E}X})\right] = \mathsf{E}|X - \mathsf{E}X|^2 \geq 0$ 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 $(\Omega, \mathcal{A}, \mathsf{P})$.

Definition 6. Let $\{X_t,\, t \in T\}$ be a stochastic process such that $\mathsf{E}X_t := \mu_t$ exists for all $t \in T$. Then the function $\{\mu_t, t \in T\}$ 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, i.e., $\mu_t = 0$ for all $t \in T$.

Definition 7. Let $\{X_t,\, t \in T\}$ be a process with finite second order moments, i.e., $\mathsf{E}|X_t|^2 < \infty$, $\forall t \in T$. Then a (complex-valued) function defined on $T \times T$ by

$$R(s,t) = \mathsf{E}\left[(X_s - \mu_s)(\overline{X}_t - \overline{\mu}_t)\right]$$

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. *The autocorrelation function* of the process $\{X_t,\, t \in T\}$ with positive variances is defined by

$$r(s,t) = \frac{R(s,t)}{\sqrt{R(s,s)}\sqrt{R(t,t)}}, \quad s,t \in T.$$

Definition 9. A stochastic process $\{X_t,\, t \in T\}$ is called *Gaussian*, if for any $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$, the vector $(X_{t_1}, \dots, X_{t_n})^\top$ is normally distributed $\mathcal{N}_n(\mathbf{m_t}, \mathbf{V_t})$, where $\mathbf{m_t} = (\mathsf{E}X_{t_1}, \dots, \mathsf{E}X_{t_n})^\top$ and

$$\mathbf{V_t} = \begin{pmatrix} \mathrm{var}X_{t_1} & \mathrm{cov}(X_{t_1}, X_{t_2}) & \dots & \mathrm{cov}(X_{t_1}, X_{t_n}) \\ \mathrm{cov}(X_{t_2}, X_{t_1}) & \mathrm{var}X_{t_2} & \dots & \mathrm{cov}(X_{t_2}, X_{t_n}) \\ \dots & \dots & \ddots & \dots \\ \mathrm{cov}(X_{t_n}, X_{t_1}) & \mathrm{cov}(X_{t_n}, X_{t_2}) & \dots & \mathrm{var}X_{t_n} \end{pmatrix}.$$

1.4 Strict and weak stationarity

Definition 10. A stochastic process $\{X_t,\, t \in T\}$ is said to be *strictly stationary*, if for any $n \in \mathbb{N}$, for any $x_1, \dots, x_n \in \mathbb{R}$ and for any $t_1, \dots, t_n$ and h such that $t_k \in T, t_k + h \in T, 1 \le k \le n$,

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) = F_{t_1+h,\dots,t_n+h}(x_1, \dots, x_n).$$

Definition 11. A stochastic process $\{X_t,\, t \in T\}$ with finite second order 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 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*.

The autocovariance function of a weakly stationary process is a function of one variable:

$$R(t) := R(t, 0),\ t \in T.$$

The autocorrelation function in such case is

$$r(t) = \frac{R(t)}{R(0)},\ t \in T.$$

Theorem 2. *Any strictly stationary stochastic process $\{X_t,\, t \in T\}$ with finite second order moments is also weakly stationary.*

Proof. If $\{X_t,\, t \in T\}$ is strictly stationary with finite second order moments, X_t are equally distributed for all $t \in T$ with the mean value

$$\mathsf{E}X_t = \mathsf{E}X_{t+h},\ \forall t \in T,\ \forall h:\ t+h \in T.$$

Especially, for $h = -t$, $\mathsf{E}X_t = \mathsf{E}X_0 = \text{const}$.
Similarly, (X_t, X_s) are equally distributed and

$$\mathsf{E}\,[X_t\, X_s] = \mathsf{E}\,[X_{t+h}\, X_{s+h}]\ \ \forall s,t \in T,\ \forall h:\ s+h \in T,\ t+h \in T.$$

Especially, for $h = -t$, $\mathsf{E}\,[X_t X_s] = \mathsf{E}\,[X_0 X_{s-t}]$ is a function of $s - t$. □

Example 1. Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of i.i.d. random variables with a distribution function F. Since for all $n \in \mathbb{N}$ and all $t_1, \dots, t_n, h \in \mathbb{Z}$,

$$\begin{aligned}
F_{t_1,\dots,t_n}(x_1,\dots,x_n) &= \mathsf{P}\,[X_{t_1} \le x_1, \dots, X_{t_n} \le x_n] = \\
&= \prod_{i=1}^{n} \mathsf{P}\,[X_{t_i} \le x_i] = \prod_{i=1}^{n} F(x_i), \\
F_{t_1+h,\dots,t_n+h}(x_1,\dots,x_n) &= \mathsf{P}\,[X_{t_1+h} \le x_1, \dots, X_{t_n+h} \le x_n] = \\
&= \prod_{i=1}^{n} \mathsf{P}\,[X_{t_i+h} \le x_i] = \prod_{i=1}^{n} F(x_i),
\end{aligned}$$

$\{X_t, t \in \mathbb{Z}\}$ is strictly stationary.

Example 2. Let $\{X_t,\, t \in \mathbb{Z}\}$ be a sequence defined by $X_t = (-1)^t X$, where X is a random variable such that

$$X = \begin{cases} -\frac{1}{4} & \text{with probability } \frac{3}{4}, \\ \ \ \frac{3}{4} & \text{with probability } \frac{1}{4}. \end{cases}.$$

Then $\{X_t,\, t \in \mathbb{Z}\}$ is weakly stationary, since

$$\begin{aligned}
\mathsf{E}X_t &= 0, \\
\operatorname{var} X_t &= \sigma^2 = \frac{3}{16}, \\
R(s,t) &= \sigma^2(-1)^{s+t} = \sigma^2(-1)^{s-t},
\end{aligned}$$

but it is not strictly stationary (variables X and $-X$ are not equally distributed).

Theorem 3. *Any 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 $\mathsf{E}X_t = \mu,\ \operatorname{cov}(X_t, X_s) = R(t-s) = \operatorname{cov}(X_{t+h}, X_{s+h}), t, s \in T$, thus, for all $n \in \mathbb{N}$ and all $t_1, \ldots, t_n, h \in \mathbb{Z}$,

$$\mathsf{E}(X_{t_1}, \ldots, X_{t_n}) = \mathsf{E}(X_{t_1+h}, \ldots, X_{t_n+h}) = (\mu, \ldots, \mu) := \boldsymbol{\mu},$$

$$\operatorname{var}(X_{t_1}, \ldots, X_{t_n}) = \operatorname{var}(X_{t_1+h}, \ldots, X_{t_n+h}) := \boldsymbol{\Sigma}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} R(0) & R(t_2-t_1) & \ldots & R(t_n-t_1) \\ R(t_2-t_1) & R(0) & \ldots & R(t_n-t_2) \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & R(0) \end{pmatrix}.$$

Since the joint distribution of a normal vector is uniquely defined by the vector of mean values and the variance matrix, $(X_{t_1}, \ldots, X_{t_n}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $(X_{t_1+h}, \ldots, X_{t_n+h}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ from which the strict stationarity of $\{X_t,\, t \in T\}$ follows. □

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

$$\begin{aligned} |R(s,t)| &= |\mathsf{E}(X_s - \mathsf{E}X_s)(\overline{X_t} - \overline{\mathsf{E}X_t})| \leq (\mathsf{E}|X_s - \mathsf{E}X_s|^2)^{\frac{1}{2}}(\mathsf{E}|X_t - \mathsf{E}X_t|^2)^{\frac{1}{2}} \\ &= \sqrt{R(s,s)}\sqrt{R(t,t)}. \end{aligned}$$

□

Thus, for the autocovariance function of a weakly stationary process we have $R(0) \geq 0$ and $|R(t)| \leq R(0)$.

Definition 12. Let f 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 $t_1, \ldots, t_n \in T$,

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

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 $t_j - t_k \in T$,

$$\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} g(t_j - t_k) \geq 0.$$

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

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 $c_1 = 1, c_2 = \mathrm{i}(= \sqrt{-1})$.

□

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 order moments. Then its autocovariance function is positive semidefinite on $T \times T$.*

Proof. W.l.o.g., suppose that the process is centered. Then for any $n \in \mathbb{N}$, complex constants $c_1, \ldots, c_n$ and $t_1, \ldots, t_n \in T$

$$0 \leq \mathsf{E} \left| \sum_{j=1}^{n} c_j X_{t_j} \right|^2 = \mathsf{E} \left[\sum_{j=1}^{n} c_j X_{t_j} \overline{\sum_{k=1}^{n} c_k X_{t_k}} \right]$$
$$= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} \mathsf{E}(X_{t_j} \overline{X_{t_k}}) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} R(t_j, t_k).$$

□

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 order 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, for any $n \in \mathbb{N}$ and any real $t_1, \dots, t_n \in T$, the matrix

$$\mathbf{V_t} = \begin{pmatrix} R(t_1,t_1) & R(t_1,t_2) & \dots & R(t_1,t_n) \\ R(t_2,t_1) & R(t_2,t_2) & \dots & R(t_2,t_n) \\ \dots & \dots & \dots & \dots \\ R(t_n,t_1) & R(t_n,t_2) & \dots & R(t_n,t_n) \end{pmatrix}$$

is positive semidefinite. The function

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

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

□

Example 3. Decide whether the 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 the positive semidefinite function. Consider $n \in \mathbb{N}$, $c_1, \dots, c_n \in \mathbb{C}$ a $t_1, \dots, t_n \in \mathbb{R}$. Then we have

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

The function $\cos t$ is positive semidefinite, and according to Theorem 6 there exists a (Gaussian) stochastic process $\{X_t,\ t \in T\}$, the 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 $c_1, \dots, c_n$ and $t_1, \dots, t_n \in T$

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c_j\bar{c}_k h(t_j, t_k) = \sum_{j=1}^{n}\sum_{k=1}^{n} c_j\bar{c}_k [f(t_j, t_k) + g(t_j, t_k)]$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{n} c_j\bar{c}_k f(t_j, t_k) + \sum_{j=1}^{n}\sum_{k=1}^{n} c_j\bar{c}_k g(t_j, t_k) \geq 0.$$

□

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

Proof. It follows from Theorems 6–8. □

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 the state space $(S, \mathcal{E})$, if for any $t_0, t_1, \dots, t_n,\ 0 \leq t_0 < t_1 < \dots < t_n$,

$$P(X_{t_n} \leq x | X_{t_{n-1}}, \dots, X_{t_0}) = P(X_{t_n} \leq x | X_{t_{n-1}}) \ \text{ a. s.} \tag{1}$$

for all $x \in \mathbb{R}$.

Relation (1) is called *the Markovian property*. Simple cases are discrete state Markov processes, i.e., discrete and continuous time Markov chains.

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

$$\mathbf{Q} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}, \quad \alpha > 0,\ \beta > 0.$$

Treat the stationarity of this process.

We know that all the finite dimensional distributions of a continuous time Markov chain are determined by the initial distribution $\mathbf{p}(0) = \{p_j(0),\, j \in S\}^T$ and the transition probability matrix $\mathbf{P}(t) = \{p_{ij}(t),\, i, j \in S\}$. In our case, $\mathbf{P}(t) = \exp(\mathbf{Q}t)$, where

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

(see, e.g., Prášková and Lachout, 2012, pp. 93–95) and due to the initial distribution, the absolute distribution is

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

Then we have

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

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

On the other hand, if the initial distribution is the stationary distribution of the Markov chain, i.e., such probability distribution that satisfies $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T\mathbf{P}(t)$, then $\{X_t,\, t \geq 0\}$ is the strictly stationary process (Prášková and Lachout, 2012, Theorem 3.12).

In our case, the solution of $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T\mathbf{P}(t)$ gives

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

and from here we get the constant mean $\mathsf{E}X_t = \frac{\alpha}{\alpha+\beta}$ and the autocovariance function

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

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 λ 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 λt. Increments $X_t - X_s,\ s < t$ have the Poisson distribution with the parameter $\lambda(t-s)$. The 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 σ^2 is a positive constant. Especially, for any $t \geq 0$, $\mathsf{E}W_t = 0$ and $\operatorname{var} W_t = \sigma^2 t$.

The Wiener process is Markov 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). A system of σ-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. $\forall \alpha \in \mathbb{C}, \forall x, y \in H : \langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

3. $\forall x, y, z \in H\ \langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

4. $\forall x \in H$ is $\langle x, x\rangle \geq 0$; $\langle x, x\rangle = 0 \Leftrightarrow x = 0$ (0 is zero element in H).

The number

$$\|x\| := \sqrt{\langle x, x\rangle}, \quad \forall x \in H$$

is called *the norm* of an element x.

Theorem 9. *For the norm $\|x\| := \sqrt{\langle x, x\rangle}$ the following properties hold:*

1. $\|x\| \geq 0\ \forall x \in H$ *and* $\|x\| = 0 \Leftrightarrow x = 0$.

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. $\forall x, y \in H$: $|\langle x, y\rangle| \leq \|x\| \cdot \|y\| = \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}$
   (the Cauchy-Schwarz inequality).

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

3.2 Convergence in norm

Definition 20. We say that a sequence $\{x_n,\, n \in \mathbb{N}\}$ of elements of an inner product space H *converges in norm* to and 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

$$||x_n|| \to ||x||$$
$$\langle x_n, y_n\rangle \to \langle x, y\rangle.$$

Proof. From the triangle inequality we get

$$\begin{aligned} ||x|| &\leq ||x - y|| + ||y|| \\ ||y|| &\leq ||y - x|| + ||x| \\ \big|\, ||x|| - ||y||\, \big| &\leq\ ||x - y||. \end{aligned}$$

From here we get the first assertion since

$$\big|\, ||x_n|| - ||x||\, \big| \leq\ ||x_n - x||.$$

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

$$\begin{aligned} |\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n - x + x, y_n - y + y\rangle - \langle x, y\rangle| \\ &\leq |\langle x_n - x, y_n - y\rangle| + |\langle x, y_n - y\rangle| + \\ &+ |\langle x_n - x, y\rangle| \\ &\leq \|x_n - x\| \cdot \|y_n - y\| + \|x\| \cdot \|y_n - y\| + \\ &+ \|x_n - x\| \cdot \|y\|. \end{aligned}$$

□

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 the set of all random variables with finite second order moments defined on a probability space $(\Omega, \mathcal{A}, \mathsf{P})$. We can easily verify that $\mathcal{L}$ is the vector space:

1. $\forall X, Y \in \mathcal{L},\ X + Y \in \mathcal{L}$ since

$$\mathsf{E}|X+Y|^2 \leq 2\left(\mathsf{E}|X|^2 + \mathsf{E}|Y|^2\right) < \infty.$$

2. $\forall X \in \mathcal{L}$ and $\forall \alpha \in \mathbb{C}$, $\alpha X \in \mathcal{L}$, since

$$\mathsf{E}|\alpha X|^2 = |\alpha|^2 \cdot \mathsf{E}|X|^2 < \infty.$$

3. The null element of $\mathcal{L}$ is the random variable identically equal to zero.

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

$$X \sim Y \Longleftrightarrow \mathsf{P}\left[X = Y\right] = 1$$

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

$$\langle X, Y\rangle = \mathsf{E}\left[\, X\overline{Y}\,\right], \quad \forall X \in \widetilde{X}, Y \in \widetilde{Y},$$

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

The space of classes of equivalence on $\mathcal{L}$ with the above relation $\langle ., .\rangle$ is denoted $L_2(\Omega, \mathcal{A}, \mathsf{P})$. The relation $\langle X, Y\rangle$ satisfies the properties of the inner product on $L_2(\Omega, \mathcal{A}, \mathsf{P})$: For every $X, Y, Z\ \in L_2(\Omega, \mathcal{A}, \mathsf{P})$ and every $\alpha \in \mathbb{C}$ it holds

1. $\langle \alpha X, Y\rangle = \mathsf{E}\left[\alpha X\overline{Y}\right] = \alpha\mathsf{E}\left[X\overline{Y}\right] = \alpha\langle X, Y\rangle$.
2. $\langle X + Y, Z\rangle = \mathsf{E}\left[(X+Y)\overline{Z}\right] = \mathsf{E}\left[X\overline{Z}\right] + \mathsf{E}\left[Y\overline{Z}\right] = \langle X, Z\rangle + \langle Y, Z\rangle$.
3. $\langle X, X\rangle = \mathsf{E}\left[X\overline{X}\right] = \mathsf{E}|X|^2 \geq 0$.
4. $\langle X, X\rangle = \mathsf{E}|X|^2 = 0 \Leftrightarrow X \sim 0$.

4.2 Mean square convergence

We have defined $L_2(\Omega, \mathcal{A}, \mathsf{P})$ to be the space of classes of equivalence on $\mathcal{L}$ with the inner product

$$\langle X, Y \rangle = \mathsf{E}\left[X\overline{Y} \right],$$

the norm is therefore defined by

$$||X|| := \sqrt{\mathsf{E}|X|^2}$$

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

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

$$||X_n - X||^2 = \mathsf{E}|X_n - X|^2 \to 0 \text{ as } n \to \infty.$$

Notation: $X = \text{l. i. m. } X_n$ (*limit in the (squared) mean*).

Theorem 11. *The space $L_2(\Omega, \mathcal{A}, P)$ is complete.*

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

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

Convention: A stochastic process $\{X_t,\, t \in T\}$ such that $\mathsf{E}|X_t|^2 < \infty$ 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 $(\Omega, \mathcal{A}, \mathsf{P})$. 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 the 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 $L_2(\Omega, \mathcal{A}, \mathsf{P})$ 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 $\{X_t^h,\, t \in T\}_{h\in S}, T \subset \mathbb{R}, S \subset \mathbb{R}$, be a collection of stochastic processes in $L_2(\Omega, \mathcal{A}, \mathsf{P})$ (shortly: second order processes). We say that processes $\{X_t^h,\, t \in T\}_{h\in S}$ *converge in mean square* to a second order process $\{X_t,\, t \in T\}$ as $h \to h_0$, if

$$\forall t \in T:\ X_t^h \underset{h\to h_0}{\longrightarrow} X_t \quad \text{in mean square, i.e., } \mathsf{E}|X_t^h - X_t|^2 \underset{h\to h_0}{\longrightarrow} 0.$$

Briefly, we write

$$\{X_t^h,\, t \in T\}_{h\in S} \underset{h\to h_0}{\longrightarrow} \{X_t,\, t \in T\} \quad \text{in mean square.}$$

Theorem 12. *Centered second order processes* $\{X_t^h,\, t \in T\}_{h\in S}$ *converge in mean square to a centered second order process* $\{X_t,\, t \in T\}$ *as* $h \to h_0$ *if and only if*

$$\mathsf{E}\left[X_t^h \overline{X_t^{h'}} \right] \to b(t) \ \text{ as } h, h' \to h_0,$$

where $b(.)$ *is a finite function on* T.

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

Proof. 1. Let $\{X_t^h,\, t \in T\}_{h\in S} \underset{h\to h_0}{\longrightarrow} \{X_t,\, t \in T\}$ in mean square. Then $\forall t, t' \in T$

$$\begin{aligned} X_t^h &\underset{h\to h_0}{\longrightarrow} X_t \ \text{ in mean square} \\ X_{t'}^{h'} &\underset{h'\to h_0}{\longrightarrow} X_{t'} \ \text{ in mean square.} \end{aligned}$$

From the continuity of the inner product we get, as $h, h' \to h_0$,

$$\mathsf{E}\left[X_t^h \overline{X_{t'}^{h'}} \right] \longrightarrow \mathsf{E}\left[X_t \overline{X_{t'}} \right].$$

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

$$\mathsf{E}\left[X_t^h \overline{X_t^{h'}} \right] \longrightarrow \mathsf{E}\left[X_t \overline{X_t} \right] = \mathsf{E}|X_t|^2 := b(t) < \infty,$$

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

$$\mathsf{E}\left[X_t^h \overline{X_{t'}^h} \right] \underset{h \to h_0}{\longrightarrow} \mathsf{E}\left[X_t \overline{X_{t'}} \right] \quad \text{as } t, t' \in T,$$

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

2. Let $\{X_t^h,\, t \in T\}_{h \in S}$ be centered second order processes for which

$$\mathsf{E}\left[X_t^h \overline{X_t^{h'}} \right] \to b(t) < \infty \text{ as } h, h' \to h_0 \text{ a } \forall t \in T.$$

Then

$$\|X_t^h - X_t^{h'}\|^2 \to 0, \text{ as } h, h' \to h_0, \forall t \in T$$

since $\forall t \in T$

$$\begin{aligned} \|X_t^h - X_t^{h'}\|^2 &= \mathsf{E}\left[(X_t^h - X_t^{h'})(\overline{X_t^h} - \overline{X_t^{h'}}) \right] \\ &= \mathsf{E}\left[X_t^h \overline{X_t^h} \right] - \mathsf{E}\left[X_t^{h'} \overline{X_t^h} \right] - \mathsf{E}\left[X_t^h \overline{X_t^{h'}} \right] + \\ &+ \mathsf{E}\left[X_t^{h'} \overline{X_t^{h'}} \right] \longrightarrow b(t) - b(t) - b(t) + b(t) = 0 \end{aligned}$$

as $h, h' \to h_0$.
We have proved that processes $\{X_t^h,\, t \in T\}_{h \in S}$ satisfy the Cauchy property for any $t \in T$. Due to the completeness of $L_2(\Omega, \mathcal{A}, \mathsf{P})$, $\forall t \in T, \exists X_t \in L_2(\Omega, \mathcal{A}, \mathsf{P})$ such that $X_t^h \to X_t$ in mean square as $h \to h_0$, thus $\mathsf{E}|X_t|^2 < \infty$ $\forall t \in T$. Therefore there exists a limit process $\{X_t,\, t \in T\} \in L_2(\Omega, \mathcal{A}, \mathsf{P})$. We prove that $\{X_t,\, t \in T\}$ is centered:

$$\mathsf{E}X_t = \mathsf{E}X_t - \mathsf{E}X_t^h + \mathsf{E}X_t^h = \mathsf{E}\left[X_t - X_t^h \right].$$

Then

$$|\mathsf{E}X_t| = \left| \mathsf{E}\left[X_t - X_t^h \right] \right| \le \sqrt{\mathsf{E}\left| X_t - X_t^h \right|^2} \to 0$$

as $h \to h_0$, $\forall t \in T$. □

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 point* $t_0 \in T$, if

$$\mathsf{E}|X_t - X_{t_0}|^2 \to 0 \quad \text{as } t \to t_0.$$

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

$$\mathsf{P}\Big[\,|X_t - X_{t_0}| > \varepsilon\Big] \leq \varepsilon^{-2} \cdot \mathsf{E}\,|X_t - X_{t_0}|^2\,.$$

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 $s = t$.*

Proof. 1. Let $\{X_t,\ t \in T\}$ be a centered mean square continuous process. We prove that its autocovariance function is continuous at every point of $T \times T$. Since $\mathsf{E}X_t = 0$, we have $\forall s_0, t_0 \in T$ a $s \to s_0,\ t \to t_0$

$$\begin{aligned}|R(s,t) - R(s_0,t_0)| &= \left|\mathsf{E}\left[\,X_s\overline{X}_t\,\right] - \mathsf{E}\left[\,X_{s_0}\overline{X}_{t_0}\,\right]\right| \\ &= |\langle X_s, X_t\rangle - \langle X_{s_0}, X_{t_0}\rangle| \to 0,\end{aligned}$$

which follows from the continuity of the inner product, since $X_t \to X_{t_0}$ as $t \to t_0$ and $X_s \to X_{s_0}$ as $s \to s_0$, due to the continuity of the process.

2. Let $R(s,t)$ be continuous at points $[s,t]$ such that $s = t$. Then $\forall t_0 \in T$

$$\begin{aligned}\mathsf{E}|X_t - X_{t_0}|^2 &= \mathsf{E}\left[(X_t - X_{t_0})(\overline{X}_t - \overline{X}_{t_0})\right] = \\ &= \mathsf{E}\left[X_t\overline{X}_t\right] - \mathsf{E}\left[X_t\overline{X}_{t_0}\right] - \mathsf{E}\left[X_{t_0}\overline{X}_t\right] + \mathsf{E}\left[X_{t_0}\overline{X}_{t_0}\right] \\ &= R(t,t) - R(t,t_0) - R(t_0,t) + R(t_0,t_0).\end{aligned}$$

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 14. *Let $\{X_t,\ t \in T\}$ be a second order 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 at points $[s,t]$, such that $s = t$.*

Proof.

$$\begin{aligned}\mathsf{E}\,|X_t - X_{t_0}|^2 &= \mathsf{E}\,|X_t - \mu_t + \mu_t - X_{t_0}|^2 = \\ &= \mathsf{E}\left[|X_t - \mu_t - (X_{t_0} - \mu_{t_0}) + \mu_t - \mu_{t_0}|^2\right]\end{aligned}$$

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

$$\begin{aligned} \mathsf{E}\,|X_t - X_{t_0}|^2 &= \mathsf{E}\,|Y_t - Y_{t_0} + \mu_t - \mu_{t_0}|^2 \leq \\ &\leq 2\mathsf{E}\,|Y_t - Y_{t_0}|^2 + 2\,|\mu_t - \mu_{t_0}|^2 \end{aligned}$$

□

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 $\mathsf{E}X_t = 0$, $t \in \mathbb{R}$ 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, $\mathsf{E}W_t = 0$, $R(s,t) = \mathsf{E}W_s W_t = \sigma^2 \cdot \min\{s,t\}$. The process is not weakly neither strictly stationary (though Gaussian).
The process je centered, $R(s,t)$ is continuous (thus at $[s,t]$ with $s = t$). The process 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, $X_t \sim \mathrm{Po}(\lambda)$. Since $\mathsf{E}X_t = \mu_t = \lambda t$, $t \geq 0$ and $\mathrm{cov}(X_s, X_t) = \lambda \cdot \min\{s,t\}$, the process is not weakly stationary.
Since μ_t is continuous, $R(s,t)$ is continuous, the process is mean square continuous.

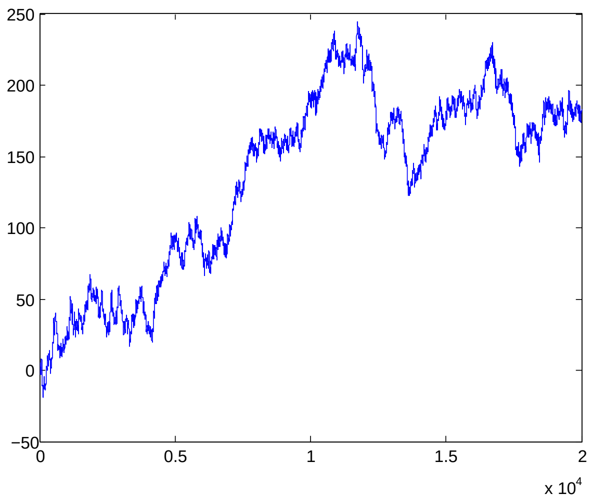

Figure 1: A trajectory of a Wiener process

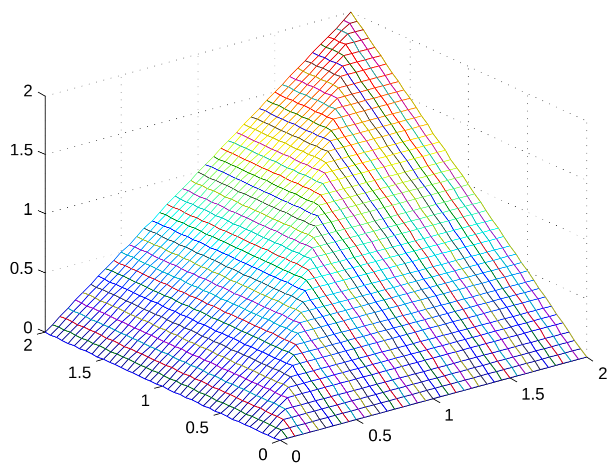

Figure 2: The autocovariance function of a Wiener process

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 $t_0 \in T$* if there exists the mean square limit

$$\underset{h\to 0}{\mathrm{l.\,i.\,m.}}\ \frac{X_{t_0+h} - X_{t_0}}{h} := X'_{t_0}.$$

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*

$$\lim_{h,h'\to 0} \frac{1}{hh'}\left[R(s+h,t+h') - R(s,t+h') - R(s+h,t) + R(s,t)\right].$$

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

$$\lim_{h,h'\to 0} \mathsf{E}\left[\frac{X_{t+h} - X_t}{h} \cdot \frac{\overline{X_{t+h'}} - \overline{X_t}}{h'}\right] =$$
$$\lim_{h,h'\to 0} \frac{1}{hh'}\left[R(t+h,t+h') - R(t,t+h') - R(t+h,t) + R(t,t)\right].$$

□

Remark 3. A sufficient condition the generalized second-order partial derivative of $R(s,t)$ to exist is the following one: Let $[s,t]$ be an interior point in $T \times T$. If there exist

$$\frac{\partial^2 R(s,t)}{\partial s\, \partial t} \text{ and } \frac{\partial^2 R(s,t)}{\partial t\, \partial s}$$

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 the 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

$$\mathsf{E}\left|\frac{X_{t+h}-X_t}{h}-\frac{X_{t+h'}-X_t}{h'}\right|^2 \to 0 \quad \text{as } h \to 0,\ h' \to 0$$

$\forall t \in T$. It holds since

$$\begin{aligned}\mathsf{E}\left|\frac{X_{t+h}-X_t}{h}-\frac{X_{t+h'}-X_t}{h'}\right|^2 \leq 2\mathsf{E}\left|\frac{Y_{t+h}-Y_t}{h}-\frac{Y_{t+h'}-Y_t}{h'}\right|^2 \\ +2\left|\frac{\mu_{t+h}-\mu_t}{h}-\frac{\mu_{t+h'}-\mu_t}{h'}\right|^2,\end{aligned}$$

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. □

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

$$\frac{\partial^2 \cos(s-t)}{\partial s \partial t} \text{ and } \frac{\partial^2 \cos(s-t)}{\partial t \partial s} \text{ exist and they are continuous.}$$

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

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

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 $D_n = \{t_{n,0}, t_{n,1}, \dots, t_{n,n}\}$, where $a = t_{n,0} < t_{n,1} < \dots \dots < t_{n,n} = b$, $\forall n \in \mathbb{N}$ 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

$$I_n := \sum_{i=0}^{n-1} X_{t_{n,i}} \left(t_{n,i+1} - t_{n,i}\right), \ n \in \mathbb{N}.$$

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 the Riemann integral of the process $\{X_t, \, t \in [a,b]\}$ to be

$$\int_a^b X_t dt = \int_a^b \left(X_t - \mu_t\right) \, dt + \int_a^b \mu_t dt,$$

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

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 Δ_m, Δ_n of which converge to zero as $m, n \to \infty$. Put

$$I_m := \sum_{j=0}^{m-1} \left(s_{m,j+1} - s_{m,j}\right) X_{s_{m,j}}$$
$$I_n := \sum_{k=0}^{n-1} \left(t_{n,k+1} - t_{n,k}\right) X_{t_{n,k}}.$$

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

$$\begin{aligned} \mathsf{E}\left[I_m \overline{I_n}\right] &= \lim \mathsf{E}\left\{ \left[\sum_{j=0}^{m-1} X_{s_{m,j}} (s_{m,j+1} - s_{m,j})\right] \cdot \left[\sum_{k=0}^{n-1} \overline{X_{t_{n,k}}} (t_{n,k+1} - t_{n,k})\right] \right\} \\ &= \lim \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} R(s_{m,j}, t_{n,k})(s_{m,j+1} - s_{m,j})(t_{n,k+1} - t_{n,k}) \end{aligned}$$

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. The 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

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

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. a) Let $\mu,\ \nu$ be finite measures on the σ-field of 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 Borel $B \subset (-\pi,\pi)$ and $\mu(\{-\pi\} \cup \{\pi\}) = \nu(\{-\pi\} \cup \{\pi\})$.

b) 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 Anděl (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 the autocovariance function of a stationary random sequence if and only if for any $t \in \mathbb{Z}$,*

$$R(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda) \tag{2}$$

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.

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 $t_1, \ldots, t_n \in \mathbb{Z}$

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

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

2. Let R be the autocovariance function of a stationary random sequence; then it must be positive semidefinite, i.e.,

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

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

$$\begin{aligned}\varphi_n(\lambda) &= \frac{1}{2\pi n}\sum_{j=1}^{n}\sum_{k=1}^{n} e^{-i(j-k)\lambda}R(j-k)\\ &= \frac{1}{2\pi n}\sum_{\kappa=-n+1}^{n-1}\ \sum_{j=max(1,\,\kappa+1)}^{min(n,\,\kappa+n)} e^{-i\kappa\lambda}R(\kappa)\\ &= \frac{1}{2\pi n}\sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda}R(\kappa)(n-|\kappa|).\end{aligned}$$

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

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

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

$$\begin{aligned}F_n(\pi) &= \int_{-\pi}^{\pi}\varphi_n(\lambda)d\lambda\\ &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\left[\sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda}R(\kappa)(n-|\kappa|)\right]d\lambda\\ &= \frac{1}{2\pi n}\sum_{\kappa=-n+1}^{n-1} R(\kappa)(n-|\kappa|)\int_{-\pi}^{\pi} e^{-i\kappa\lambda}d\lambda = R(0),\end{aligned}$$

since the last integral is $2\pi\delta(\kappa)$.

The sequence $\{F_n,\ n \in \mathbb{N}\}$ is a sequence of non-decreasing functions, $0 \le F_n(x) \le R(0) < \infty$ 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 non-decreasing bounded right-continuous function and $F(x) = 0,\ x \le -\pi,\ F(x) = R(0),\ x > \pi$.

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

$$\int_{-\pi}^{\pi} e^{it\lambda}dF_{n_k}(\lambda) \longrightarrow \int_{-\pi}^{\pi} e^{it\lambda}dF(\lambda) \quad \text{as } k \to \infty,\ n_k \to \infty.$$

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

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

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

To prove the uniqueness, suppose that $R(t) = \int_{-\pi}^{\pi} e^{it\lambda} dG(\lambda)$, 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 μ_F a μ_G are finite measures on Borel subsets of the interval $[-\pi, \pi]$ induced by functions F a G, respectively. The rest of the proof follows from Lemma 1 since $\mu_F(B) = \mu_G(B)$ for any $B \subset (-\pi, \pi)$ and $\mu_F(\{-\pi\} \cup \{\pi\}) = \mu_G(\{-\pi\} \cup \{\pi\})$. □

Formula (2) is called *the spectral decomposition (representation) of an autocovariance function* of a stationary random sequence. The function F is called *the spectral distribution function* of a stationary 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 *the spectral density.* Obviously $f = F'$.

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

$$R(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{Z}. \tag{3}$$

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*

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda), \quad t \in \mathbb{R}, \tag{4}$$

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

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

Function F is determined uniquely.

Proof. 1. Let R be a complex-valued function on $\mathbb{R}$ that satisfies (4), where F is a 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 7, 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. □

Function F from Theorem 20 is called the spectral distribution function of a stationary mean square continuous stochastic process. If the spectral distribution function in (4) is absolutely continuous, its derivative f is again called the spectral density and (4) can be written in the form

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{R}. \tag{5}$$

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 argument $t \in \mathbb{Z}$, 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 K be such that $\sum_{t=-\infty}^{\infty} |K(t)| < \infty$ and $f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-it\lambda} K(t)$. Since the series $\sum_{t=-\infty}^{\infty} e^{-it\lambda} K(t)$ converges absolutely and uniformly for $\lambda \in [-\pi, \pi]$, we can interchange the integration and the summation and for any $t \in \mathbb{Z}$ we get

$$\begin{aligned} \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda &= \int_{-\pi}^{\pi} e^{it\lambda} \left[\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} K(k) \right] d\lambda \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[K(k) \int_{-\pi}^{\pi} e^{i(t-k)\lambda} d\lambda \right] \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} K(k) 2\pi \delta(t-k) = K(t). \end{aligned}$$

□

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. Since $\sum_{t=-\infty}^{\infty} |R(t)| < \infty$ it follows from the previous theorem that

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

where

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

To prove that f is the spectral density, due to the uniqueness of the spectral decomposition (3), it suffices to prove that $f(\lambda) \geq 0$ for every $\lambda \in [-\pi, \pi]$.

We know from the proof of Theorem 19 that for every $\lambda \in [-\pi, \pi]$,

$$\varphi_n(\lambda) = \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda} R(\kappa)(n - |\kappa|) \geq 0.$$

We will show that $f(\lambda) = \lim_{n\to\infty} \varphi_n(\lambda)$.

We have, as $n \to \infty$,

$$
\begin{aligned}
|f(\lambda) - \varphi_n(\lambda)| &\leq \left| \frac{1}{2\pi} \sum_{|k| \geq n} e^{-ik\lambda} R(k) \right| \\
&+ \left| \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} e^{-i\kappa\lambda} R(\kappa)|\kappa| \right| \\
&\leq \frac{1}{2\pi} \sum_{|k| \geq n} |R(k)| + \frac{1}{2\pi n} \sum_{\kappa=-n+1}^{n-1} |R(\kappa)||\kappa| \longrightarrow 0
\end{aligned}
$$

where we have used the assumption on the absolute summability of the autocovariance function and the Kronecker lemma.[1] □

Formula (6) is called *the 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 the spectral density of the process exists and it holds*

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

The proof is quite analogous to the computation of a probability density function by using the inverse Fourier transformation of the characteristic function (see, e. g. Štěpán, 1987, IV.5.3.)

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 σ^2. The autocovariance function is $\operatorname{cov}(X_s, X_t) = \sigma^2 \delta(s-t) = R(s-t)$, the sequence is weakly stationary, and since $\sum_{t=-\infty}^{\infty} |R(t)| = \sigma^2 < \infty$ the spectral density exists and according to inverse formula (6)

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

The spectral distribution function of the white noise sequence is

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

Notation: $WN(0, \sigma^2)$ (white noise)

[1] $\sum_{k=1}^{\infty} a_k < \infty \Rightarrow \frac{1}{n} \sum_{k=1}^{n} k\, a_k \to 0$ as $n \to \infty$.

Example 17. Consider a stationary sequence with the autocovariance function $R(t) = a^{|t|}$, $t \in \mathbb{Z}$, $|a| < 1$. Since

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

the spectral density exists and according to inverse formula (6)

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} a^{|k|} \\ &= \frac{1}{2\pi} \sum_{k=0}^{\infty} e^{-ik\lambda} a^k + \frac{1}{2\pi} \sum_{k=-\infty}^{-1} e^{-ik\lambda} a^{-k} \\ &= \frac{1}{2\pi} \sum_{k=0}^{\infty} \left(ae^{-i\lambda}\right)^k + \frac{1}{2\pi} \sum_{k=1}^{\infty} \left(ae^{i\lambda}\right)^k \\ &= \frac{1}{2\pi} \frac{1}{1-ae^{-i\lambda}} + \frac{1}{2\pi} \frac{ae^{i\lambda}}{1-ae^{i\lambda}} \\ &= \frac{1}{2\pi} \frac{1-a^2}{|1-ae^{-i\lambda}|^2} = \frac{1}{2\pi} \frac{1-a^2}{1-2a\cos\lambda + a^2}. \end{aligned}$$

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

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} R(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} ce^{-\alpha|t|} dt \\ &= \frac{c}{2\pi} \int_{-\infty}^{\infty} (\cos\lambda t - i\sin\lambda t) e^{-\alpha|t|} dt \\ &= \frac{c}{\pi} \int_{0}^{\infty} \cos(\lambda t) e^{-\alpha t} dt = \frac{c\alpha}{\pi} \frac{1}{\alpha^2 + \lambda^2} \end{aligned}$$

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

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

$$\begin{aligned} F(\lambda) &= 0, & \lambda < -1, \\ &= \frac{1}{2}, & -1 \le \lambda < 1, \\ &= 1, & \lambda \ge 1. \end{aligned}$$

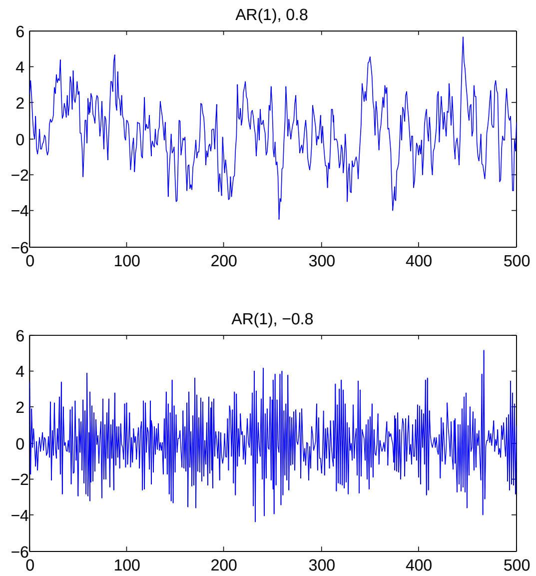

Figure 3: Trajectories of a process with the autocovariance function $R(t) = a^{|t|}$, up: $a = 0,8$, down $a = -0,8$

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 values 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 $(t_1, t_2] \cap (t_3, t_4] = \emptyset$,

$$\mathsf{E}(X_{t_2} - X_{t_1})(\overline{X}_{t_4} - \overline{X}_{t_3}) = 0.$$

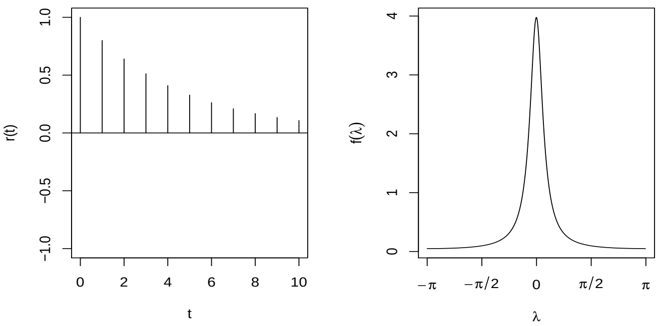

Figure 4: Autocovariance function $R(t) = a^{|t|}$ (left) and the spectral density (right), $a = 0,8$

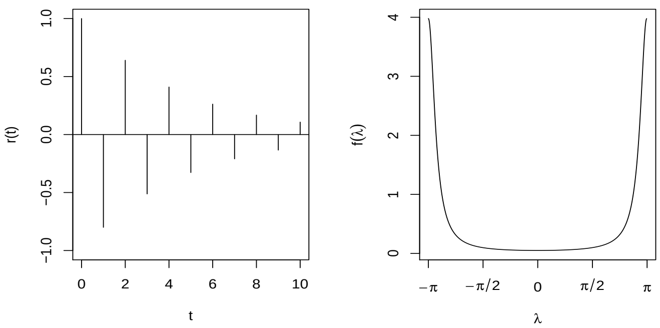

Figure 5: Autocovariance function $R(t) = a^{|t|}$ (left) and the spectral density (right), $a = -0,8$

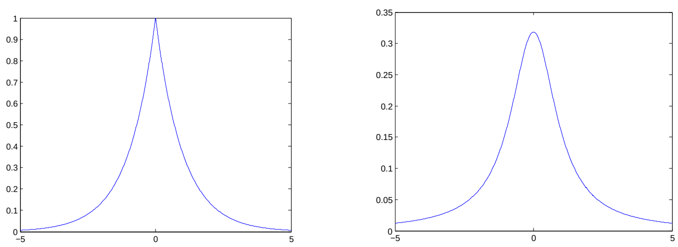

Figure 6: Autocovariance function $R(t) = ce^{-\alpha|t|}, t \in \mathbb{R}$ (left) and the spectral density (right), $c = 1, \alpha = 1$

We also say that the increments of the process are orthogonal random variables.

In what follows we will consider only centered right-mean square continuous processes i.e., such that $\mathsf{E}|X_t - X_{t_0}|^2 \to 0$ 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*

$$\begin{aligned} F(\lambda) &= 0, & \lambda &\le a, \\ &= F(b), & \lambda &\ge b, \\ F(\lambda_2) - F(\lambda_1) &= \mathsf{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2, & a &\le \lambda_1 < \lambda_2 \le b. \end{aligned} \tag{8}$$

Proof. Define function

$$\begin{aligned} F(\lambda) &= \mathsf{E}|Z_\lambda - Z_a|^2, \quad \lambda \in [a,b] \\ &= 0, \quad \lambda \le a, \\ &= F(b), \quad \lambda \ge b. \end{aligned}$$

We will 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

$$\begin{aligned} F(\lambda_2) &= \mathsf{E}|Z_{\lambda_2} - Z_a|^2 = \mathsf{E}|Z_{\lambda_2} - Z_{\lambda_1} + Z_{\lambda_1} - Z_a|^2 \\ &= \mathsf{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 + \mathsf{E}|Z_{\lambda_1} - Z_a|^2 \\ &\quad + \mathsf{E}(Z_{\lambda_2} - Z_{\lambda_1})(\overline{Z_{\lambda_1}} - \overline{Z_a}) + \mathsf{E}(Z_{\lambda_1} - Z_a)(\overline{Z_{\lambda_2}} - \overline{Z_{\lambda_1}}) \\ &= \mathsf{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 + F(\lambda_1) \end{aligned}$$

since the increments $Z_{\lambda_2} - Z_{\lambda_1}$ a $Z_{\lambda_1} - Z_a$ are orthogonal. From here we have

$$F(\lambda_2) - F(\lambda_1) = \mathsf{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 \ge 0,$$

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

Now, 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 $G(\lambda) = G(\lambda) - G(a) = \mathsf{E}|Z_\lambda - Z_a|^2 = F(\lambda) - F(a) = F(\lambda)$, which proves the uniqueness of function F. □

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 na $[0, T]$ is a centered mean square continuous Gaussian process with independent and stationary increments, therefore with orthogonal increments, such that $W_0 = 0$, $W_s - W_t \sim \mathcal{N}(0, \sigma^2|s-t|)$, $0 \le s, t, \le T$. The associate distribution function on $[0, T]$ is

$$\begin{aligned} F(\lambda) &= 0, \qquad \lambda \le 0, \\ &= \mathsf{E}|W_\lambda - W_0|^2 = \sigma^2\lambda, \quad 0 \le \lambda \le T, \\ &= \sigma^2 T, \quad \lambda \ge T. \end{aligned}$$

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

The process $\{\widetilde{W_\lambda},\ \lambda \in [\pi, \pi]\}$ is Gaussian process with orthogonal increments and the associated distribution function

$$\begin{aligned} F(\lambda) &= 0, & \lambda &\le -\pi, \\ &= \frac{\sigma^2}{2\pi}(\lambda + \pi), & \lambda &\in [-\pi, \pi], \\ &= \sigma^2, & \lambda &\ge \pi \end{aligned}$$

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 μ_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

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

Recall the basic properties of this space.

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 μ_F) [2] is defined by

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

[2] $f \sim g$ if $f = g$ μ_F-almost everywhere

- Norm in $L_2(F)$ is given by $\|f\| = \left[\int_a^b |f(\lambda)|^2 dF(\lambda)\right]^{\frac{1}{2}}$;
- Convergence in $L_2(F)$ means that

  $f_n \to f$ in $L_2(F)$ as $n \to \infty$, if $\|f_n - f\| \to 0$, i. e.,

$$\int_a^b |f_n(\lambda) - f(\lambda)|^2 dF(\lambda) \to 0, \qquad n \to \infty;$$

- 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., . for $a = \lambda_0 < \lambda_1 < \cdots < \lambda_n = b$

$$f(\lambda) = \sum_{k=1}^{n} c_k J_{(\lambda_{k-1},\ \lambda_k]}(\lambda), \tag{9}$$

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

$$\int_{(a,b]} f(\lambda) dZ(\lambda) := \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}), \tag{10}$$

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 mean square right-continuous process with orthogonal increments and the associated distribution function F, let f, g be simple functions in $L_2(F)$, α, β complex-valued constants. Then*

1. $\mathsf{E} \int_a^b f(\lambda)dZ(\lambda) = 0$.
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. $\mathsf{E} \int_a^b f(\lambda)dZ(\lambda)\overline{\int_a^b g(\lambda)dZ(\lambda)} = \int_a^b f(\lambda)\overline{g(\lambda)}dF(\lambda)$.

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

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

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

2. W. l. o. g, let

$$f(\lambda) = \sum_{k=1}^{n} c_k J_{(\lambda_{k-1},\, \lambda_k]}(\lambda), \quad g(\lambda) = \sum_{k=1}^{n} d_k J_{(\lambda_{k-1},\, \lambda_k]}(\lambda).$$

Then

$$\begin{aligned} &\int_a^b [\alpha f(\lambda) + \beta g(\lambda)] dZ(\lambda) = \sum_{k=1}^{n} (\alpha c_k + \beta d_k)(Z_{\lambda_k} - Z_{\lambda_{k-1}}) \\ &= \alpha \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) + \beta \sum_{k=1}^{n} d_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) \\ &= \alpha \int_a^b f(\lambda) dZ(\lambda) + \beta \int_a^b g(\lambda) dZ(\lambda). \end{aligned}$$

3. Let

$$f(\lambda) = \sum_{k=1}^{n} c_k J_{(\lambda_{k-1},\, \lambda_k]}(\lambda), \quad g(\lambda) = \sum_{k=1}^{n} d_k J_{(\lambda_{k-1},\, \lambda_k]}(\lambda).$$

Then

$$\begin{aligned} &\mathsf{E} \int_a^b f(\lambda) dZ(\lambda) \overline{\int_a^b g(\lambda) dZ(\lambda)} \\ &= \mathsf{E} \sum_{k=1}^{n} c_k \left(Z_{\lambda_k} - Z_{\lambda_{k-1}} \right) \overline{\sum_{k=1}^{n} d_k \left(Z_{\lambda_k} - Z_{\lambda_{k-1}} \right)} \\ &= \sum_{k=1}^{n} c_k \overline{d_k} \mathsf{E} \left| Z_{\lambda_k} - Z_{\lambda_{k-1}} \right|^2 \\ &= \sum_{k=1}^{n} c_k \overline{d_k} \left(F(\lambda_k) - F(\lambda_{k-1}) \right) = \int_a^b f(\lambda) \overline{g(\lambda)} dF(\lambda). \end{aligned}$$

□

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), it means that 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 $L_2(\Omega, \mathcal{A}, P)$ because

$$
\begin{aligned}
\mathsf{E}|I(f_m) - I(f_n)|^2 &= \mathsf{E}(I(f_m) - I(f_n))\overline{(I(f_m) - I(f_n))} \\
&= \mathsf{E}\int_a^b (f_m(\lambda) - f_n(\lambda))dZ(\lambda)\overline{\int_a^b (f_m(\lambda) - f_n(\lambda))dZ(\lambda)} \\
&= \int_a^b (f_m(\lambda) - f_n(\lambda))\overline{(f_m(\lambda) - f_n(\lambda))}dF(\lambda) \\
&= \int_a^b |f_m(\lambda) - f_n(\lambda)|^2 dF(\lambda) \longrightarrow 0
\end{aligned}
$$

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

$$
I(f) = \operatorname*{l.\,i.\,m.}_{n\to\infty} I(f_n) := \int_a^b f(\lambda)dZ(\lambda), \tag{11}
$$

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

Notice that $I(f)$ 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 $g_n \to f$ v $L_2(F)$. Then $I(f_n), I(g_n)$ have mean square limits I, J, respectively.

Define sequence $\{h_n\} = \{f_1, g_1, f_2, g_2, \dots\}$ which is simple and $h_n \to f$ in $L_2(F)$. Then $I(h_n) \to K$ in mean square. Since selected subsequences $\{I(f_n)\}$ a $\{I(g_n)\}$ 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 $\mathsf{E}I(f) = \mathsf{E}\int_a^b f(\lambda)dZ(\lambda) = 0$.*
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*

$$
\mathsf{E}I(f)\overline{I(g)} = \int_a^b f(\lambda)\overline{g(\lambda)}dF(\lambda). \tag{12}
$$

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

2. Let $f, g \in L_2(F)$ and let $\{f_n,\, n \in \mathbb{N}\}$, 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

$$\begin{aligned}
&\int_a^b |\alpha f_n(\lambda) + \beta g_n(\lambda) - (\alpha f(\lambda) + \beta g(\lambda))|^2 dF(\lambda) \\
&\le 2|\alpha|^2 \int_a^b |f_n(\lambda) - f(\lambda)|^2 dF(\lambda) \\
&+ 2|\beta|^2 \int_a^b |g_n(\lambda) - g(\lambda)|^2 dF(\lambda) \to 0
\end{aligned}$$

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

$$\begin{aligned}
&\mathsf{E}|\alpha I(f_n) + \beta I(g_n) - (\alpha I(f) + \beta I(g))|^2 \\
&= \mathsf{E}|\alpha(I(f_n) - I(f)) + \beta(I(g_n) - I(g))|^2 \\
&\le 2|\alpha|^2 \mathsf{E}|I(f_n) - I(f)|^2 + 2|\beta|^2 \mathsf{E}|I(g_n) - I(g)|^2 \longrightarrow 0
\end{aligned}$$

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

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 $L_2(\Omega, \mathcal{A}, P)$ we have

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

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

$$\begin{aligned}\mathsf{E}I(f_n)\overline{I(g_n)} &= \int_a^b f_n(\lambda)\overline{g_n(\lambda)}dF(\lambda) = \langle f_n, g_n\rangle \to \langle f, g\rangle \\ &= \int_a^b f(\lambda)\overline{g(\lambda)}dF(\lambda),\end{aligned}$$

from here (12) follows.

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

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

from which (13) follows. □

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

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

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

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

$$\int_{-\infty}^{\infty} f(\lambda)dZ(\lambda) := \mathrm{l.i.m.} \int_a^b f(\lambda)dZ(\lambda) \text{ as } a \to -\infty,\ b \to \infty.$$

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-mean square 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 to be

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

Then

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

it means that $e_t \in L_2(F)$ and $X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)$ is well defined random variable.

According to Theorem 26 we have

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

2. $$\begin{aligned} \mathsf{E}|X_t|^2 &= \mathsf{E}\left|\int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)\right|^2 = \mathsf{E}\int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)\overline{\int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)} \\ &= \int_{-\pi}^{\pi} |e^{it\lambda}|^2 dF(\lambda) = \int_{-\pi}^{\pi} dF(\lambda) < \infty. \end{aligned}$$

3. $$\begin{aligned} \operatorname{cov}(X_{t+h}, X_t) &= \mathsf{E}\int_{-\pi}^{\pi} e^{i(t+h)\lambda} dZ(\lambda)\overline{\int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)} \\ &= \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda) := R(h). \end{aligned}$$

From here we can conclude that

- $\operatorname{cov}(X_{t+h}, X_t) = R(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda)$ 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, Theorem 19)
- from the uniqueness of the spectral decomposition (2) it follows that F is the spectral distribution function of the sequence $\{X_t,\ t \in \mathbb{Z}\}$.

□

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

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} d\widetilde{W}(\lambda), \ t \in \mathbb{Z}$$

are centered, uncorrelated with the same variance σ^2; the sequence $\{X_t, t \in \mathbb{Z}\}$ is the Gaussian white noise, see also Examples 16 and 22.

Example 24. Consider a sequence of functions $\{f_{tk}, \ t \in \mathbb{Z}\}$ on $[-\pi, \pi]$ defined by

$$f_{tk}(\lambda) = \sum_{j=1}^{k} e^{it\lambda_j} J_{(\lambda_{j-1}, \lambda_j]}(\lambda),$$

where $-\pi = \lambda_0 < \lambda_1 < \cdots < \lambda_k = \pi$ and $k \in \mathbb{N}$ is given. Let $\{Z_\lambda, \ \lambda \in [-\pi, \pi]\}$ be a centered right-mean square continuous process with orthogonal increments on $[-\pi, \pi]$ with the associated distribution function F. We can see that f_{tk} are simple functions in $L_2(F)$, thus we can compute

$$X_{tk} := \int_{-\pi}^{\pi} f_{tk}(\lambda) dZ(\lambda) = \sum_{j=1}^{k} e^{it\lambda_j} (Z_{\lambda_j} - Z_{\lambda_{j-1}}) = \sum_{j=1}^{k} e^{it\lambda_j} \widetilde{Z}_j.$$

Here $\widetilde{Z}_j = Z_{\lambda_j} - Z_{\lambda_{j-1}}, \ j = 1, \ldots, k$, are uncorrelated random variables with zero mean and the variance $\mathsf{E}|\widetilde{Z}_j|^2 = \mathsf{E}|Z_{\lambda_j} - Z_{\lambda_{j-1}}|^2 = F(\lambda_j) - F(\lambda_{j-1}) := \sigma_j^2$. Then we have

$$\mathsf{E}X_{tk} = 0, \ \mathsf{E}X_{(t+h),k}\overline{X_{tk}} = \sum_{j=1}^{k} e^{ih\lambda_j} \sigma_j^2 := R_k(h).$$

We see that $\{X_{tk}, t \in \mathbb{Z}\}$ is stationary and its autocovariance function has the spectral decomposition

$$R_k(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF_{Xk}(\lambda)$$

where F_{Xk} is the spectral distribution function of $\{X_{tk}, t \in \mathbb{Z}\}$; it has jumps at points λ_j such that $F_{Xk}(\lambda_j) - F_{Xk}(\lambda_{j-1}) = \sigma_j^2$. On the other hand, $\sigma_j^2 = F(\lambda_j) - F(\lambda_{j-1})$. Since $F(-\pi) = F_{Xk}(-\pi) = 0$, F equals to F_{Xk} at least at points $\lambda_j, \ j = 0, 1, \ldots, k$.

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*

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda), \quad t \in \mathbb{Z} \tag{14}$$

and

$$\mathsf{E}|Z(\lambda) - Z(-\pi)|^2 = F(\lambda), \quad -\pi \leq \lambda \leq \pi.$$

Proof. Brockwell and Davis (1991), Theorem 4.8.2 or Prášková (2016), Theorem 4.4. □

Relation (14) is called *the spectral decomposition of a stationary random sequence.*

Remark 7. Theorem 28 says that 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) λ_{j-1} and λ_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*

$$X_t = \int_{-\infty}^{\infty} e^{it\lambda} dZ(\lambda),\ \ t\in\mathbb{R}, \tag{15}$$

and the associated distribution function of the process $\{Z_\lambda,\, \lambda\in\mathbb{R}\}$ is the spectral distribution function of the process $\{X_t,\, t\in\mathbb{R}\}$.

Proof. Priestley (1981), Chap. 4.11 □

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*

$$U = \int_{-\pi}^{\pi} \varphi(\lambda) dZ(\lambda), \tag{16}$$

where $\varphi\in L_2(F)$ and $\{Z_\lambda,\, \lambda\in[-\pi,\pi]\}$ is the orthogonal increment 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 $U = \text{l.i.m.}_{n\to\infty} U_n,\ U_n\in\mathcal{M}\{X_t,\, t\in\mathbb{Z}\}$.
a) Let $U\in\mathcal{M}\{X_t,\, t\in\mathbb{Z}\}$; then $U=\sum_{j=1}^N c_j X_{t_j}$, for $c_1,\dots,c_N\in\mathbb{C}$ and $t_1,\dots,t_N\in\mathbb{Z}$. From the spectral decomposition (14)

$$\begin{aligned} U &= \sum_{j=1}^N c_j X_{t_j} = \sum_{j=1}^N c_j \left[\int_{-\pi}^{\pi} e^{it_j\lambda} dZ(\lambda)\right] \\ &= \int_{-\pi}^{\pi} \left[\sum_{j=1}^N c_j e^{it_j\lambda}\right] dZ(\lambda) = \int_{-\pi}^{\pi} \varphi(\lambda) dZ(\lambda), \end{aligned}$$

where $\varphi(\lambda) = \sum_{j=1}^{N} c_j e^{it_j\lambda}$. Obviously, φ is a finite linear combination of functions from $L_2(F)$, thus $\varphi \in L_2(F)$.
b) Let $U = \text{l.i.m.}_{n\to\infty} U_n,\ U_n \in \mathcal{M}\{X_t,\, t \in \mathbb{Z}\}$. According to a)

$$U_n = \int_{-\pi}^{\pi} \varphi_n(\lambda) dZ(\lambda), \quad \varphi_n \in L_2(F).$$

Here, $\{U_n\}$ 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

$$\mathsf{E}|U_m - U_n|^2 = \mathsf{E}\left|\int_{-\pi}^{\pi} \varphi_m(\lambda)dZ(\lambda) - \int_{-\pi}^{\pi} \varphi_n(\lambda)dZ(\lambda)\right|^2 =$$

$$\mathsf{E}\left|\int_{-\pi}^{\pi} [\varphi_m(\lambda) - \varphi_n(\lambda)]dZ(\lambda)\right|^2 = \int_{-\pi}^{\pi} |\varphi_m(\lambda) - \varphi_n(\lambda)|^2 dF(\lambda).$$

Hence, there exists $\varphi \in L_2(F)$ such that $\varphi_n \to \varphi$ v $L_2(F)$. By (13)

$$U_n = \int_{-\pi}^{\pi} \varphi_n(\lambda)dZ(\lambda) \to \int_{-\pi}^{\pi} \varphi(\lambda)dZ(\lambda) \ \text{ v } L_2(\Omega, \mathcal{A}, P),$$

thus $U = \int_{-\pi}^{\pi} \varphi(\lambda)dZ(\lambda)$.
2. Let U be a random variable that satisfies (16). Since $\varphi \in L_2(F)$, there exists a sequence of trigonometric polynomials $\varphi_n(\lambda) = \sum_{k=-n}^{n} c_k^{(n)} e^{i\lambda t_k^{(n)}}$ on $[-\pi, \pi]$ such that $\varphi_n \to \varphi$ v $L_2(F)$. According to (13)

$$\int_{-\pi}^{\pi} \varphi(\lambda)dZ(\lambda) = \underset{n\to\infty}{\text{l. i. m.}} \int_{-\pi}^{\pi} \varphi_n(\lambda)dZ(\lambda),$$

hence

$$\begin{aligned} U &= \int_{-\pi}^{\pi} \varphi(\lambda)dZ(\lambda) = \underset{n\to\infty}{\text{l. i. m.}} \int_{-\pi}^{\pi} \left[\sum_{k=-n}^{n} c_k^{(n)} e^{i\lambda t_k^{(n)}}\right] dZ(\lambda) \\ &= \underset{n\to\infty}{\text{l. i. m.}} \sum_{k=-n}^{n} c_k^{(n)} \left[\int_{-\pi}^{\pi} e^{i\lambda t_k^{(n)}} dZ(\lambda)\right] \\ &= \underset{n\to\infty}{\text{l. i. m.}} \sum_{k=-n}^{n} c_k^{(n)} X_{t_k^{(n)}} \in \mathcal{H}\{X_t,\, t \in \mathbb{Z}\}. \end{aligned}$$

□

8 Linear models of time series

8.1 White noise

Recall that the white noise sequence $\mathrm{WN}(0, \sigma^2)$ is defined as a sequence $\{Y_t,\, t \in \mathbb{Z}\}$ of uncorrelated random variables with mean zero and variance $0 < \sigma^2 < \infty$, the autocovariance function

$$R_Y(t) = \sigma^2 \delta(t), \quad t \in \mathbb{Z}$$

and the spectral density

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

Moreover,

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

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

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

that is same as the spectral distribution function $F_Y(\lambda)$ of $\{Y_t,\, t \in \mathbb{Z}\}$.

8.2 Moving average sequences

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

$$X_t = b_0 Y_t + b_1 Y_{t-1} + \cdots + b_n Y_{t-n}, \quad t \in \mathbb{Z}, \tag{17}$$

where $\{Y_t,\, t \in \mathbb{Z}\}$ is a white noise $\mathrm{WN}(0, \sigma^2)$ 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)

Remark 8. In special case, $b_i = \frac{1}{n+1}, i = 0, \ldots, n.$

The basic properties of MA(n) :

1. $\mathsf{E}X_t = 0$ for all $t \in \mathbb{Z}$.

2. The autocovariance function (computation in *time domain*): For $t \geq 0$,

$$\begin{aligned}
\operatorname{cov}(X_{s+t}, X_s) = \mathsf{E} X_{s+t}\overline{X_s} &= \mathsf{E}\left(\sum_{j=0}^{n} b_j Y_{s+t-j} \sum_{k=0}^{n} \overline{b_k Y_{s-k}}\right) \\
&= \sum_{j=0}^{n}\sum_{k=0}^{n} b_j \overline{b_k} \mathsf{E}(Y_{s+t-j}\overline{Y_{s-k}}) \\
&= \sigma^2 \sum_{j=0}^{n}\sum_{k=0}^{n} b_j \overline{b_k} \delta(t-j+k) \\
&= \sigma^2 \sum_{k=0}^{n-t} b_{t+k}\overline{b_k}, \quad 0 \leq t \leq n \\
&= 0, \quad t > n.
\end{aligned}$$

For $t \leq 0$ we proceed analogously. Since $\operatorname{cov}(X_{s+t}, X_s)$ depends on t, only, we can conclude that the sequence is weakly stationary.

3. Spectral decomposition. By using the spectral decomposition of the white noise we obtain

$$\begin{aligned}
X_t = \sum_{j=0}^{n} b_j Y_{t-j} &= \sum_{j=0}^{n} b_j \left[\int_{-\pi}^{\pi} e^{i(t-j)\lambda} dZ_Y(\lambda)\right] \\
&= \int_{-\pi}^{\pi} \left[\sum_{j=0}^{n} b_j e^{i(t-j)\lambda}\right] dZ_Y(\lambda) \\
&= \int_{-\pi}^{\pi} e^{it\lambda} \left[\sum_{j=0}^{n} b_j e^{-ij\lambda}\right] dZ_Y(\lambda) \\
&= \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ_Y(\lambda),
\end{aligned}$$

where $g(\lambda) = \sum_{j=0}^{n} b_j e^{-ij\lambda} \in L_2(F)$. From the properties of the stochastic integral we again get $\mathsf{E}X_t = 0$, and for the autocovariance function (in *spectral domain*) we have

$$\begin{aligned}
\mathsf{E} X_{s+t}\overline{X_s} &= \mathsf{E} \int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) dZ_Y(\lambda) \overline{\int_{-\pi}^{\pi} e^{is\lambda} g(\lambda) dZ_Y(\lambda)} \\
&= \int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) e^{-is\lambda} \overline{g(\lambda)} dF_Y(\lambda) \\
&= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 f_Y(\lambda) d\lambda \\
&= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 \frac{\sigma^2}{2\pi} d\lambda = R_X(t),
\end{aligned}$$

that is again a function of t which confirms the weak stationarity. Due to the uniqueness of the spectral decomposition of the autocovariance function (Theorem 19) we can conclude that function $|g(\lambda)|^2\frac{\sigma^2}{2\pi}$ is the spectral density of the sequence $\{X_t,\, t\in\mathbb{Z}\}$.

We have just 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*

$$\begin{aligned} R_X(t) &= \sigma^2\sum_{k=0}^{n-t} b_{k+t}\overline{b_k}, && 0\le t\le n, \\ &= \overline{R_X(-t)}, && -n\le t\le 0, \\ &= 0, && |t|>n. \end{aligned} \tag{18}$$

The spectral density f_X of sequence (17) exists and is given by

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

Remark 9. For real-valued constants $b_0,\dots,b_n$ the autocovariance function of the sequence MA(n) takes form

$$\begin{aligned} R_X(t) &= \sigma^2\sum_{k=0}^{n-|t|} b_k b_{k+|t|}, && |t|\le n, \\ &= 0, && |t|>n. \end{aligned} \tag{20}$$

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 $\sum_{j=0}^\infty c_jY_{t-j}$ 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

$$X_t = \underset{n\to\infty}{\text{l. i. m.}} \sum_{j=0}^{n} c_jY_{t-j}.$$

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

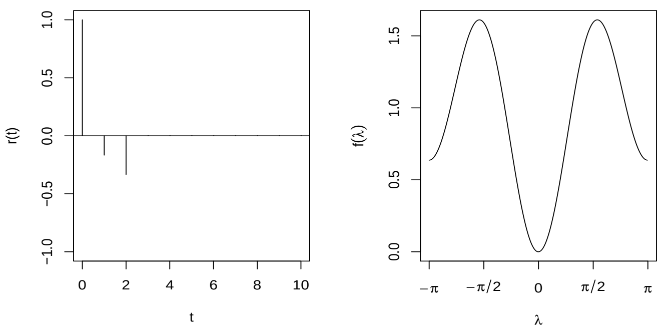

Figure 7: Autocorrelation function (left) and spectral density (right) of the MA(2) sequence $X_t = Y_t + Y_{t-1} - 2\,Y_{t-2}$. Y_t : Gaussian white noise $\sim N(0,1)$

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

W. l. o. g., assume that $m < n$. Since Y_k are uncorrelated with a constant variance σ^2 we easily get

$$\mathsf{E}\left|\sum_{j=0}^{m} c_j Y_{t-j} - \sum_{k=0}^{n} c_k Y_{t-k}\right|^2 = \mathsf{E}\left|\sum_{j=m+1}^{n} c_j Y_{t-j}\right|^2$$
$$= \sum_{j=m+1}^{n} |c_j|^2 \mathsf{E}|Y_{t-j}|^2 = \sigma^2 \sum_{j=m+1}^{n} |c_j|^2 \to 0$$

as $m, n \to \infty$ which means that there exists a mean square limit of the sequence $\{\sum_{j=0}^{n} c_j Y_{t-j}\}$, that we denote by $\sum_{j=0}^{\infty} c_j Y_{t-j}$.

2. Since $\mathsf{E}|Y_{t-j}| \le (\mathsf{E}|Y_{t-j}|^2)^{\frac{1}{2}} = \sqrt{\sigma^2} < \infty$, we can see that

$$\sum_{j=0}^{\infty} \mathsf{E}|c_j Y_{t-j}| = \sum_{j=0}^{\infty} |c_j| \mathsf{E}|Y_{t-j}| \le \sigma \sum_{j=0}^{\infty} |c_j| < \infty,$$

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

$$\mathsf{E}\left|\sum_{j=m+1}^{n} c_j X_{t-j}\right|^2 \le \mathsf{E}\left(\sum_{j=m+1}^{n} |c_j||X_{t-j}|\right)^2$$
$$= \sum_{j=m+1}^{n}\sum_{k=m+1}^{n} |c_j||c_k|\mathsf{E}|X_{t-j}||X_{t-k}|.$$

The weak stationarity and the Schwarz inequality imply

$$\mathsf{E}|X_{t-j}||X_{t-k}| \le (\mathsf{E}|X_{t-j}|^2)^{\frac{1}{2}}(\mathsf{E}|X_{t-k}|^2)^{\frac{1}{2}} = R(0),$$

thus,

$$\mathsf{E}\left|\sum_{j=m+1}^{n} c_j X_{t-j}\right|^2 \le R(0)\left(\sum_{j=m+1}^{n} |c_j|\right)^2 \to 0$$

as $m, n \to \infty$. We have proved the mean square convergence.

2. From the weak stationarity we also get

$$\sum_{j=0}^{\infty} \mathsf{E}|c_j X_{t-j}| = \sum_{j=0}^{\infty} |c_j|\mathsf{E}|X_{t-j}| \le \sqrt{R(0)}\sum_{j=0}^{\infty} |c_j| < \infty,$$

from which the rest of the proof follows. □

Definition 32. Let $\{Y_t,\ t \in \mathbb{Z}\}$ be a white noise $\mathrm{WN}(0, \sigma^2)$ 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

$$X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z} \tag{21}$$

is called *causal linear process.*

Remark 10. The causality means that the random variable X_t depends on $Y_s, s \le t$ (contemporary and past variables, only). Sometimes we also use notation $\mathrm{MA}(\infty)$.

Remark 11. A weaker condition $\sum_{j=0}^{\infty} |c_j|^2 < \infty$ implies the mean square convergence in the series defined by (21), only.

Theorem 34. *The 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*

$$\begin{aligned} R_X(t) &= \sigma^2 \sum_{k=0}^{\infty} c_{k+t}\overline{c_k}, && t \ge 0, \\ &= \overline{R_X(-t)}, && t \le 0. \end{aligned} \tag{22}$$

The spectral density f_X of sequence (21) exists and takes values

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

Proof. Notice that

- $\{X_t^{(n)}, t \in \mathbb{Z}\}$ where $X_t^{(n)} := \sum_{j=0}^{n} c_j Y_{t-j}$ is 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). It means that $\{X_t,\ t \in \mathbb{Z}\}$ is centered, since it is the mean square limit of the centered sequence.
- 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) and the stationarity of sequence (21). Further, notice that the sequence $\{X_t^{(n)},\ t \in \mathbb{Z}\}$ has the spectral decomposition

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

If we denote $g(\lambda) = \sum_{j=0}^{\infty} c_j e^{-ij\lambda}$, we have

$$\int_{-\pi}^{\pi} |g_n(\lambda) - g(\lambda)|^2 dF_Y(\lambda) = \int_{-\pi}^{\pi} \left|\sum_{j=n+1}^{\infty} c_j e^{-ij\lambda}\right|^2 dF_Y(\lambda)$$
$$\leq \int_{-\pi}^{\pi} \left(\sum_{j=n+1}^{\infty} |c_j|\right)^2 f_Y(\lambda) d\lambda = \sigma^2 \left(\sum_{j=n+1}^{\infty} |c_j|\right)^2 \to 0.$$

Thus, $g_n \to g$ in $L_2(F_Y)$. According to Theorem 26,

$$X_t^{(n)} = \int_{-\pi}^{\pi} e^{it\lambda} g_n(\lambda) dZ_Y(\lambda) \longrightarrow \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ_Y(\lambda)$$

in mean square as $n \to \infty$ and simultaneously, $X_t^{(n)} \to X_t$ in mean square, which means that

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

Further, the computation

$$\mathsf{E}X_{s+t}\overline{X_s} = \mathsf{E}\int_{-\pi}^{\pi} e^{i(s+t)\lambda} g(\lambda) dZ_Y(\lambda) \overline{\int_{-\pi}^{\pi} e^{is\lambda} g(\lambda) dZ_Y(\lambda)}$$
$$= \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 dF_Y(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} |g(\lambda)|^2 \frac{\sigma^2}{2\pi} d\lambda$$

results in 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).

□

Example 25. Let us consider a causal linear process such that

$$X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z}, \quad c_j = \varphi^j, \ |\varphi| < 1$$

and $Y_t \sim WN(0, \sigma^2$.

The process is centered, weakly stationary, with the autocovariance function

$$\begin{aligned} R_X(t) &= \sigma^2 \frac{\varphi^t}{1-\varphi^2}, & t \geq 0, \\ &= R_X(-t), & t \leq 0. \end{aligned}$$

The spectral density is

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \varphi^j e^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\lambda}|^2}, \qquad \lambda \in [-\pi, \ \pi].$$

Further, we can write

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} \varphi^j Y_{t-j} = Y_t + \sum_{j=1}^{\infty} \varphi^j Y_{t-j} = Y_t + \varphi \sum_{j=1}^{\infty} \varphi^{j-1} Y_{t-j} \\ &= Y_t + \varphi \sum_{k=0}^{\infty} \varphi^k Y_{t-1-k} \\ &= \varphi X_{t-1} + Y_t. \end{aligned} \tag{24}$$

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 to be an *autoregressive sequence of order* n, notation AR(n), if it satisfies equation

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_n X_{t-n} + Y_t, \quad t \in \mathbb{Z}, \tag{25}$$

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.

Equivalently, X_t can be defined by

$$X_t + a_1 X_{t-1} + \cdots + a_n X_{t-n} = \sum_{j=0}^{n} a_j X_{t-j} = Y_t, \tag{26}$$

where $a_0 = 1$.

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

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

Using this operator, 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

$$c(B) = \sum_{k=-\infty}^{\infty} c_k B^k. \tag{27}$$

This operator has usual properties of algebraic power series.

Theorem 35. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be the autoregressive sequence of order n defined by (26). If all the roots of the polynomial $a(z) = 1 + a_1 z \ + \cdots + a_n z^n$ 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

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}\frac{1}{|\sum_{j=0}^{n} a_j e^{-ij\lambda}|^2}, \quad \lambda \in [-\pi, \pi], \tag{28}$$

where $a_0 = 1$.

Proof. Consider the AR(n) sequence,

$$X_t + a_1 X_{t-1} + \cdots + a_n X_{t-n} = a(B)X_t = Y_t.$$

If all the roots $z_i, i = 1, \ldots, n$, of the polynomial $a(z) = 1 + a_1 z \ + \cdots + a_n z^n$ are outside the unit circle, then $a(z) \neq 0$ for $|z| \leq 1$. Since $|z_i| \geq \min_{1 \leq i \leq n} |z_i| \geq 1 + \delta > 1$ for a $\delta > 0$, $a(z) \neq 0$ for $|z| < 1 + \delta$, and $c(z) = \frac{1}{a(z)}$ is holomorphic in the region $|z| < 1 + \delta$ and has a representation

$$c(z) = \sum_{j=0}^{\infty} c_j z^j, |z| < 1 + \delta. \tag{29}$$

The series in (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 the causal linear process, that satisfies Theorem 34. It is centered and weakly stationary, with the autocovariance function (22) and spectral density

$$\begin{aligned} f_X(\lambda) &= \frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{\infty} e^{-ik\lambda} c_k\right|^2 = \frac{\sigma^2}{2\pi}\left|c(e^{-i\lambda})\right|^2 \\ &= \frac{\sigma^2}{2\pi}\frac{1}{|a(e^{-i\lambda})|^2} = \frac{\sigma^2}{2\pi}\frac{1}{|\sum_{j=0}^{n} a_j e^{-ij\lambda}|^2}. \end{aligned}$$

□

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:

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

where $A_1, \dots, A_n$ are constants that can be determined. For $|z| \le 1$ and $|z_j| > 1$,

$$\begin{aligned}
\frac{A_j}{z_j - z} &= \frac{A_j}{z_j\left(1 - \frac{z}{z_j}\right)} = \frac{A_j}{z_j}\sum_{k=0}^{\infty}\left(\frac{z}{z_j}\right)^k, \\
c(z) &= \sum_{j=1}^{n}\frac{A_j}{z_j - z} = \sum_{j=1}^{n}\frac{A_j}{z_j}\sum_{k=0}^{\infty}\left(\frac{z}{z_j}\right)^k \\
&= \sum_{k=0}^{\infty} z^k \sum_{j=1}^{n}\frac{A_j}{z_j^{k+1}} = \sum_{k=0}^{\infty} c_k z^k, \\
c_k &= \sum_{j=1}^{n}\frac{A_j}{z_j^{k+1}}.
\end{aligned}$$

Since for all $i = 1, \dots, n$ we have $|z_i| \ge 1 + \delta > 1$, 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 solving the system of equations

$$\begin{aligned}
c_0 &= 1, \\
c_1 + a_1 c_0 &= 0, \\
c_2 + a_1 c_1 + a_2 c_0 &= 0, \\
&\dots \\
c_p + a_1 c_{p-1} + \dots + a_n c_{p-n} &= 0, \quad p = n, n+1, \dots,
\end{aligned}$$

that we get 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

$$c_p + a_1 c_{p-1} + \dots + a_n c_{p-n} = 0$$

for $p \ge n$ can be solved as a system of homogeneous difference equations of order n with constant coefficients, and initial conditions $c_0, c_1 \dots, c_{n-1}$.

Yule-Walker equations

The autocovariance function of a stationary and real-valued autoregression sequence can be alternatively computed by using so-called Yule-Walker equations. Let us consider a sequence $\{X_t,\, t \in \mathbb{Z}\}$,

$$X_t + a_1 X_{t-1} + \dots + a_n X_{t-n} = Y_t, \tag{30}$$

that satisfies conditions of Theorem 35, with real-valued coefficients $a_1, \dots, a_n$ and with $\{Y_t,\, t \in \mathbb{Z}\}$ being the real white noise $\mathrm{WN}(0, \sigma^2)$. Since the sequence $\{X_t,\, t \in \mathbb{Z}\}$ is a real-valued causal linear process and Y_t are uncorrelated, it can be easily proved that $\mathsf{E} X_s Y_t = \langle X_s, Y_t \rangle = 0$ for $s < t$.

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

$$\mathsf{E} X_t Y_t + a_1 \mathsf{E} X_{t-1} Y_t + \dots + a_n \mathsf{E} X_{t-n} Y_t = \mathsf{E} Y_t^2,$$

thus,

$$\mathsf{E} X_t Y_t = \sigma^2.$$

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

$$\mathsf{E} X_t X_{t-k} + a_1 \mathsf{E} X_{t-1} X_{t-k} + \dots + a_n \mathsf{E} X_{t-n} X_{t-k} = \mathsf{E} Y_t X_{t-k},$$

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

$$R(0) + a_1 R(1) + \dots + a_n R(n) = \sigma^2, \qquad k = 0, \qquad (31)$$

$$R(k) + a_1 R(k-1) + \dots + a_n R(n-k) = 0, \qquad k \geq 1. \qquad (32)$$

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 $k = 1, \dots, n-1:$

$$\begin{aligned} r(1) + a_1 + a_2 r(1) + a_3 r(2) + \dots + a_n r(n-1) &= 0, \\ r(2) + a_1 r(1) + a_2 + a_3 r(1) + \dots + a_n r(n-2) &= 0, \\ &\dots \\ r(n-1) + a_1 r(n-2) + \dots + a_n r(1) &= 0. \end{aligned}$$

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

$$r(k) + a_1 r(k-1) + \dots + a_n r(n-k) = 0, \ k \geq n,$$

with the characteristic polynomial

$$\lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n = L(\lambda).$$

In this way we get the solution $r(t)$ for $t \geq 0$. For a real-valued sequence, $r(t) = r(-t)$. If we insert $R(k) = r(k)R(0)$ into (31) we get the equation for $R(0)$:

$$R(0)[1 + a_1 r(1) + \cdots + a_n r(n)] = \sigma^2,$$

thus

$$R(0) = \frac{\sigma^2}{1 + a_1 r(1) + \cdots + a_n r(n)}. \tag{33}$$

Remark 12. If $z_i, i = 1, \ldots, n$, are the roots of the polynomial $a(z) = 1 + a_1 z + \cdots + a_n z^n$, then $\lambda_i = z_i^{-1}, i = 1, \ldots, n$, are the roots of the polynomial $L(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. 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. Consider the AR(1) sequence

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

Polynomial $a(z) = 1 + az$ has the root $-\frac{1}{a}$ that is outside the unit circle; it means that $\{X_t,\ t \in \mathbb{Z}\}$ is a weakly stationary causal linear process. The Yule - Walker equations for the autocovariance function $R_X(t) = R(t)$ are now

$$\begin{aligned} R(0) + aR(1) &= \sigma^2, \\ R(k) + aR(k-1) &= 0, \quad k \geq 1, \end{aligned}$$

A general solution to the difference equation for the autocorrelation function is $r(k) = c(-a)^k$, the initial condition is $r(0) = 1 = c$. 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 the AR(2) sequence

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

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 the causal linear process that is weakly stationary. The Yule-Walker equations are

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

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

The equations for the autocorrelation function are of the form

$$\begin{aligned} r(1) - \frac{3}{4} + \frac{1}{8}r(1) &= 0, \\ r(k) - \frac{3}{4}r(k-1) + \frac{1}{8}r(k-2) &= 0, \qquad k \geq 2. \end{aligned} \tag{34}$$

Solving the first equation we get $r(1) = \frac{2}{3}$. For $k \geq 2$ we solve the 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-valued 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

$$\begin{aligned} c_1 + c_2 &= r(0), \\ \lambda_1 c_1 + \lambda_2 c_2 &= r(1), \end{aligned}$$

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

$$\begin{aligned} r(k) &= \frac{5}{3}\left(\frac{1}{2}\right)^k - \frac{2}{3}\left(\frac{1}{4}\right)^k, \quad k = 0, 1, \ldots \\ r(k) &= r(-k), \quad k = -1, -2, \ldots \end{aligned}$$

The value of $R(0)$ can be obtained from (33).

8.5 ARMA sequences

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

$$X_t + a_1X_{t-1} + \cdots + a_mX_{t-m} = Y_t + b_1Y_{t-1} + \cdots + b_nY_{t-n}, \quad t \in \mathbb{Z}, \tag{35}$$

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 we can write

$$X_t = \varphi_1X_{t-1} + \cdots + \varphi_mX_{t-m} + Y_t + \theta_1Y_{t-1} + \cdots + \theta_nY_{t-n}.$$

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

$$a(B)X_t = b(B)Y_t. \tag{36}$$

Theorem 36. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be the 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

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\sum_{j=0}^{n} b_j e^{-ij\lambda}|^2}{|\sum_{k=0}^{m} a_k e^{-ik\lambda}|^2}, \quad \lambda \in [-\pi, \pi], \tag{37}$$

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, for $|z| \leq 1$ it holds

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

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

$$h(B)a(B)X_t = X_t = h(B)b(B)Y_t = c(B)Y_t,$$

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

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

$$\begin{aligned} f_X(\lambda) &= \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} c_j e^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| c(e^{-i\lambda}) \right|^2 = \frac{\sigma^2}{2\pi} \left| \frac{b(e^{-i\lambda})}{a(e^{-i\lambda})} \right|^2 \\ &= \frac{\sigma^2}{2\pi} \frac{|\sum_{j=0}^{n} b_j e^{-ij\lambda}|^2}{|\sum_{k=0}^{m} a_k e^{-ik\lambda}|^2}. \end{aligned}$$

□

Remark 13. 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 the 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 the coefficients with the same powers of z we obtain

$$c_0 = 1, \; c_j = (-a)^{j-1}(b - a), \; j \geq 1.$$

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 $R(0)$:

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

For $k \geq 1$,

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

The spectral density is

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|1+be^{-i\lambda}|^2}{|1+ae^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \cdot \frac{1+2b\cos\lambda + b^2}{1+2a\cos\lambda + a^2}, \quad \lambda \in [-\pi, \pi].$$

We can also use an analogy of the Yule-Walker equations. Multiplying (35) by $X_{t-k}, k \geq 0$ and taking the expectation we get

$$\mathsf{E}X_t X_{t-k} + \sum_{j=1}^{m} a_j \mathsf{E}X_{t-j}X_{t-k} = \mathsf{E}Y_t X_{t-k} + \sum_{j=1}^{n} b_j \mathsf{E}Y_{t-j}X_{t-k}.$$

From Theorem 36 and properties of the white noise,

$$\mathsf{E}X_{t-j}Y_{t-k} = \begin{cases} \sigma^2 c_{k-j}, & k \geq j, \\ 0, & k < j \end{cases}$$

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

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

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

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

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

$$\begin{aligned} R(0) + aR(1) &= \sigma^2 + b(b-a)\sigma^2, \\ R(1) + aR(0) &= \sigma^2 b, \\ R(k) + aR(k-1) &= 0, \quad k \geq 2. \end{aligned}$$

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:

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

which is the same result as before.

Definition 35. Let $\{X_t,\, t\in\mathbb{Z}\}$ be the 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

$$Y_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \quad t\in\mathbb{Z}. \tag{40}$$

Let us study conditions under which an ARMA sequence is invertible.

Theorem 37. *Let $\{X_t,\, t\in\mathbb{Z}\}$ be the stationary ARMA(m,n) random sequence defined by (36). Let the polynomials $a(z)$ and $b(z)$ have no common roots and the polynomial $b(z)=1+b_1z+\cdots+b_nz^n$ 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

$$d(z) = \sum_{j=0}^{\infty} d_j z^j = \frac{a(z)}{b(z)}, \quad |z|\le 1.$$

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. □

Remark 14. Let us notice that the equation $d(z)b(z)=a(z)$ with polynomials $a(z) = 1+a_1z+\cdots+a_mz^m$, $b(z)=1+b_1z+\cdots+b_nz^n$, respectively, implies $d_0=1$. Relation (40) can be written as

$$X_t + \sum_{j=1}^{\infty} d_j X_{t-j} = Y_t, \quad t\in\mathbb{Z}. \tag{41}$$

The invertible ARMA(m,n) sequence can be thus expressed as an AR(∞) 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

$$X_t = \sum_{j=-\infty}^{\infty} c_j Y_{t-j}, \quad t \in \mathbb{Z}. \tag{42}$$

The sequence $\{c_j,\, j \in \mathbb{Z}\}$ is called *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 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 $\sum_{-\infty}^{\infty} |c_k| < \infty$. 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*

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

and the spectral density

$$f_X(\lambda) = |\Psi(\lambda)|^2 f_Y(\lambda), \quad \lambda \in [-\pi, \pi], \tag{43}$$

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

$$\begin{aligned} X_t^{(n)} &= \sum_{k=-n}^{n} c_k Y_{t-k} = \sum_{k=-n}^{n} c_k \int_{-\pi}^{\pi} e^{i(t-k)\lambda} dZ_Y(\lambda) \\ &= \int_{-\pi}^{\pi} e^{it\lambda} \sum_{k=-n}^{n} c_k e^{-ik\lambda} dZ_Y(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} h_n(\lambda) dZ_Y(\lambda), \end{aligned}$$

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

$$X_t = \underset{n\to\infty}{\text{l. i. m.}}\ X_t^{(n)} = \int_{-\pi}^{\pi} e^{it\lambda}\Psi(\lambda)dZ_Y(\lambda)$$

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 Theorem 12 the autocovariance functions $\{X_t^{(n)},\, t \in \mathbb{Z}\}$ converge to the autocovariance function of $\{X_t,\, t \in \mathbb{Z}\}$, and so

$$\begin{aligned} \mathsf{E}X_{s+t}\overline{X_s} &= \lim_{n\to\infty} \mathsf{E}X_{s+t}^n\overline{X_s^n} = \lim_{n\to\infty} \sum_{j=-n}^{n}\sum_{k=-n}^{n} c_j\overline{c_k}\mathsf{E}(Y_{s+t-j}\overline{Y_{s-k}}) \\ &= \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_j\overline{c_k}R_Y(t-j+k) := R_X(t). \end{aligned}$$

Since $\mathsf{E}X_{s+t}\overline{X_s} = R_X(t)$ is a function of one variable, only, $\{X_t,\, t \in \mathbb{Z}\}$ is weakly stationary.

It also holds

$$\begin{aligned} R_X(t) &= \mathsf{E}\int_{-\pi}^{\pi} e^{i(t+s)\lambda}\Psi(\lambda)dZ_Y(\lambda)\overline{\int_{-\pi}^{\pi} e^{is\lambda}\Psi(\lambda)dZ_Y(\lambda)} \\ &= \int_{-\pi}^{\pi} e^{it\lambda}|\Psi(\lambda)|^2 dF_Y(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda}|\Psi(\lambda)|^2 f_Y(\lambda)d\lambda \end{aligned}$$

and from the spectral decomposition of the autocovariance function (Theorem 19) it follows that the function

$$|\Psi(\lambda)|^2 f_Y(\lambda) := f_X(\lambda)$$

is the spectral density of the sequence $\{X_t,\, t \in \mathbb{Z}\}$. □

Example 30. Let $\{Y_t,\, t \in \mathbb{Z}\}$ be a white noise 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 the 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 value μ is *mean square ergodic* or it satisfies *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

$$\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{\text{P}} \mu,$$

i.e., $\{X_t,\ t \in \mathbb{Z}\}$ satisfies *the weak law of large numbers* for stationary sequences.

Theorem 39. *A stationary random sequence $\{X_t,\, t \in \mathbb{Z}\}$ with mean value μ and auto-covariance function R is mean square ergodic if and only if*

$$\frac{1}{n} \sum_{t=1}^{n} R(t) \to 0 \;\; \textit{as } n \to \infty. \tag{45}$$

Proof. W.l.o.g. put $\mu = 0$ (otherwise we consider $\widetilde{X}_t := X_t - \mu$). Consider the spectral decomposition

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

where $\{Z_\lambda,\, \lambda \in [-\pi, \pi]\}$ is the 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

$$\frac{1}{n}\sum_{t=1}^{n} X_t = \frac{1}{n}\sum_{t=1}^{n}\Big(\int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)\Big) = \int_{-\pi}^{\pi}\Big(\frac{1}{n}\sum_{t=1}^{n} e^{it\lambda}\Big) dZ(\lambda)$$
$$= \int_{-\pi}^{\pi} h_n(\lambda) dZ(\lambda),$$

where

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

Further, let us consider function

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

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

$$Z_0 = 0 \ \text{ a.s.} \iff \frac{1}{n}\sum_{t=1}^{n} R(t) \to 0 \text{ as } n \to \infty. \tag{46}$$

From Theorem 26 we have $\mathsf{E}Z_0 = 0$; thus $Z_0 = 0$ a. s. if and only if $\mathsf{E}|Z_0|^2 = 0$. Further from Theorem 26,

$$\mathsf{E}|Z_0|^2 = \mathsf{E}\Big|\int_{-\pi}^{\pi} h(\lambda) dZ(\lambda)\Big|^2 = \int_{-\pi}^{\pi} |h(\lambda)|^2 dF(\lambda).$$

From the spectral decomposition of the autocovariance function and the Lebesgue theorem

$$\frac{1}{n}\sum_{t=1}^{n} R(t) = \frac{1}{n}\sum_{k=1}^{n}\Big[\int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda)\Big] = \int_{-\pi}^{\pi}\Big(\frac{1}{n}\sum_{t=1}^{n} e^{it\lambda}\Big) dF(\lambda)$$
$$= \int_{-\pi}^{\pi} h_n(\lambda) dF(\lambda) \to \int_{-\pi}^{\pi} h(\lambda) dF(\lambda) = \int_{-\pi}^{\pi} |h(\lambda)|^2 dF(\lambda) \tag{47}$$

The rest of the proof follows from (46) and(47). □

Example 32. Let us consider the AR(1) process

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

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,

$$\frac{1}{n}\sum_{t=1}^{n} R_X(t) = \frac{\sigma^2}{1-\varphi^2}\frac{1}{n}\sum_{t=1}^{n}\varphi^t = \frac{1}{n}\frac{\sigma^2}{1-\varphi^2}\frac{\varphi(1-\varphi^n)}{1-\varphi} \to 0$$

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 expected value μ 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 $\mathsf{E}Y = 0,\ varY = \sigma^2 \in (0,\infty)$, and $\mathsf{E}X_t\overline{Y} = 0\ \forall t \in \mathbb{Z}$.

Then $\mathsf{E}Z_t = \mathsf{E}X_t + \mathsf{E}Y = \mu$ for all $t \in \mathbb{Z}$ and

$$\mathsf{E}(Z_{s+t}-\mu)(\overline{Z_t-\mu}) = R_X(s) + \sigma^2 := R_Z(s),$$

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 value μ and autocovariance function R, such that $\sum_{t=-\infty}^{\infty} |R(t)| < \infty$. Then, as $n \to \infty$*

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \to \mu \textit{ in mean square,} \tag{48}$$

$$n \, var \overline{X}_n \to \sum_{k=-\infty}^{\infty} R(k). \tag{49}$$

Proof. 1. $\sum_{k=-\infty}^{\infty} |R(k)| < \infty \Rightarrow R(k) \to 0$ 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.

2. We have

$$\begin{aligned}
\operatorname{var} \overline{X}_n &= \operatorname{var}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) \\
&= \frac{1}{n^2}\Big[\sum_{k=1}^{n} \operatorname{var} X_k + \sum\sum_{1\le j\ne k\le n} \operatorname{cov}(X_j, X_k)\Big] \\
&= \frac{1}{n^2}\Big[nR(0) + 2\sum_{j=1}^{n-1}(n-j)R(j)\Big] \\
&= \frac{1}{n}\Big[R(0) + 2\sum_{j=1}^{n-1}\Big(1-\frac{j}{n}\Big)R(j)\Big] \\
&= \frac{1}{n}\sum_{j=-n+1}^{n-1}\left(1-\frac{|j|}{n}\right)R(j).
\end{aligned} \tag{50}$$

Thus,

$$n \operatorname{var} \overline{X}_n = \sum_{j=-n+1}^{n-1} R(j) - \frac{2}{n}\sum_{j=1}^{n-1} jR(j).$$

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

Remark 15. From Theorem 40 we also get

$$\lim_{n\to\infty} n \operatorname{var} \overline{X}_n = \sum_{k=-\infty}^{\infty} R(k) = 2\pi f(0)$$

where f is the spectral density of the sequence $\{X_t,\, t \in \mathbb{Z}\}$.

.

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

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

Remark 16. The existence of the integral $\int_0^\tau X_t dt$ is guarranted by Theorem 18 since the autocovariance function of the stationary mean square continuous process is continuous and the expected value μ is constant.

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*

$$\frac{1}{\tau}\int_0^\tau R(t)dt \to 0 \quad as \quad \tau \to \infty.$$

Proof. Rozanov (1963), Chap. 1, § 6. □

Theorem 42. *Let $\{X_t,\ t \in \mathbb{R}\}$ be a real-valued stationary, mean square continuous process, with mean value μ and autocovariance function R, such that $\int_{-\infty}^{\infty} |R(t)|dt < \infty$.*
Then, as $\tau \to \infty$

$$\overline{X}_\tau = \frac{1}{\tau}\int_0^\tau X_t dt \to \mu \text{ in mean square,} \tag{51}$$

$$\tau\, var \overline{X}_\tau \to \int_{-\infty}^{\infty} R(t)dt. \tag{52}$$

Proof. Bosq and Nguyen (1996), Theorems 9.11 and 15.1. □

Example 34. Let $\{X_t, t \in \mathbb{R}\}$ be a stationary centered stochastic process with the autocovariance function

$$R(t) = ce^{-\alpha|t|},\ t \in \mathbb{R},\ \alpha > 0,\ c > 0.$$

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 $\tau\, \text{var}\, \overline{X}_\tau \to \frac{2c}{\alpha}$.

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* $n \to \infty$, $X_n \xrightarrow{\text{D}} X,\ Y_n \xrightarrow{\text{P}} 0.$ *Then* $X_n + Y_n \xrightarrow{\text{D}} X$ *as* $n \to \infty$.

Proof. Brockwell and Davis (1991), Proposition 6.3.3. □

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* ψ *be random variables such that*

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

Then

$$\xi_n \xrightarrow{\text{D}} \psi \text{ as } n \to \infty.$$

Proof. Brockwell and Davis (1991), Proposition 6.3.9. □

Theorem 45. *(Lévy-Lindeberg CLT) Let* $\{Y_t, t \in \mathbb{Z}\}$ *be a sequence of independent identically distributed random variables with mean* μ *and finite positive variance* σ^2. *Let* $\overline{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$. *Then, as* $n \to \infty$

$$\sqrt{n}\frac{\overline{Y}_n - \mu}{\sigma} \xrightarrow{\text{D}} \mathcal{N}(0, 1). \tag{53}$$

Proof. Brockwell and Davis (1991), Theorem 6.4.1. □

Theorem 46. *(Cramér-Wold Theorem) Let* $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots,$ *be* k*-dimensional random vectors. Then*

$$\mathbf{X}_n \xrightarrow{\text{D}} \mathbf{X} \quad \text{as } n \to \infty$$

if and only if for every $\mathbf{c} \in \mathbb{R}_k$

$$\mathbf{c}'\mathbf{X}_n \xrightarrow{\text{D}} \mathbf{c}'\mathbf{X} \quad \text{pro } n \to \infty.$$

Proof. Brockwell and Davis (1991), Proposition 6.3.1. □

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 (i. i. d.) random variables with zero mean and finite positive variance σ^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 $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{\text{D}} N(0, \Delta^2), \tag{54}$$

where $\Delta^2 = \sigma^2 \left(\sum_{j=0}^{m} b_j \right)^2$.

Proof. We can write

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big(\sum_{j=0}^{m} b_j Y_{t-j} \Big) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{b_1}{\sqrt{n}} \sum_{t=1}^{n} Y_{t-1} + \cdots + \frac{b_m}{\sqrt{n}} \sum_{t=1}^{n} Y_{t-m} \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{b_1}{\sqrt{n}} \Big(\sum_{t=1}^{n} Y_t + Y_0 - Y_n \Big) + \ldots \\
&+ \frac{b_m}{\sqrt{n}} \Big(\sum_{t=1}^{n} Y_t + \sum_{k=-m+1}^{0} Y_k - \sum_{j=n-m+1}^{n} Y_j \Big) \\
&= \Big(\sum_{j=0}^{m} b_j \Big) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{1}{\sqrt{n}} \xi_n,
\end{aligned}$$

where

$$\xi_n = \sum_{s=1}^{m} Y_{1-s} \Big(\sum_{j=s}^{m} b_j \Big) - \sum_{s=0}^{m-1} Y_{n-s} \Big(\sum_{j=s+1}^{m} b_j \Big)$$

is a finite linear combination of $2m$ i. i. d. random variables $Y_0, Y_{-1}, \ldots, Y_{-m+1}$ and $Y_n, Y_{n-1}, \ldots, Y_{n-m+1}$ with zero mean and variance σ^2.

According to Theorem 45, $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \xrightarrow{\text{D}} \mathcal{N}(0, \sigma^2)$ as $n \to \infty$. From here

$$\Big(\sum_{j=0}^{m} b_j \Big) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \xrightarrow{\text{D}} \mathcal{N}(0, \Delta^2), \text{ where } \Delta^2 = \sigma^2 \Big(\sum_{j=0}^{m} b_j \Big)^2. \tag{55}$$

Now, using Theorem 43 it suffices to proof that $\frac{1}{\sqrt{n}}\xi_n \xrightarrow{\text{P}} 0$ as $n \to \infty$. But it holds, since as $n \to \infty$

$$P\left(\left|\frac{1}{\sqrt{n}}\xi_n\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\mathsf{E}\left(\frac{1}{n}\xi_n^2\right) = \frac{1}{\epsilon^2}\frac{\sigma^2 \cdot \text{const}}{n} \to 0.$$

□

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 i. i. d. random variables with zero mean and finite positive variance σ^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 $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(X_t - \mu) \xrightarrow{\text{D}} \mathcal{N}(0, \Delta^2),$$

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 $k \in \mathbb{N}$

$$S_{kn} \xrightarrow{\text{D}} \psi_k, \tag{56}$$

where $\psi_k \sim \mathcal{N}(0, \Delta_k^2)$, $\Delta_k^2 = \sigma^2(\sum_{j=0}^k b_j)^2$. Further, from the assumptions of the theorem it follows that

$$\Delta_k^2 = \sigma^2 \left(\sum_{j=0}^k b_j \right)^2 \to \sigma^2 \left(\sum_{j=0}^\infty b_j \right)^2 = \Delta^2,$$

as $k \to \infty$, and thus

$$\psi_k \xrightarrow{\mathrm{D}} \mathcal{N}(0, \Delta^2). \tag{57}$$

According to the Chebyshev inequality

$$\begin{aligned} P(|\xi_n - S_{kn}| > \epsilon) = P(|D_{kn}| > \epsilon) &\leq \frac{1}{\epsilon^2} \mathrm{var} D_{kn} \\ &= \frac{1}{\epsilon^2} \mathrm{var} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n V_{kt} \right). \end{aligned}$$

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

$$\begin{aligned} &P(|\xi_n - S_{kn}| > \epsilon) \leq \frac{1}{\epsilon^2} \mathrm{var}\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^n V_{kt} \Big) \\ &= \frac{1}{\epsilon^2} \sum_{j=-n+1}^{n-1} R_V(j) \Big(1 - \frac{|j|}{n} \Big) \leq \frac{1}{\epsilon^2} \sum_{j=-n+1}^{n-1} |R_V(j)| \\ &= \frac{1}{\epsilon^2} \Big[R_V(0) + 2 \sum_{j=1}^{n-1} |R_V(j)| \Big] \\ &= \frac{\sigma^2}{\epsilon^2} \Big[\sum_{j=k+1}^\infty b_j^2 + 2 \sum_{j=1}^{n-1} \Big| \sum_{\nu=k+1}^\infty b_\nu b_{\nu+j} \Big| \Big] \\ &\leq \frac{\sigma^2}{\epsilon^2} \Big[\sum_{j=k+1}^\infty b_j^2 + 2 \sum_{j=1}^{n-1} \sum_{\nu=k+1}^\infty |b_\nu| |b_{\nu+j}| \Big] \\ &\leq \frac{\sigma^2}{\epsilon^2} \Big[\sum_{j=k+1}^\infty |b_j|^2 + 2 \sum_{\nu=k+1}^\infty |b_\nu| \sum_{j=1}^\infty |b_{\nu+j}| \Big] = \frac{\sigma^2}{\epsilon^2} \Big(\sum_{j=k+1}^\infty |b_j| \Big)^2, \end{aligned}$$

so that

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

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 $n \to \infty$,

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

□

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

$$X_t = \mu + \sum_{j=0}^{\infty} a^j Y_{t-j}, \ t \in \mathbb{Z}.$$

Since $\sum_{j=0}^{\infty} a^j \neq 0$, it holds, as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{\text{D}} \mathcal{N}(0, \Delta^2), \quad \Delta^2 = \sigma^2 \frac{1}{(1-a)^2}.$$

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

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 $(\dots, X_{t-1}, X_t)$ and $(X_{t+m+1}, X_{t+m+2}, \dots)$ are independent.

Remark 17. A sequence of i. i. d. 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

$$X_t = Y_t Y_{t+m}, \ t \in \mathbb{Z},$$

for some $m \in \mathbb{N}$. Then

- $\mathsf{E}X_t = \mathsf{E}(Y_t Y_{t+m}) = 0$,
- $\mathsf{E}X_s X_t = \mathsf{E}(Y_t Y_{t+m} Y_s Y_{s+m}) = 0$ pro $t \neq s$.

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 $n \to \infty$,

$$n \, var\overline{X}_n \longrightarrow \Delta_m^2, \tag{59}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \stackrel{\text{D}}{\longrightarrow} \mathcal{N}(0, \Delta_m^2). \tag{60}$$

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 40 we have

$$\lim_{n\to\infty} n \operatorname{var}\overline{X}_n = \sum_{k=-\infty}^{\infty} R(k) = \sum_{k=-m}^{m} R(k) = \Delta_m^2.$$

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

$$\begin{aligned}
(X_1, \ldots, X_n) &= (U_1, V_1, U_2, V_2, \ldots, U_r, V_r), \\
U_j &= (X_{(j-1)k+1}, \ldots, X_{jk-m}), && j = 1, \ldots, r, \\
V_j &= (X_{jk-m+1}, \ldots, X_{jk}), && j = 1, \ldots, r.
\end{aligned}$$

$U_1, \ldots, U_r$ 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 i. i. d. Thus,

$$\sum_{t=1}^{n} X_t = \sum_{j=1}^{r} S_j + \sum_{j=1}^{r} T_j,$$

$S_j,\ j = 1, \ldots, r$, are i. i. d. (S_j is the sum of elements of the vector U_j,)
$T_j,\ j = 1, \ldots, r$, are i. i. d. (the sum of elements of the vectors V_j).

For $k > 2m$ we have $\mathsf{E}S_1 = 0,\ \mathsf{E}T_1 = 0$ and

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

Similarly, utilizing the strict stationarity we have

$$\begin{aligned}\operatorname{var} T_1 &= \operatorname{var}(X_{k-m+1} + \cdots + X_k) = \operatorname{var}(X_1 + \cdots + X_m) \\ &= \sum_{\nu=-m+1}^{m-1} (m - |\nu|) R(\nu) = \delta_m^2.\end{aligned}$$

Now, we can write

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t := \xi_n = S_{kn} + D_{kn}, \tag{61}$$

where

$$S_{kn} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n/k} S_j = \frac{1}{\sqrt{k}} \frac{1}{\sqrt{r}} \sum_{j=1}^{r} S_j, \tag{62}$$

$$D_{kn} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n/k} T_j = \frac{1}{\sqrt{k}} \frac{1}{\sqrt{r}} \sum_{j=1}^{r} T_j. \tag{63}$$

Form the Lévy-Lindeberg theorem, as $r \to \infty$,

$$\frac{1}{\sqrt{r}} \sum_{j=1}^{r} S_j \xrightarrow{\text{D}} \mathcal{N}(0, \Delta_{mk}^2).$$

For a fixed k and $r \to \infty$ also $n \to \infty$, so that

$$S_{kn} \xrightarrow{\text{D}} \psi_k, \tag{64}$$

where ψ_k has the normal distribution $\mathcal{N}\left(0, \frac{\Delta_{mk}^2}{k}\right)$. As $k \to \infty$,

$$\frac{\Delta_{mk}^2}{k} \to \sum_{j=-m}^{m} R(j) = \Delta_m^2,$$

$$\psi_k \xrightarrow{\text{D}} \mathcal{N}(0, \Delta_m^2). \tag{65}$$

From the Chebyshev inequality,

$$\begin{aligned}P(|\xi_n - S_{kn}| > \epsilon) = P(|D_{kn}| > \epsilon) &\le \frac{1}{\epsilon^2} \cdot \frac{1}{n} \operatorname{var}\left(\sum_{j=1}^{r} T_j\right) \\ &= \frac{1}{\epsilon^2} \cdot \frac{1}{k} \operatorname{var} T_1 = \frac{1}{\epsilon^2 k} \delta_m^2.\end{aligned}$$

Thus,

$$\lim_{k\to\infty} \overline{\lim_{n\to\infty}}\, P(|\xi_n - S_{kn}| > \epsilon) = 0 \tag{66}$$

and the proof follows from (64), (65), (66) and Theorem 44. □

Example 38. Consider the sequence $\{X_t,\, t \in \mathbb{Z}\}$

$$X_t = \mu + Y_t + a_1 Y_{t-1} + a_2 Y_{t-2},\ t \in \mathbb{Z},$$

where Y_t are i. i. d., $\mathsf{E}Y_t = 0, \operatorname{var} \mathrm{Y_t} = \sigma^2 > 0$.

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

$$\begin{aligned}
&R(0) = \sigma^2(1 + a_1^2 + a_2^2),\\
&R(1) = \sigma^2(a_1 + a_1 a_2) = R(-1),\\
&R(2) = \sigma^2 a_2 = R(-2),\\
&R(k) = 0, \quad |k| > 2.
\end{aligned}$$

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, $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \Delta_m^2)$, provided $\Delta_m^2 \neq 0$.

Example 39. Let $\{Y_t,\, t \in \mathbb{Z}\}$ be a sequence of i. i. d. random variables, $\mathsf{E}Y_t = 0, \operatorname{var} \mathrm{Y_t} = \sigma^2, \mathrm{EY_t^4} < \infty$. Prove that for every $k > 0$ as $n \to \infty$ it holds

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t^2 - \sigma^2) &\xrightarrow{\mathrm{D}} \mathcal{N}(0, \tau^2),\\
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t Y_{t+k} &\xrightarrow{\mathrm{D}} \mathcal{N}(0, \sigma^4),\\
\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} Y_t Y_{t+k} &\xrightarrow{\mathrm{D}} \mathcal{N}(0, \sigma^4),\\
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_t &\xrightarrow{\mathrm{D}} \mathcal{N}_k(\mathbf{0}, \sigma^4 \mathbf{I}),
\end{aligned}$$

where $\tau^2 = \operatorname{var} Y_1^2$, $\mathbf{X}_t = (Y_t Y_{t+1}, \ldots, Y_t Y_{t+k})'$ and $\mathbf{I}$ is the identity matrix of order k.

Solution.

1. Y_t^2 are i. i. d., $\mathsf{E}Y_t^2 = \sigma^2$, $\operatorname{var} Y_t^2 = \tau^2$. The Central limit theorem (Theorem 45) implies that

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

2. Denote $X_t := Y_t Y_{t+k}$ for $k > 0$. The sequence $\{X_t,\, t \in \mathbb{Z}\}$ is strictly stationary, $\mathsf{E}X_t = 0,\ \mathsf{E}X_t^2 = \sigma^4$, X_t are mutually uncorrelated but k-dependent. By Theorem 49,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{\text{D}} \mathcal{N}(0, \Delta_k^2),$$

where $\Delta_k^2 = \sum_{j=-k}^{k} R_X(j) = \sigma^4$.

3. We can write

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

From step 2, as $n \to \infty$,

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

Form the Chebyshev inequality, as $n \to \infty$,

$$P\left(\Big|\frac{1}{\sqrt{n}} \sum_{t=n-k+1}^{n} Y_t Y_{t+k}\Big| > \epsilon\right) = P\left(\Big|\frac{1}{\sqrt{n}} \sum_{t=n-k+1}^{n} X_t\Big| > \epsilon\right)$$
$$\leq \frac{1}{\epsilon^2\, n} \mathsf{E}\Big(\sum_{t=n-k+1}^{n} X_t\Big)^2 = \frac{1}{\epsilon^2\, n} \sum_{t=n-k+1}^{n} \mathsf{E}X_t^2 = \frac{1}{\epsilon^2} \frac{k\,\sigma^4}{n} \to 0$$

since k is fixed.

4. Define $Z_t := \mathbf{c}'\mathbf{X}_t, t \in \mathbb{Z}, \mathbf{c} \in \mathbb{R}_k$. Then

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

By Theorem 49,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t \xrightarrow{\text{D}} \mathcal{N}(0, \Delta_k^2),$$

where $\Delta_k^2 = \sum_{j=-k}^{k} R_Z(j) = \sigma^4\mathbf{c}'\mathbf{I}\mathbf{c}$. 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 $M^\perp$. □

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 $x = \widehat{x} + (x - \widehat{x})$, such that $\widehat{x} \in M$ and $x - \widehat{x} \in M^\perp$. Further*

$$||x - \widehat{x}|| = \min_{y \in M} ||x - y|| \tag{67}$$

and

$$||x||^2 = ||\widehat{x}||^2 + ||x - \widehat{x}||^2. \tag{68}$$

Proof. Rudin (2003), Theorem 4.11, or Brockwell and Davis (1992), Theorem 2.3.1. □

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

$$x = P_M x + (x - P_M x) = P_M x + (I - P_M)x. \tag{69}$$

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 $\alpha, \beta \in \mathbb{C}$, $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$.*

2. *If $x \in M$, then $P_M x = x$.*

3. *If* $x \in M^\perp$*, then* $P_M x = 0$.

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* $||x_n - x|| \to 0$ *as* $n \to \infty$*, then* $||P_M x_n - P_M x|| \to 0$.

Proof.

1. By using (69) we get

$$\begin{aligned}\alpha x + \beta y &= \alpha(P_M x + (x - P_M x)) + \beta(P_M y + (y - P_M y)) \\ &= \alpha P_M x + \beta P_M y + \alpha(x - P_M x) + \beta(y - P_M y).\end{aligned}$$

Obviously,

$$\alpha P_M x + \beta P_M y \in M, \ \alpha(x - P_M x) + \beta(y - P_M y) \in M^\perp$$

since M and $M^\perp$ are linear subspaces, and thus, $\alpha P_M x + \beta P_M y = P_M(\alpha x + \beta y)$.

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)

$$||P_M x_n - P_M x||^2 = ||P_M(x_n - x)||^2 \leq ||x_n - x||^2.$$

□

10.2 Prediction based on finite history

Let us consider the following problem: We have random variables $X_1, \ldots, X_n$ with zero mean and finite second order moments. Utilizing observations $X_1, \ldots, X_n$ we want to forecast X_{n+h}, where $h > 0$. We would like 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

$$\mathsf{E} \left| X_{n+h} - g\left(X_1, \ldots, X_n\right)\right|^2.$$

It is well known that the best approximation is given by the conditional mean value

$$g(X_1, \dots, X_n) = \mathsf{E}(X_{n+h}|X_1, \dots, X_n).$$

Indeed, if (for the simplicity, we consider only real-valued random variables) we denote $(X_1, \dots, X_n)' = \mathbf{X}_n$, we can write

$$\begin{aligned}
&\mathsf{E}\left(X_{n+h} - g(\mathbf{X}_n)\right)^2 \\
&= \mathsf{E}\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n) + \mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)^2 \\
&= \mathsf{E}\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)^2 + \mathsf{E}\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)^2 \\
&+ 2\mathsf{E}\left[\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)\right],
\end{aligned}$$

where the last summand is

$$\begin{aligned}
&\mathsf{E}\left[\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)\right] \\
&= \mathsf{E}\left[\mathsf{E}\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)|\mathbf{X}_n\right] \\
&= \mathsf{E}\left[\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)\right] = 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathsf{E}(X_{n+h} - g(\mathbf{X}_n))^2 \\
&= \mathsf{E}\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)^2 + \mathsf{E}\left(\mathsf{E}(X_{n+h}|\mathbf{X}_n) - g(\mathbf{X}_n)\right)^2 \\
&\geq \mathsf{E}\left(X_{n+h} - \mathsf{E}(X_{n+h}|\mathbf{X}_n)\right)^2
\end{aligned}$$

with equality for $g(\mathbf{X}_n) = \mathsf{E}(X_{n+h}|\mathbf{X}_n)$.

In the next, we will confine ourselves to linear functions of $X_1, \dots, X_n$. 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 $X_1, \dots, X_n$ will be denoted by $\widehat{X}_{n+h}(n)$.

Direct method

Let $H := \mathcal{H}\{X_1, \dots, X_n, \dots, X_{n+h}\}$ be the Hilbert space generated by centered random variables $X_1, \dots, X_{n+h}$ and $H_1^n := \mathcal{H}\{X_1, \dots, X_n\}$ be the Hilbert subspace generated by random variables $X_1, \dots, X_n$.

The best linear prediction of X_{n+h} is the random variable

$$\widehat{X}_{n+h}(n) = \sum_{j=1}^{n} c_j X_j \in H_1^n, \tag{70}$$

such that

$$\mathsf{E}|X_{n+h} - \widehat{X}_{n+h}(n)|^2 = \|X_{n+h} - \widehat{X}_{n+h}(n)\|^2$$

takes minimum with respect to all linear combinations of $X_1, \dots, X_n$.

It means that

$$\widehat{X}_{n+h}(n) = P_{H_1^n}(X_{n+h}) \in H_1^n,$$

$$X_{n+h} - \widehat{X}_{n+h}(n) \perp H_1^n \tag{71}$$

and the element $\widehat{X}_{n+h}(n)$ is determined uniquely due to the projection theorem.

Since the space H_1^n is a linear span generated by $X_1, \dots, X_n$, condition (71) is satisfied if and only if

$$X_{n+h} - \widehat{X}_{n+h}(n) \perp X_j, \;\; j = 1, \dots, n,$$

i.e., if and only if

$$\mathsf{E}(X_{n+h} - \widehat{X}_{n+h}(n))\overline{X}_j = 0, \;\; j = 1, \dots, n.$$

Constants $c_1, \dots, c_n$ can be therefore obtained from the equations

$$\mathsf{E}\big(X_{n+h} - \sum_{k=1}^{n} c_k X_k\big)\overline{X}_j = 0, \;\; j = 1, \dots, n. \tag{72}$$

For $X_1, \dots X_{n+h}$ supposed to be elements of a real-valued centered stationary sequence with the autocovariance function R, system (72) is of the form

$$\sum_{k=1}^{n} c_k R(k-j) = R(n+h-j), \quad j = 1, \dots, n, \tag{73}$$

or

$$\begin{aligned} c_1R(0) + c_2R(1) + \cdots + c_nR(n-1) &= R(n+h-1), \\ c_1R(1) + c_2R(0) + \cdots + c_nR(n-2) &= R(n+h-2), \\ &\dots \\ c_1R(n-1) + c_2R(n-2) + \cdots + c_nR(0) &= R(h). \end{aligned}$$

Equivalently, system (73) can be written in the form

$$\mathbf{\Gamma}_n \mathbf{c}_n = \boldsymbol{\gamma}_{nh}$$

where $\mathbf{c}_n := (c_1, \dots, c_n)'$, $\boldsymbol{\gamma}_{nh} := (R(n+h-1), \dots, R(h))'$ and

$$\mathbf{\Gamma}_n := \begin{pmatrix} R(0) & R(1) & \dots & R(n-1) \\ R(1) & R(0) & \dots & R(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(n-1) & R(n-2) & \dots & R(0) \end{pmatrix},$$

Provided that $\mathbf{\Gamma}_n^{-1}$ exists we get $\mathbf{c}_n = \mathbf{\Gamma}_n^{-1}\boldsymbol{\gamma}_{nh}$, thus

$$\widehat{X}_{n+h}(n) = \sum_{j=1}^{n} c_j X_j = \mathbf{c}_n' \mathbf{X}_n = \boldsymbol{\gamma}_{nh}' \mathbf{\Gamma}_n^{-1} \mathbf{X}_n. \tag{74}$$

It is obvious that $\mathbf{\Gamma}_n = \operatorname{var}(X_1, \ldots, X_n) = \operatorname{var} \mathbf{X}_n = \mathsf{E}\,(\mathbf{X}_n \mathbf{X}_n')$.

The prediction error is

$$\delta_h^2 := \mathsf{E}|X_{n+h} - \widehat{X}_{n+h}(n)|^2 = \|X_{n+h} - \widehat{X}_{n+h}(n)\|^2.$$

By (68)

$$\|X_{n+h}\|^2 = \|\widehat{X}_{n+h}(n)\|^2 + \|X_{n+h} - \widehat{X}_{n+h}(n)\|^2,$$

so that

$$\delta_h^2 = \|X_{n+h}\|^2 - \|\widehat{X}_{n+h}(n)\|^2. \tag{75}$$

For a real-valued centered stationary sequence such that $\mathbf{\Gamma}_n$ is regular,

$$\begin{aligned}
\delta_h^2 &= \|X_{n+h}\|^2 - \|\widehat{X}_{n+h}(n)\|^2 = \mathsf{E}|X_{n+h}|^2 - \mathsf{E}|\widehat{X}_{n+h}(n)|^2 \\
&= R(0) - \mathsf{E}(\mathbf{c}_n' \mathbf{X}_n)^2 = R(0) - \mathbf{c}_n' \mathsf{E}\,(\mathbf{X}_n \mathbf{X}_n') \mathbf{c}_n \\
&= R(0) - \mathbf{c}_n' \mathbf{\Gamma}_n \mathbf{c}_n = R(0) - \boldsymbol{\gamma}_{nh}' \mathbf{\Gamma}_n^{-1} \mathbf{\Gamma}_n \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_{nh} \\
&= R(0) - \boldsymbol{\gamma}_{nh}' \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_{nh}.
\end{aligned} \tag{76}$$

Theorem 53. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be a real-valued centered stationary sequence with autocovariance function R, such that $R(0) > 0$ and $R(k) \to 0$ as $k \to \infty$. Then the matrix $\mathbf{\Gamma}_n = \operatorname{var}(X_1, \ldots, X_n)$ is regular for every $n \in \mathbb{N}$.*

Proof. We will prove the theorem by contradiction: suppose that $\mathbf{\Gamma}_n$ is singular for an $n \in \mathbb{N}$; then there is a nonzero vector $\mathbf{c} = (c_1, \ldots, c_n)'$ such that $\mathbf{c}' \mathbf{\Gamma}_n \mathbf{c} = 0$ and for $\mathbf{X}_n = (X_1, \ldots, X_n)'$, $\mathbf{c}' \mathbf{X}_n = 0$ a. s. holds true, since $\mathsf{E}\, \mathbf{c}' \mathbf{X}_n = 0$ and $\operatorname{var}(\mathbf{c}' \mathbf{X}_n) = \mathbf{c}' \mathbf{\Gamma}_n \mathbf{c} = 0$.

Thus there exists a positive integer $1 \le r < n$ such that $\mathbf{\Gamma}_r$ is regular and $\mathbf{\Gamma}_{r+1}$ is singular, and constants $a_1, \ldots, a_r$ such that

$$X_{r+1} = \sum_{j=1}^{r} a_j X_j.$$

By stationarity of $\{X_t,\, t \in \mathbb{Z}\}$,

$$\operatorname{var}(X_1, \ldots, X_r) = \cdots = \operatorname{var}(X_h, \ldots, X_{h+r-1}) = \mathbf{\Gamma}_r.$$

From here, for any $h \ge 1$,

$$X_{r+h} = \sum_{j=1}^{r} a_j X_{j+h-1}.$$

For every $n \geq r+1$ there exist constants $a_1^{(n)}, \dots, a_r^{(n)}$ such that $X_n = \sum_{j=1}^r a_j^{(n)} X_j = \mathbf{a}^{(n)\prime}\mathbf{X}_r$, where $\mathbf{a}^{(n)} = (a_1^{(n)}, \dots, a_r^{(n)})'$ and $\mathbf{X}_r = (X_1, \dots, X_r)'$,

$$\operatorname{var} X_n = \mathbf{a}^{(n)\prime} \operatorname{var} \mathbf{X}_r \, \mathbf{a}^{(n)} = \mathbf{a}^{(n)\prime}\mathbf{\Gamma}_r \mathbf{a}^{(n)} = R(0) > 0.$$

The matrix $\mathbf{\Gamma}_r$ is positive definite, therefore there exists a decomposition $\mathbf{\Gamma}_r = \mathbf{P\Lambda P'}$, where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of the matrix $\mathbf{\Gamma}_r$ on the diagonal and $\mathbf{PP'} = \mathbf{I}$ is the identity matrix. Since $\mathbf{\Gamma}_r$ is positive definite, all its eigenvalues are positive; w. l. o. g. assume that $0 < \lambda_1 \leq \dots \leq \lambda_r$. Then

$$R(0) = \mathbf{a}^{(n)\prime}\mathbf{P\Lambda P'}\mathbf{a}^{(n)} \geq \lambda_1 \mathbf{a}^{(n)\prime}\mathbf{PP'}\mathbf{a}^{(n)} = \lambda_1 \sum_{j=1}^r \left(a_j^{(n)}\right)^2,$$

from which for every $j = 1, \dots, r$ 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

$$\begin{aligned} 0 < R(0) = \mathsf{E}\big(X_n\big)^2 &= \mathsf{E}\Big(X_n \sum_{j=1}^r a_j^{(n)} X_j\Big) = \sum_{j=1}^r a_j^{(n)} \mathsf{E} X_n X_j \\ &= \sum_{j=1}^r a_j^{(n)} R(n-j) \leq \sum_{j=1}^r |a_j^{(n)}||R(n-j)| \\ &\leq C \sum_{j=1}^r |R(n-j)|. \end{aligned}$$

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 $\mathbf{\Gamma}_n$ is regular for every $n \in \mathbb{N}$. □

Recursive methods

Let us introduce the following notation.

- Denote by $H_1^k = \mathcal{H}\{X_1, \dots, X_k\}$ the Hilbert space generated by $X_1, \dots, X_k$.
- Put $\widehat{X}_1 := 0$ and denote by $\widehat{X}_{k+1}, k \geq 1$, the one-step prediction of X_{k+1}, i. e.,

$$\widehat{X}_{k+1} := \widehat{X}_{k+1}(k) = P_{H_1^k}(X_{k+1}).$$

Then

$$H_1^n = \mathcal{H}\{X_1, \dots, X_n\} = \mathcal{H}\{X_1 - \widehat{X}_1, \dots, X_n - \widehat{X}_n\}.$$

Lemma 4. $X_1 - \widehat{X}_1, \ldots, X_n - \widehat{X}_n$ are orthogonal random variables.

Proof. Let $i < j$. Then $X_i \in H_1^i \subseteq H_1^{j-1}$ and $\widehat{X}_i \in H_1^{i-1} \subset H_1^{j-1}$, so $X_i - \widehat{X}_i \in H_1^{j-1}$. Further, $\widehat{X}_j = P_{H_1^{j-1}}(X_j)$, therefore $X_j - \widehat{X}_j \perp H_1^{j-1}$, and also

$$X_i - \widehat{X}_i \perp X_j - \widehat{X}_j.$$

□

The one-step best linear prediction of X_{k+1} computed from $X_1, \ldots, X_k$ thus can be written in the form

$$\widehat{X}_{k+1} = \sum_{j=1}^{k} \theta_{kj} \left(X_{k+1-j} - \widehat{X}_{k+1-j} \right).$$

The error of the one-step prediction of X_{k+1} is

$$v_k = \mathsf{E}|X_{k+1} - \widehat{X}_{k+1}|^2 = ||X_{k+1} - \widehat{X}_{k+1}||^2, \quad k \geq 0.$$

Theorem 54 (Innovation algorithm)**.** *Let $\{X_t,\, t \in \mathbb{Z}\}$ be a real-valued centered random sequence with autocovariance function $R(i,j)$, such that 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*

$$\begin{aligned} \widehat{X}_1 &= 0, \\ \widehat{X}_{n+1} &= \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \widehat{X}_{n+1-j} \right), \quad n \geq 1, \end{aligned} \tag{77}$$

where for $k = 0, \ldots, n-1$,

$$v_0 = R(1,1), \tag{78}$$

$$\theta_{n,n-k} = \frac{1}{v_k}\Big(R(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j} v_j \Big), \tag{79}$$

$$v_n = R(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \tag{80}$$

Proof. Define $\widehat{X}_1 := 0$, then

$$v_0 = \mathsf{E}|X_1 - \widehat{X}_1|^2 = \mathsf{E}|X_1|^2 = R(1,1).$$

Since $\widehat{X}_{n+1} = P_{H_1^n} X_{n+1}$, it must be of the form as given in (77). When multiply both sides of (77) by $X_{k+1} - \widehat{X}_{k+1}$ for $k < n$ and take the mean value we get

$$\begin{aligned}
&\mathsf{E}\widehat{X}_{n+1}(X_{k+1} - \widehat{X}_{k+1}) \\
&= \sum_{j=1}^{n} \theta_{nj} \mathsf{E}\big(X_{n+1-j} - \widehat{X}_{n+1-j}\big)\big(X_{k+1} - \widehat{X}_{k+1}\big) \\
&= \theta_{n,n-k} \mathsf{E}(X_{k+1} - \widehat{X}_{k+1})^2 = \theta_{n,n-k} v_k.
\end{aligned}$$

Since $\widehat{X}_{n+1} \in H_1^n$ and $X_{n+1} - \widehat{X}_{n+1} \perp H_1^n$ we have

$$\mathsf{E}(X_{n+1} - \widehat{X}_{n+1})(X_{k+1} - \widehat{X}_{k+1}) = 0, \ k < n,$$

$$\mathsf{E}X_{n+1}(X_{k+1} - \widehat{X}_{k+1}) = \mathsf{E}\widehat{X}_{n+1}(X_{k+1} - \widehat{X}_{k+1}) = \theta_{n,n-k} v_k. \tag{81}$$

From here we get

$$\begin{aligned}
\theta_{n,n-k} &= \frac{1}{v_k} \mathsf{E}X_{n+1}\big(X_{k+1} - \widehat{X}_{k+1}\big) \\
&= \frac{1}{v_k}\big(R(n+1, k+1) - \mathsf{E}X_{n+1}\widehat{X}_{k+1}\big).
\end{aligned}$$

Further, applying formula (77) to $\widehat{X}_{k+1}$ and replacing k by $k-j$ in formula (81) we get

$$\begin{aligned}
\mathsf{E}X_{n+1}\widehat{X}_{k+1} &= \mathsf{E}X_{n+1} \sum_{j=1}^{k} \theta_{kj} \left(X_{k+1-j} - \widehat{X}_{k+1-j}\right) \\
&= \sum_{j=1}^{k} \theta_{kj} \mathsf{E}X_{n+1} \left(X_{k+1-j} - \widehat{X}_{k+1-j}\right) \\
&= \sum_{j=1}^{k} \theta_{kj} \theta_{n,n-(k-j)} v_{k-j}.
\end{aligned}$$

Combining these results altogether we get

$$\begin{aligned}
\theta_{n,n-k} &= \frac{1}{v_k}\big(R(n+1, k+1) - \sum_{j=1}^{k} \theta_{kj}\theta_{n,n-(k-j)} v_{k-j}\big) \\
&= \frac{1}{v_k}\big(R(n+1, k+1) - \sum_{\nu=0}^{k-1} \theta_{n,n-\nu}\theta_{k,k-\nu} v_\nu\big).
\end{aligned}$$

Computation of v_n is as follows.

$$\begin{aligned}
v_n &= \mathsf{E}|X_{n+1} - \widehat{X}_{n+1}|^2 = \mathsf{E}|X_{n+1}|^2 - \mathsf{E}|\widehat{X}_{n+1}|^2 \\
&= R(n+1, n+1) - \mathsf{E}\Big|\sum_{j=1}^{n} \theta_{nj}\big(X_{n+1-j} - \widehat{X}_{n+1-j}\big)\Big|^2 \\
&= R(n+1, n+1) - \sum_{j=1}^{n} \theta_{nj}^2 \mathsf{E}(X_{n+1-j} - \widehat{X}_{n+1-j})^2 \\
&= R(n+1, n+1) - \sum_{j=1}^{n} \theta_{nj}^2 v_{n-j} \\
&= R(n+1, n+1) - \sum_{\nu=0}^{n-1} \theta_{n,n-\nu}^2 v_\nu.
\end{aligned}$$

□

Computational scheme of the innovation algorithm:

$$\begin{array}{lllll}
\widehat{X}_1 & v_0 & & & \\
\theta_{11} & \widehat{X}_2 & v_1 & & \\
\theta_{22} & \theta_{21} & \widehat{X}_3 & v_2 & \\
\theta_{33} & \theta_{32} & \theta_{31} & \widehat{X}_4 & v_3 \\
\cdots & & \cdots & & \cdots
\end{array}$$

Example 40. We have observations $X_1, \ldots, X_n$, of an $MA(1)$ random sequence that is generated by

$$X_t = Y_t + bY_{t-1},\ Y_t \sim \mathrm{WN}(0, \sigma^2),\ t \in \mathbb{Z}$$

We will find $\widehat{X}_{n+1}$ by using the innovation algorithm. We get

$$\begin{aligned}
\widehat{X}_1 &= 0, \\
v_0 &= R(0) = \sigma^2(1 + b^2), \\
\theta_{11} &= \frac{1}{v_0} R(1) = \frac{b}{1 + b^2}, \\
\widehat{X}_2 &= \theta_{11}(X_1 - \widehat{X}_1) = \theta_{11} X_1, \\
v_1 &= R(0) - \theta_{11}^2 v_0, \\
\theta_{22} &= \frac{1}{v_0} R(2) = 0, \\
\theta_{21} &= \frac{1}{v_1}(R(1) - \theta_{22}\theta_{11} v_0) = \frac{R(1)}{v_1}, \\
\widehat{X}_3 &= \theta_{21}(X_2 - \widehat{X}_2), \\
v_2 &= R(0) - \theta_{21}^2 v_1,
\end{aligned}$$

generally,

$$\begin{aligned}\theta_{nk} &= 0, \quad k = 2, \dots, n,\\ \theta_{n1} &= \frac{R(1)}{v_{n-1}},\\ \widehat{X}_{n+1} &= \theta_{n1}(X_n - \widehat{X}_n),\\ v_n &= R(0) - \theta_{n1}^2 v_{n-1}.\end{aligned}$$

Example 41. Consider an MA(q) sequence. Then $R(k) = 0$ pro $|k| > q$. By using the recursive computations we get

$$\widehat{X}_{n+1} = \sum_{j=1}^{\min(q,n)} \theta_{nj} \left(X_{n+1-j} - \widehat{X}_{n+1-j}\right), \quad n \geq 1.$$

The coefficients θ_{nj} can be determined again by using Theorem 54.

The h- step prediction, $h > 1$:

Now, we want to use the innovation algorithm to make prediction from $X_1, \dots, X_n$ for $h > 1$ steps ahead, i. e., to determine $\widehat{X}_{n+h}(n)$. Obviously, $\widehat{X}_{n+h}(n) = P_{H_1^n}(X_{n+h})$, where $H_1^n = \mathcal{H}(X_1 - \widehat{X}_1, \dots, X_n - \widehat{X}_n)$. Since $H_1^n \subset H_1^{n+1} \subset \dots \subset H_1^{n+h-1}$, it follows from the properties of the projection mapping and (77), that

$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H_1^n}(X_{n+h}) = P_{H_1^n}\left(P_{H_1^{n+h-1}}(X_{n+h})\right) = P_{H_1^n}(\widehat{X}_{n+h})\\ &= P_{H_1^n}\Big(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j})\Big)\\ &= \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} P_{H_1^n}\left(X_{n+h-j} - \widehat{X}_{n+h-j}\right)\\ &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}\left(X_{n+h-j} - \widehat{X}_{n+h-j}\right), \qquad (82)\end{aligned}$$

since $X_{n+h-j} - \widehat{X}_{n+h-j} \perp X_1^n$ pro $j < h$.

The h-step prediction error is

$$\begin{aligned}\delta_h^2 &= \mathsf{E}\left|X_{n+h} - \widehat{X}_{n+h}(n)\right|^2 = \mathsf{E}\left|X_{n+h}\right|^2 - \mathsf{E}\left|\widehat{X}_{n+h}(n)\right|^2\\ &= R(n+h, n+h) - \mathsf{E}\Big|\sum_{j=h}^{n+h-1} \theta_{n+h-1,j}\left(X_{n+h-j} - \widehat{X}_{n+h-j}\right)\Big|^2\\ &= R(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1}.\end{aligned}$$

Example 42. Le us again consider the MA(1) model from Example 40. We have shown that the one-step prediction is

$$\widehat{X}_{n+1} = P_{H_1^n}(X_{n+1}) = \theta_{n1}(X_n - \widehat{X}_n).$$

For $h > 1$ we have

$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H_1^n}(X_{n+h}) = P_{H_1^n}(\widehat{X}_{n+h}) \\ &= P_{H_1^n}(\theta_{n+h-1,1}(X_{n+h-1} - \widehat{X}_{n+h-1})) = 0,\end{aligned}$$

since $(X_{n+h-1} - \widehat{X}_{n+h-1}) \perp H_1^n$ pro $h > 1$.

Innovation algorithm for an ARMA process

Consider a causal ARMA(p,q) process

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \ t \in \mathbb{Z},$$

$Y_t \sim \mathrm{WN}(0, \sigma^2)$. We want to find $\widehat{X}_{n+1} = P_{H_1^n}(X_{n+1})$.

- First, let us consider the following transformation:

$$W_t = \begin{cases} \frac{1}{\sigma} X_t, & t = 1, 2, \ldots, m, \\ \frac{1}{\sigma}\left(X_t - \varphi_1 X_{t-1} \cdots - \varphi_p X_{t-p}\right), & t > m, \end{cases} \tag{83}$$

  where $m = \max(p,q)$. Put $\widehat{X}_1 = 0, \widehat{W}_1 = 0, \widehat{W}_k = P_{H_1^{k-1}}(W_k)$. It is clear that

$$\begin{aligned} H_1^n &= \mathcal{H}(X_1 - \widehat{X}_1, \ldots, X_n - \widehat{X}_n) = \mathcal{H}(X_1, \ldots, X_n) \\ &= \mathcal{H}(W_1, \ldots, W_n) = \mathcal{H}(W_1 - \widehat{W}_1, \ldots, W_n - \widehat{W}_n). \end{aligned}$$

- Application of the innovation algorithm to the sequence $\{W_1, \ldots, W_n\}$ gives

$$\widehat{W}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(W_{n+1-j} - \widehat{W}_{n+1-j}), & 1 \le n < m, \\ \sum_{j=1}^q \theta_{nj}(W_{n+1-j} - \widehat{W}_{n+1-j}), & n \ge m \end{cases} \tag{84}$$

  (for $t > m$, $W_t \sim \mathrm{MA}(q)$).

- Application of the projection mapping onto H_1^{t-1} to both sides of (83) results in

$$\widehat{W}_t = \begin{cases} \frac{1}{\sigma}\widehat{X}_t, & t \le m, \\ \frac{1}{\sigma}(\widehat{X}_t - \varphi_1 X_{t-1} \cdots - \varphi_p X_{t-p}), & t > m. \end{cases} \tag{85}$$

We can see that for $t \geq 1$,

$$W_t - \widehat{W}_t = \frac{1}{\sigma}\big(X_t - \widehat{X}_t\big), \qquad \mathsf{E}|W_t - \widehat{W}_t|^2 = \frac{1}{\sigma^2} v_{t-1} := w_{t-1}.$$

Therefore it holds

$$\widehat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}\big(X_{n+1-j} - \widehat{X}_{n+1-j}\big), & n < m \\ \sum_{j=1}^{q} \theta_{nj}\big(X_{n+1-j} - \widehat{X}_{n+1-j}\big) + \sum_{j=1}^{p} \varphi_j X_{n-j+1}, & n \geq m, \end{cases} \tag{86}$$

where coefficients θ_{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 $\mathsf{E}\, X_t = 0$, thus $\mathsf{E}\, W_t = 0$. For the covariances $R_W(s,t) = \mathsf{E} W_s W_t$ we get

$$R_W(s,t) = \begin{cases} \frac{1}{\sigma^2} R_X(s-t), \quad 1 \leq s,t \leq m, \\ \frac{1}{\sigma^2}\big[R_X(s-t) - \sum_{j=1}^{p} \varphi_j R_X(|s-t|-j)\big], \\ \qquad\qquad \min(s,t) \leq m, \quad m < \max(s,t) \leq 2m, \\ \frac{1}{\sigma^2} \sum_{j=0}^{q-|s-t|} \theta_j \theta_{j+|s-t|}, \quad s,t > m,\ |s-t| \leq q, \\ 0, \qquad \text{elsewhere} \end{cases} \tag{87}$$

(we have put $\theta_0 = 1$).

Innovation algorithm for an AR sequence

Let us consider a causal AR(p) process, i,e.,

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t,\ t \in \mathbb{Z},\ Y_t \sim \mathrm{WN}(0,\sigma^2)$$

- Transformation:

$$W_t = \begin{cases} \frac{1}{\sigma} X_t, & 1 \leq t \leq p, \\ \frac{1}{\sigma}(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}) = \frac{1}{\sigma} Y_t, & t > p. \end{cases} \tag{88}$$

- Innovation algorithm applied to $W_1, \ldots, W_n$:

$$\widehat{W}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}(W_{n+1-j} - \widehat{W}_{n+1-j}), & n < p, \\ 0, & n \geq p. \end{cases} \tag{89}$$

($W_{n+1} \perp H_1^n$ for $n \geq p$.)

Again, $W_t - \widehat{W}_t = \frac{1}{\sigma}(X_t - \widehat{X}_t)$ for $t \geq 1$ and from here

$$\widehat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}\big(X_{n+1-j} - \widehat{X}_{n+1-j}\big), & n < p, \\ \varphi_1 X_n + \varphi_2 X_{n-1} + \cdots + \varphi_p X_{n-p+1}, & n \geq p. \end{cases} \tag{90}$$

The autocovariance function needed for the calculation of the coefficients θ_{nj} is

$$R_W(s,t) = \begin{cases} \frac{1}{\sigma^2} R_X(s-t), & 1 \leq s, t \leq p, \\ 1, & t = s > p, \\ 0 & \text{elsewhere.} \end{cases} \tag{91}$$

The one-step prediction error for $n \geq p$ is

$$v_n = \mathsf{E}|X_{n+1} - \widehat{X}_{n+1}|^2 = \mathsf{E}Y_{n+1}^2 = \sigma^2.$$

10.3 Prediction from infinite history

Let us suppose that we know the history $X_n, X_{n-1}, \ldots$, and we want to forecast $X_{n+1}, X_{n+2}, \ldots$. We will solve this problem again by using projection in Hilbert spaces.

Consider Hilbert spaces $H = \mathcal{H}\{X_t,\, t \in \mathbb{Z}\}$ and $H_{-\infty}^n = \mathcal{H}\{\ldots X_{n-1}, X_n\}$. Then the best linear prediction $\widehat{X}_{n+h}(n)$ of X_{n+h} from the infinite history $X_n, X_{n-1}, \ldots$ is the projection of $X_{n+h} \in H$ onto $H_{-\infty}^n$, i.,e., $\widehat{X}_{n+h}(n) = P_{H_{-\infty}^n}(X_{n+h})$. The one-step prediction is again denoted by $\widehat{X}_{n+1}(n) := \widehat{X}_{n+1}$.

Prediction in a causal AR(p) process

Consider model

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t, \qquad t \in \mathbb{Z}, \tag{92}$$

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: To get $\widehat{X}_{n+1}(n)$ from $X_n, X_{n-1}, \ldots$, notice that

- $X_{n+1} = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + Y_{n+1}$
- $\varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} \in H_{-\infty}^n$,

- $Y_{n+1} \perp X_n, X_{n-1}, \dots \Rightarrow Y_{n+1} \perp H_{-\infty}^n$, (from the linearity and the continuity of the inner product)

It means that
$$\widehat{X}_{n+1} = P_{H_{-\infty}^n}(X_{n+1}) = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p}.$$
The prediction error is
$$\mathsf{E}|X_{n+1} - \widehat{X}_{n+1}|^2 = \mathsf{E}|Y_{n+1}|^2 = \sigma^2.$$

The h-step prediction, $h > 1$:
$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H_{-\infty}^n}(X_{n+h}) = P_{H_{-\infty}^n}\left(P_{H_{-\infty}^{n+h-1}}(X_{n+h})\right) \\ &= P_{H_{-\infty}^n}(\widehat{X}_{n+h}) \\ &= P_{H_{-\infty}^n}\left(\varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p}\right) \\ &= \varphi_1[X_{n+h-1}] + \varphi_2[X_{n+h-2}] + \cdots + \varphi_p[X_{n+h-p}],\end{aligned}$$
where
$$[X_{n+j}] = \begin{cases} X_{n+j}, & j \le 0 \\ \widehat{X}_{n+j}(n), & j > 0. \end{cases}$$

Example 43. Consider an AR(1) process generated by $X_t = \varphi X_{t-1} + Y_t$, where $|\varphi| < 1$ and $Y_t \sim \mathrm{WN}(0, \sigma^2)$. If we know the whole history $X_n, X_{n-1}, \dots$, we have $\widehat{X}_{n+1} = \varphi X_n$.

For $h > 1$
$$\begin{aligned}\widehat{X}_{n+h}(n) &= \varphi[X_{n+h-1}] = \varphi \widehat{X}_{n+h-1}(n) = \varphi^2 \widehat{X}_{n+h-2}(n) = \dots \\ &= \varphi^h X_n.\end{aligned}$$

The prediction error is
$$\begin{aligned}\mathsf{E}|X_{n+h} - \widehat{X}_{n+h}|^2 &= \mathsf{E}|X_{n+h}|^2 - \mathsf{E}|\widehat{X}_{n+h}(n)|^2 \\ &= R_X(0) - \mathsf{E}\left|\varphi^h X_n\right|^2 = R_X(0)\left(1 - \varphi^{2h}\right) \\ &= \sigma^2 \frac{1 - \varphi^{2h}}{1 - \varphi^2}.\end{aligned}$$

Prediction in a causal and invertible ARMA(p, q) process

Consider a causal and invertible process
$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \quad t \in \mathbb{Z}, \tag{93}$$
where $Y_t \sim \mathrm{WN}(0, \sigma^2)$.

- Due to 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 here it follows that $Y_t \perp X_s$ for every $s < t$.
- Due to 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,

$$X_t = -\sum_{j=1}^{\infty} d_j X_{t-j} + Y_t. \tag{94}$$

Since

$$-\sum_{j=1}^{\infty} d_j X_{t-j} = \mathrm{l.\,i.\,m.}_{\ N\to\infty}\Big(-\sum_{j=1}^{N} d_j X_{t-j}\Big) \in H_{-\infty}^{t-1},$$

and $Y_t \perp H_{-\infty}^{t-1}$, from decomposition (94) it follows that the best linear prediction of X_{n+1} based on the whole history $X_n, X_{n-1}, \dots$, is

$$\widehat{X}_{n+1} = -\sum_{j=1}^{\infty} d_j X_{n+1-j}. \tag{95}$$

The prediction error is

$$\mathsf{E}|X_{n+1} - \widehat{X}_{n+1}|^2 = \mathsf{E}|Y_{n+1}|^2 = \sigma^2.$$

From the uniqueness of the decomposition $X_{n+1} = \widehat{X}_{n+1} + Y_{n+1}$ and from formula (93) we can also see that

$$\widehat{X}_{n+1} = \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} + \theta_1 Y_n + \cdots + \theta_q Y_{n+1-q},$$

thus, if we use the relation $Y_t = X_t - \widehat{X}_t$, (noticing that $\widehat{X}_t = P_{H_{-\infty}^{t-1}} X_t$) we have

$$\begin{aligned}\widehat{X}_{n+1} &= \varphi_1 X_n + \cdots + \varphi_p X_{n+1-p} \\ &\quad + \theta_1\big(X_n - \widehat{X}_n\big) + \cdots + \theta_q\big(X_{n+1-q} - \widehat{X}_{n+1-q}\big).\end{aligned}$$

The h-step prediction for $h > 1$ is

$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H_{-\infty}^{n}}(X_{n+h}) = P_{H_{-\infty}^{n}}(P_{H_{-\infty}^{n+h-1}}(X_{n+h})) = P_{H_{-\infty}^{n}}(\widehat{X}_{n+h}) \\ &= P_{H_{-\infty}^{n}}\big(\varphi_1 X_{n+h-1} + \cdots + \varphi_p X_{n+h-p} \\ &+ \theta_1 Y_{n+h-1} + \cdots + \theta_q Y_{n+h-q}\big) \\ &= \varphi_1[X_{n+h-1}] + \cdots + \varphi_p[X_{n+h-p}] \\ &+ \theta_1[Y_{n+h-1}] + \cdots + \theta_q[Y_{n+h-q}],\end{aligned}$$

where

$$[X_{n+j}] = \begin{cases} X_{n+j}, & j \le 0, \\ \widehat{X}_{n+j}(n), & j > 0 \end{cases}$$

and

$$[Y_{n+j}] = \begin{cases} X_{n+j} - \widehat{X}_{n+j}, & j \le 0, \\ 0, & j > 0. \end{cases}$$

If we use (95) we have

$$\widehat{X}_{n+h}(n) = P_{H^n_{-\infty}} \left(-\sum_{j=1}^{\infty} d_j X_{n+h-j} \right) = -\sum_{d=1}^{\infty} [X_{n-h+j}].$$

Alternatively, from the causality we get

$$\widehat{X}_{n+h}(n) = P_{H^n_{-\infty}}(X_{n+h}) = P_{H^n_{-\infty}}\Big(\sum_{j=0}^{\infty} c_j Y_{n+h-j}\Big).$$

Then, from properties of the projection mapping,

$$P_{H^n_{-\infty}}\Big(\sum_{j=0}^{\infty} c_j Y_{n+h-j}\Big) = \sum_{j=0}^{\infty} c_j P_{H^n_{-\infty}}(Y_{n+h-j}),$$

and thus we can express the $h-$ step prediction also as

$$\widehat{X}_{n+h}(n) = \sum_{j=h}^{\infty} c_j Y_{n+h-j}.$$

The prediction error can be then easily computed by

$$\mathsf{E}\big|X_{n+h} - \widehat{X}_{n+h}(n)\big|^2 = \mathsf{E}\Big|\sum_{j=0}^{h-1} c_j Y_{n+h-j}\Big|^2 = \sigma^2 \sum_{j=0}^{h-1} |c_j|^2.$$

Example 44. Consider the MA(1) model

$$X_t = Y_t + \theta Y_{t-1},\ t \in \mathbb{Z}, \quad Y_t \sim \mathrm{WN}(0, \sigma^2), \quad |\theta| < 1.$$

In this case, $\{X_t,\ t \in \mathbb{Z}\}$ is invertible, $Y_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$, and the best linear prediction is

$$\widehat{X}_{n+1}(n) = \widehat{X}_{n+1} = -\sum_{j=1}^{\infty} (-\theta)^j X_{n+1-j} = \theta Y_n = \theta(X_n - \widehat{X}_n).$$

The prediction error is $\mathsf{E}|X_{n+1} - \widehat{X}_{n+1}|^2 = \mathsf{E}Y_{n+1}^2 = \sigma^2$.

For $h > 1$,

$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H^n_{-\infty}}\big(P_{H^{n+h-1}_{-\infty}}(X_{n+h})\big) = P_{H^n_{-\infty}}(\widehat{X}_{n+h})\\ &= \theta P_{H^n_{-\infty}}(Y_{n+h-1}) = 0,\end{aligned}$$

since for $h \geq 2$, $Y_{n+h-1} \perp H^n_{-\infty}$. The h-step prediction error is

$$\mathsf{E}|X_{n+h} - \widehat{X}_{n+h}(n)|^2 = \mathsf{E}|X_{n+h}|^2 = R_X(0) = \sigma^2(1+\theta^2).$$

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 $\widehat{X}_{n+h}(n-1) = P_{H^{n-1}_{-\infty}}(X_{n+h})$, in other words, we want to find $\widehat{X}_{n+h}(n-1) \in H^{n-1}_{-\infty} \subset \mathcal{H}\{X_t,\, t \in \mathbb{Z}\}$, such that $X_{n+h} - \widehat{X}_{n+h}(n-1) \perp H^{n-1}_{-\infty}$.
- 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 $\widehat{X}_{n+h}(n-1)$ should be of the form

$$\widehat{X}_{n+h}(n-1) = \int_{-\pi}^{\pi} e^{in\lambda}\Phi_h(\lambda) dZ(\lambda), \tag{96}$$

where $\Phi_h(\lambda) \in L_2(F)$. Condition $X_{n+h} - \widehat{X}_{n+h}(n-1) \perp H^{n-1}_{-\infty}$ will be met if

$$X_{n+h} - \widehat{X}_{n+h}(n-1) \perp X_{n-j}, \quad j = 1, 2, \ldots$$

thus for $j = 1, 2, \ldots$,

$$\mathsf{E}\big(X_{n+h} - \widehat{X}_{n+h}(n-1)\big)\overline{X}_{n-j} = 0,$$

or

$$
\begin{aligned}
&\mathsf{E}\left(X_{n+h} - \widehat{X}_{n+h}(n-1)\right)\overline{X}_{n-j} = \\
&= R(h+j) - \mathsf{E}\int_{-\pi}^{\pi} e^{in\lambda}\Phi_h(\lambda)dZ(\lambda)\overline{\int_{-\pi}^{\pi} e^{i(n-j)\lambda}dZ(\lambda)} \\
&= R(h+j) - \int_{-\pi}^{\pi} e^{in\lambda}\Phi_h(\lambda)e^{-i(n-j)\lambda}dF(\lambda) \\
&= \int_{-\pi}^{\pi} e^{i(h+j)\lambda}dF(\lambda) - \int_{-\pi}^{\pi} e^{ij\lambda}\Phi_h(\lambda)dF(\lambda) \\
&= \int_{-\pi}^{\pi} e^{i(h+j)\lambda}f(\lambda)d\lambda - \int_{-\pi}^{\pi} e^{ij\lambda}\Phi_h(\lambda)f(\lambda)d\lambda \\
&= \int_{-\pi}^{\pi} e^{ij\lambda}\left(e^{ih\lambda} - \Phi_h(\lambda)\right)f(\lambda)d\lambda = 0. \qquad (97)
\end{aligned}
$$

Denote

$$\Psi_h(\lambda) := \left(e^{ih\lambda} - \Phi_h(\lambda)\right)f(\lambda).$$

Then (97) can be written as

$$\int_{-\pi}^{\pi} e^{ij\lambda}\Psi_h(\lambda)d\lambda = 0, \quad j = 1, 2, \ldots. \qquad (98)$$

It follows from condition (98) that the Fourier expansion of the function Ψ_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 $L_2(F)$,

$$
\begin{aligned}
\widehat{X}_{n+h}(n-1) &= \int_{-\pi}^{\pi} e^{in\lambda}\Big[\sum_{k=1}^{\infty} a_k e^{-ik\lambda}\Big] dZ(\lambda)\\
&= \underset{N\to\infty}{\mathrm{l.\,i.\,m.}} \int_{-\pi}^{\pi} e^{in\lambda}\Big[\sum_{k=1}^{N} a_k e^{-ik\lambda}\Big] dZ(\lambda)\\
&= \underset{N\to\infty}{\mathrm{l.\,i.\,m.}} \sum_{k=1}^{N} a_k \Big[\int_{-\pi}^{\pi} e^{i(n-k)\lambda} dZ(\lambda)\Big]\\
&= \underset{N\to\infty}{\mathrm{l.\,i.\,m.}} \sum_{k=1}^{N} a_k X_{n-k} = \sum_{k=1}^{\infty} a_k X_{n-k}.
\end{aligned}
$$

Theorem 55. *Let $\{X_t,\, t \in \mathbb{Z}\}$ be a real-valued centered stationary random sequence with the autocovariance function R and the spectral density $f(\lambda) = f^*(e^{i\lambda})$, where f^* is a rational function of a complex-valued variable.*
Let Φ_h^* be a function of a complex-valued variable z holomorphic for $|z| \geq 1$ and such that $\Phi_h^*(\infty) = 0$.
Let

$$\Psi_h^*(z) = \left(z^h - \Phi_h^*(z)\right) f^*(z), \ \ z \in \mathbb{C},$$

be a function holomorphic for $|z| \leq 1$. Then the best linear prediction of X_{n+h} from $X_{n-1}, X_{n-2}, \dots$ is

$$\widehat{X}_{n+h}(n-1) = \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) dZ(\lambda),$$

where $\Phi_h(\lambda) = \Phi_h^*\left(e^{i\lambda}\right)$ 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

$$
\begin{aligned}
\delta_h^2 &= \mathsf{E}|X_{n+h} - \widehat{X}_{n+h}(n-1)|^2\\
&= R(0) - \int_{-\pi}^{\pi} |\Phi_h(\lambda)|^2 f(\lambda) d(\lambda) \qquad (99)\\
&= R(0) - \int_{-\pi}^{\pi} e^{-ih\lambda} \Phi_h(\lambda) f(\lambda) d\lambda. \qquad (100)
\end{aligned}
$$

Proof. Anděl (1976), Chap. X, Theorem 8. □

The function Φ_h is called to be *spectral characteristic of prediction* of X_{n+h} from $X_{n-1}, X_{n-2}, \dots$

Example 45. Consider again the autoregressive sequence

$$X_t = \varphi X_{t-1} + Y_t,\ |\varphi| < 1,\ \varphi \neq 0, Y_t \sim \text{WN}(0, \sigma^2).$$

We want to determine prediction $\widehat{X}_{n+h}(n-1), h \geq 0$ on the basis of $X_{n-1}, X_{n-2}, \ldots$, in the spectral domain.

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

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1-\varphi e^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \frac{1}{(1-\varphi e^{-i\lambda})(1-\varphi e^{i\lambda})} = f^*\left(e^{i\lambda}\right),$$

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

$$\Psi_h^*(z) = \left(z^h - \Phi_h^*(z)\right) f^*(z) = \frac{\sigma^2}{2\pi} \frac{z(z^h - \Phi_h^*(z))}{(1-\varphi z)(z-\varphi)},\ z \in \mathbb{C},$$

should be holomorphic for $|z| \geq 1$. Since $|\varphi| < 1$, it must be $z^h - \Phi_h^*(z) = 0$ for $z = \varphi$, otherwise Ψ_h^* has a pole at $z = \varphi$.

Thus,

$$\Phi_h^*(\varphi) = \varphi^h. \tag{101}$$

Put $\Phi_h^*(z) := \frac{\gamma}{z}$, where γ is a constant. Then Φ_h^* is holomorphic for $|z| \geq 1$ and $\Phi_h^*(\infty) = 0$. The function Ψ_h^* is holomorphic for $|z| \leq 1$.

The value of constant γ follows from (101): $\varphi^h = \Phi_h^*(\varphi) = \frac{\gamma}{\varphi}$, thus $\gamma = \varphi^{h+1}$.

Functions

$$\begin{aligned} \Phi_h^*(z) &= \frac{\varphi^{h+1}}{z} = \varphi^{h+1} z^{-1},\ z \in \mathbb{C}, \\ \Psi_h^*(z) &= \frac{\sigma^2}{2\pi} \frac{z^{h+1} - \varphi^{h+1}}{(z-\varphi)(1-\varphi z)},\ z \in \mathbb{C}, \end{aligned}$$

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

$$\begin{aligned} \widehat{X}_{n+h}(n-1) &= \int_{-\pi}^{\pi} e^{in\lambda} \Phi_h(\lambda) dZ(\lambda) \\ &= \int_{-\pi}^{\pi} e^{in\lambda} \varphi^{h+1} e^{-i\lambda} dZ(\lambda) \\ &= \int_{-\pi}^{\pi} e^{i(n-1)\lambda} dZ(\lambda) \varphi^{h+1} = \varphi^{h+1} X_{n-1} \end{aligned}$$

which is the same result as obtained in the time domain.

For the prediction error, from (99) we get

$$\begin{aligned}\delta_h^2 &= \mathsf{E}\big|X_{t+h} - \widehat{X}_{t+h}(t-1)\big|^2 = \|X_{t+h}\|^2 - \|\widehat{X}_{t+h}(t-1)\|^2\\ &= R(0) - \mathsf{E}\big|\widehat{X}_{t+h}(t-1)\big|^2 = R(0) - \mathsf{E}\left|\int_{-\pi}^{\pi} e^{it\lambda}\Phi_h(\lambda)dZ(\lambda)\right|^2\\ &= R(0) - \int_{-\pi}^{\pi} \big|e^{it\lambda}\Phi_h(\lambda)\big|^2 f(\lambda)d\lambda =\\ &= R(0) - |\varphi|^{2(h+1)}\int_{-\pi}^{\pi} f(\lambda)d\lambda = R(0)\left(1-\varphi^{2(h+1)}\right),\end{aligned}$$

which is again in accordance with the result in the time domain.

12 Filtration of signal and noise

Let us consider a sequence $\{X_t,\, t\in\mathbb{Z}\}$, said 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 extract 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 $V_t = X_t + Y_t$ for $t\in\mathbb{Z}$. 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 $V_1,\dots,V_n$ to be known observations.

On the basis of $V_1,\dots,V_n$ we want to find the best linear approximation of X_s in the form $\widehat{X}_s = \sum_{j=1}^n c_jV_j$, with coefficients $c_1,\dots,c_n$ that minimize the mean square error $\mathsf{E}|X_s - \widehat{X}_s|^2$.

Denote $H_1^n = \mathcal{H}\{V_1,\dots,V_n\} \subset L_2(\Omega,\mathcal{A},P)$. Then the best linear approximation $\widehat{X}_s$ of X_s is the projection of $X_s \in L_2(\Omega,\mathcal{A},P)$ onto H_1^n, i.e., $\widehat{X}_s \in H_1^n$ and $X_s - \widehat{X}_s \perp H_1^n$. Since $H_1^n = \mathcal{H}\{V_1,\dots,V_n\} = \mathcal{M}\{V_1,\dots,V_n\}$, it suffices to find constants $c_1,\dots,c_n$ such that

$$\widehat{X}_s = \sum_{j=1}^n c_jV_j$$

and

$$X_s - \widehat{X}_s \perp V_t,\ t = 1,\dots,n,$$

or

$$\mathsf{E}(X_s - \widehat{X}_s)V_t = 0,\ t = 1, \ldots, n. \tag{102}$$

Since $V_t = X_t + Y_t$ for all t and X_t, Y_t are uncorrelated, we can see that $\mathsf{E}X_sV_t = \mathsf{E}X_sX_t = R_X(s-t)$, and equations (102) can be written in the form

$$R_X(s-t) - \sum_{j=1}^{n} c_j R_V(j-t) = 0,\ t = 1, \ldots, n. \tag{103}$$

The variable $\widehat{X}_s$ 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

$$\begin{aligned}\delta^2 &= \mathsf{E}|X_s - \widehat{X}_s|^2 = \|X_s - \widehat{X}_s\|^2 = \|X_s\|^2 - \|\widehat{X}_s\|^2 \\ &= R_X(0) - \mathsf{E}\Big|\sum_{j=1}^{n} c_j V_j\Big|^2 = R_X(0) - \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k R_V(j-k).\end{aligned}$$

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 $t \in \mathbb{Z}$. Then the best linear filtration of X_s from $\{V_t,\, t \in \mathbb{Z}\}$ is*

$$\widehat{X}_s = \int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda)\,dZ_V(\lambda),$$

where

$$\Phi(\lambda) = \frac{f_X(\lambda)}{f_V(\lambda)}, \qquad \lambda \in [-\pi, \pi], \tag{104}$$

$f_V = f_X + f_Y$ is the spectral density of $\{V_t,\, t \in \mathbb{Z}\}$ and $Z_V = \{Z_\lambda,\, \lambda \in [-\pi, \pi]\}$ is the orthogonal increment process from the spectral decomposition of the sequence $\{V_t,\, t \in \mathbb{Z}\}$. The filtration error is

$$\delta^2 = \int_{-\pi}^{\pi} \frac{f_X(\lambda)f_Y(\lambda)}{f_X(\lambda) + f_Y(\lambda)}\,d\lambda = \int_{-\pi}^{\pi} \Phi(\lambda)f_Y(\lambda)\,d\lambda.$$

Function Φ is called *spectral characteristic of filtration.*

Remark 18. Notice that if $\Phi(\lambda) = \sum_{k=-\infty}^{\infty} a_k e^{ik\lambda}$, where $\sum_{k=-\infty}^{\infty} |a_k| < \infty$, then $\widehat{X}_s = \sum_{k=-\infty}^{\infty} a_k V_{s-k}$.

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 $f_V = f_X + f_Y$.

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 $\widehat{X}_s = P_H(X_s)$.

Let Φ be the function defined in (104). First, we will show that

$$\widehat{X}_s := \int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda) dZ_V(\lambda) \in H.$$

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 a f_V are continuous functions and f_V takes in $[-\pi, \pi]$ positive values, only. Thus,

$$\int_{-\pi}^{\pi} |\Phi(\lambda)|^2 dF_V(\lambda) = \int_{-\pi}^{\pi} \left|\frac{f_X(\lambda)}{f_V(\lambda)}\right|^2 f_V(\lambda) d\lambda = \int_{-\pi}^{\pi} \frac{|f_X(\lambda)|^2}{f_V(\lambda)} d\lambda < \infty$$

and $\widehat{X}_s \in H$.

Further, $\widehat{X}_s$ will be the projection of X_s onto H if $(X_s - \widehat{X}_s) \perp H$, i.e., if $(X_s - \widehat{X}_s) \perp V_t$ for all $t \in \mathbb{Z}$. For any $t \in \mathbb{Z}$ we have

$$\begin{aligned}
\mathsf{E}\left(X_s - \widehat{X}_s\right)\overline{V}_t &= \mathsf{E}X_s\overline{V}_t - \mathsf{E}\widehat{X}_s\overline{V}_t \\
&= \mathsf{E}X_s\left(\overline{X}_t + \overline{Y}_t\right) - \mathsf{E}\left(\int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda) dZ_V(\lambda)\overline{\int_{-\pi}^{\pi} e^{it\lambda} dZ_V(\lambda)}\right) \\
&= \mathsf{E}X_s\overline{X}_t - \int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda) e^{-it\lambda} dF_V(\lambda) \\
&= R_X(s-t) - \int_{-\pi}^{\pi} e^{i\lambda(s-t)}\Phi(\lambda) f_V(\lambda) d\lambda \\
&= R_X(s-t) - \int_{-\pi}^{\pi} e^{i\lambda(s-t)} f_X(\lambda) d\lambda \\
&= R_X(s-t) - R_X(s-t) = 0.
\end{aligned}$$

We have proved that $\widehat{X}_s$ is the best linear filtration.

Let us determine the filtration error:

$$\begin{aligned}
\delta^2 &= \|X_s\|^2 - \|\widehat{X}_s\|^2 = R_X(0) - \mathsf{E}|\widehat{X}_s|^2 \\
&= \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \mathsf{E}\Big|\int_{-\pi}^{\pi} e^{is\lambda}\Phi(\lambda)dZ_V(\lambda)\Big|^2 \\
&= \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \int_{-\pi}^{\pi} \big|\Phi(\lambda)\big|^2 dF_V(\lambda) \\
&= \int_{-\pi}^{\pi} f_X(\lambda)d\lambda - \int_{-\pi}^{\pi} \Big|\frac{f_X(\lambda)}{f_V(\lambda)}\Big|^2 f_V(\lambda)d\lambda \\
&= \int_{-\pi}^{\pi} \frac{f_X(\lambda)f_Y(\lambda)}{f_X(\lambda)+f_Y(\lambda)}d\lambda = \int_{-\pi}^{\pi} \Phi(\lambda)f_Y(\lambda)d\lambda.
\end{aligned}$$

□

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 σ_W^2, and $\{Y_t,\, t \in \mathbb{Z}\}$ is another white noise sequence with zero mean and variance σ_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

$$f_V(\lambda) = \frac{\sigma^2}{2\pi}\frac{|1-\theta e^{-i\lambda}|^2}{|1-\varphi e^{-i\lambda}|^2}, \quad \lambda \in [-\pi,\pi], \tag{105}$$

where $\sigma^2 = \frac{\varphi}{\theta}\sigma_Y^2,\ \theta$ 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 φ, and

$$c = \frac{\sigma_W^2 + \sigma_Y^2(1+\varphi^2)}{\varphi\sigma_Y^2}.$$

(See Prášková, 2016, Problem 8.1 for some hints.) Then

$$\begin{aligned}
\Phi(\lambda) = \frac{f_X(\lambda)}{f_V(\lambda)} &= \frac{\sigma_W^2}{\sigma^2}\frac{1}{|1-\theta e^{-i\lambda}|^2} = \frac{\sigma_W^2}{\sigma^2}\Big|\sum_{k=0}^{\infty}\theta^k e^{-ik\lambda}\Big|^2 \\
&= \frac{\sigma_W^2}{\sigma^2}\frac{1}{1-\theta^2}\sum_{k=-\infty}^{\infty}\theta^{|k|}e^{-ik\lambda}
\end{aligned}$$

for all $\lambda \in [-\pi, \pi]$.

The best linear filtration of X_s from $\{V_t,\, t \in \mathbb{Z}\}$ is

$$\widehat{X}_s = \frac{\sigma_W^2}{\sigma^2}\frac{1}{1-\theta^2}\sum_{k=-\infty}^{\infty}\theta^{|k|}V_{s-k}. \tag{106}$$

The filtration error is

$$\begin{aligned}\delta^2 = \mathsf{E}|X_s - \widehat{X}_s|^2 &= \int_{-\pi}^{\pi}\phi(\lambda)f_Y(\lambda)d\lambda \\ &= \frac{\sigma_Y^2}{2\pi}\frac{\sigma_W^2}{\sigma^2}\int_{-\pi}^{\pi}\frac{1}{|1-\theta e^{-i\lambda}|^2}d\lambda = \frac{\sigma_Y^2\sigma_W^2}{\sigma^2}\frac{1}{1-\theta^2}.\end{aligned}$$

Remark 19. It follows from (105) that f_V has the same form as the spectral density of an ARMA(1,1) sequence. The mixture of the AR(1) sequence $\{X_t,\, t \in \mathbb{Z}\}$ and the 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 to be

$$Z_t - \varphi Z_{t-1} = U_t - \theta U_{t-1},\ t \in \mathbb{Z},$$

where $\varphi \neq 0,\ |\varphi| < 1,\ |\theta| < 1$ and $\{U_t,\, t \in \mathbb{Z}\}$ is a white noise with the variance $\sigma^2 = \frac{\varphi}{\theta}\sigma_Y^2$. Parameter θ can be determined as given above.

Remark 20. Function Φ is the transfer function of the linear filter $\{\frac{\sigma_W^2}{\sigma^2}\frac{1}{1-\theta^2}\theta^{|k|},\ k \in \mathbb{Z}\}$.

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

$$\alpha(k) = \begin{cases} \operatorname{corr}(X_1, X_{k+1}) = \frac{\operatorname{cov}(X_1,X_2)}{\sqrt{\operatorname{var}X_1}\sqrt{\operatorname{var}X_2}}, & k = 1, \\ \operatorname{corr}\left(X_1 - \widetilde{X}_1, X_{k+1} - \widetilde{X}_{k+1}\right), & k > 1, \end{cases}$$

where $\widetilde{X}_1$ is the linear projection of X_1 onto Hilbert space $H_2^k = \mathcal{H}\{X_2, \ldots, X_k\}$ and $\widetilde{X}_{k+1}$ is the linear projection of X_{k+1} onto H_2^k.

From the properties of the projection mapping it follows that $\widetilde{X}_1 = c_2X_2 + \cdots + c_kX_k$ where constants $c_2, \ldots, c_k$ are determined by conditions $\mathsf{E}(X_1 - \widetilde{X}_1)X_j = 0,\ j = 2, \ldots, k$. The same holds for $\widetilde{X}_{k+1}$. We can see that $\alpha(k)$ represents the correlation coefficient of residuals $X_1 - \widetilde{X}_1$ and $X_{k+1} - \widetilde{X}_{k+1}$ of the best linear approximation of the variables X_1 and X_{k+1} by random variables $X_2, \dots, X_k$.

The stationarity of the sequence $\{X_t,\, t \in \mathbb{Z}\}$ implies that for $h \in \mathbb{N}$, $\operatorname{corr}(X_1 - \widetilde{X}_1, X_{k+1} - \widetilde{X}_{k+1}) = \operatorname{corr}(X_h - \widetilde{X}_h, X_{k+h} - \widetilde{X}_{k+h})$, where $\widetilde{X}_h, \widetilde{X}_{k+h}$ are linear projections of random variables X_h, X_{k+h} onto the Hilbert space $\mathcal{H}\{X_{h+1}, \dots, X_{h+k-1}\}$. Therefore, $\alpha(k)$ is also the correlation coefficient of X_h and X_{h+k} after the linear dependence $X_{h+1}, \dots, X_{h+k-1}$ was eliminated.

Example 47. Consider the causal AR(1) process

$$X_t = \varphi X_{t-1} + Y_t,$$

where $|\varphi| < 1$ and $Y_t \sim \mathrm{WN}(0, \sigma^2)$.

According to the definition, $\alpha(1) = \operatorname{corr}(X_1, X_2) = r_X(1) = \varphi$. For $k > 1$,

$$\alpha(k) = \frac{\mathsf{E}(X_1 - \widetilde{X}_1)(X_{k+1} - \widetilde{X}_{k+1})}{\sqrt{\mathsf{E}(X_1 - \widetilde{X}_1)^2}\sqrt{\mathsf{E}(X_{k+1} - \widetilde{X}_{k+1})^2}}.$$

Due to the causality, $\widetilde{X}_{k+1} = P_{H_2^k}(X_{k+1}) = \varphi X_k$ and $X_{k+1} - \widetilde{X}_{k+1} = Y_{k+1} \perp H_2^k$. Further, it follows from causality that $Y_{k+1} \perp X_1$, thus $\mathsf{E}(X_1 - \widetilde{X}_1)(X_{k+1} - \widetilde{X}_{k+1}) = \mathsf{E}(X_1 - \widetilde{X}_1)Y_{k+1} = 0$, from which we conclude that $\alpha(k) = 0$ for $k > 1$.

Remark 21. 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 the MA(1) process

$$X_t = Y_t + bY_{t-1},$$

where $|b| < 1$ and $Y_t \sim \mathrm{WN}(0, \sigma^2)$. 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, $\alpha(2) = \operatorname{corr}(X_1 - \widetilde{X}_1, X_3 - \widetilde{X}_3)$. To determine $\widetilde{X}_1$, notice that $\widetilde{X}_1 = P_{H_2^2} X_1 = cX_2$ and $\left(X_1 - \widetilde{X}_1\right) \perp \widetilde{X}_1$. Thus $\mathsf{E}(X_1 - cX_2)X_2 = 0$, and $c = \frac{R_X(1)}{R_X(0)} = \frac{b}{1+b^2}$. We have $\widetilde{X}_1 = \frac{b}{1+b^2}X_2$. Quite analogously we get $\widetilde{X}_3 = \frac{b}{1+b^2}X_2$, i.e., $\widetilde{X}_1 = \widetilde{X}_3$. We have

$$\alpha(2) = \operatorname{corr}\left(X_1 - \frac{b}{1+b^2}X_2, X_3 - \frac{b}{1+b^2}X_2\right).$$

Obviously,

$$\mathsf{E}\Big(X_1 - \frac{b}{1+b^2}X_2\Big)\Big(X_3 - \frac{b}{1+b^2}X_2\Big) = R_X(2) - \frac{2b}{1+b^2}R_X(1) + \frac{b^2}{(1+b^2)^2}R_X(0)$$
$$= -\frac{\sigma^2 b^2}{1+b^2},$$

similarly,

$$\mathsf{E}\Big(X_1 - \frac{b}{1+b^2}X_2\Big)^2 = R_X(0) + \frac{b^2}{(1+b^2)^2}R_X(0) - \frac{2b}{1+b^2}R_X(1) = \frac{\sigma^2(1+b^2+b^4)}{1+b^2},$$

and combining these results we conclude that

$$\alpha(2) = -\frac{b^2}{1+b^2+b^4}.$$

Generally, it can be shown that

$$\alpha(k) = -\frac{(-b)^k(1-b^2)}{1-b^{2(k+1)}},\ k \geq 1.$$

Definition 42 (An alternative definition of the partial correlation function)**.** Let $\{X_t,\, t \in \mathbb{Z}\}$ be a centered stationary sequence, let $P_{H_1^k}(X_{k+1})$ be the best linear prediction of X_{k+1} on the basis of $X_1, \dots, X_k$. If $H_1^k = \mathcal{H}\{X_1, \dots, X_k\}$, and $P_{H_1^k}(X_{k+1}) = \varphi_1 X_k + \dots + \varphi_k X_1$, 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*

$$\begin{aligned} \alpha(1) &= r(1), \\ \alpha(k) &= \frac{\begin{vmatrix} 1 & r(1) & \dots & r(k-2) & r(1) \\ r(1) & 1 & \dots & r(k-3) & r(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(k-1) & r(k-2) & \dots & r(1) & r(k) \end{vmatrix}}{\begin{vmatrix} 1 & r(1) & \dots & r(k-1) \\ r(1) & 1 & \dots & r(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(k-1) & r(k-2) & \dots & 1 \end{vmatrix}} \qquad k > 1, \end{aligned} \tag{107}$$

where r is the autocorrelation function of the sequence $\{X_t,\, t \in \mathbb{Z}\}$.

Proof. Denote $H_1^k = \mathcal{H}\{X_1, \dots, X_k\}$, $H_2^k = \mathcal{H}\{X_2, \dots, X_k\}$, $\widehat{X}_{k+1} = P_{H_1^k}(X_{k+1})$, $\widetilde{X}_1 = P_{H_2^k}(X_1)$, $\widetilde{X}_{k+1} = P_{H_2^k}(X_{k+1})$.

Since $X_1 = \widetilde{X}_1 + (X_1 - \widetilde{X}_1)$, where $\widetilde{X}_1 \in H_2^k$, $X_1 - \widetilde{X}_1 \perp H_2^k$, and $\widetilde{X}_{k+1} \in H_2^k$,

$$\mathsf{E}\widetilde{X}_{k+1}(X_1 - \widetilde{X}_1) = 0. \tag{108}$$

Consider

$$\widehat{X}_{k+1} = \varphi_1 X_k + \cdots + \varphi_k X_1.$$

Then

$$\widehat{X}_{k+1} = \big[\varphi_1 X_k + \cdots + \varphi_{k-1} X_2 + \varphi_k \widetilde{X}_1\big] + \big[\varphi_k (X_1 - \widetilde{X}_1)\big],$$

and the random variables in the brackets are mutually orthogonal. Then $\big[\varphi_k (X_1 - \widetilde{X}_1)\big]$ can be considered to be the projection of $\widehat{X}_{k+1}$ onto the Hilbert space $\widetilde{H} = \mathcal{H}\{X_1 - \widetilde{X}_1\} \subset H_1^k$. It is also the projection of X_{k+1} onto the space $\widetilde{H}$ and

$$\begin{aligned}\mathsf{E}\big(X_{k+1} - \varphi_k(X_1 - \widetilde{X}_1)\big)\big(X_1 - \widetilde{X}_1\big) &= 0 \\ &= \mathsf{E}X_{k+1}\big(X_1 - \widetilde{X}_1\big) - \varphi_k \mathsf{E}\big(X_1 - \widetilde{X}_1\big)^2.\end{aligned}$$

From here and from (108) we get

$$\varphi_k = \frac{\mathsf{E}X_{k+1}\big(X_1 - \widetilde{X}_1\big)}{\mathsf{E}\big(X_1 - \widetilde{X}_1\big)^2} = \frac{\mathsf{E}\big(X_{k+1} - \widetilde{X}_{k+1}\big)\big(X_1 - \widetilde{X}_1\big)}{\mathsf{E}\big(X_1 - \widetilde{X}_1\big)^2}. \tag{109}$$

Since $\mathsf{E}(X_1 - \widetilde{X}_1)^2 = \mathsf{E}(X_{k+1} - \widetilde{X}_{k+1})^2$, which holds from the fact that for a stationary sequence, $\operatorname{var}(X_2, \dots, X_k) = \operatorname{var}(X_k, \dots, X_2)$, we get from (109) that

$$\varphi_k = \operatorname{corr}(X_1 - \widehat{X}_1, X_{k+1} - \widehat{X}_{k+1}) = \alpha(k).$$

Now we will verify (107). We know that $\widehat{X}_{k+1} = \varphi_1 X_k + \varphi_2 X_{k-1} + \cdots + \varphi_k X_1 \in H_1^k$, $X_{k+1} - \widehat{X}_{k+1} \perp H_1^k$ and therefore

$$\mathsf{E}\big(X_{k+1} - (\varphi_1 X_k + \cdots + \varphi_k X_1)\big) X_{k+1-j} = 0, \quad j = 1, 2, \dots, k,$$

which is a system of equations

$$\begin{aligned} R(1) - \varphi_1 R(0) - \cdots - \varphi_k R(k-1) &= 0 \\ R(2) - \varphi_1 R(1) - \cdots - \varphi_k R(k-2) &= 0 \\ &\vdots \\ R(k) - \varphi_1 R(k-1) - \cdots - \varphi_k R(0) &= 0. \end{aligned}$$

Dividing each equation by $R(0)$, we get the system of equations

$$\begin{aligned} \varphi_1 + \varphi_2 r(1) + \dots + \varphi_k r(k-1) &= r(1) \\ \varphi_1 r(1) + \varphi_2 + \dots + \varphi_k r(k-2) &= r(2) \\ &\vdots \\ \varphi_1 r(k-1) + \varphi_2 r(k-2) + \dots + \varphi_k &= r(k), \end{aligned}$$

or, in the matrix form,

$$\begin{pmatrix} 1 & r(1) & \dots & r(k-1) \\ r(1) & 1 & \dots & r(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(k-1) & r(k-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_k \end{pmatrix} = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(k) \end{pmatrix},$$

The ratio of determinants (107) gives the solution for φ_k. □

Example 49. Consider again the causal AR(1) process

$$X_t = \varphi X_{t-1} + Y_t,$$

where $|\varphi| < 1$ and $Y_t \sim \mathrm{WN}(0, \sigma^2)$. Let us compute the partial autocorrelation function according to formula (107). We get

$$\alpha(k) = \frac{\begin{vmatrix} 1 & \varphi & \dots & \varphi^{k-2} & \varphi \\ \varphi & 1 & \dots & \varphi^{k-3} & \varphi^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{k-1} & \varphi^{k-2} & \dots & \varphi & \varphi^k \end{vmatrix}}{\begin{vmatrix} 1 & \varphi & \dots & \varphi^{k-1} \\ \varphi & 1 & \dots & \varphi^{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi^{k-1} & \varphi^{k-2} & \dots & 1 \end{vmatrix}} \qquad k > 1, \tag{110}$$

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 $\mathsf{E}X_t = \mu$ and autocovariance function $R(s,t) = R(s-t)$.

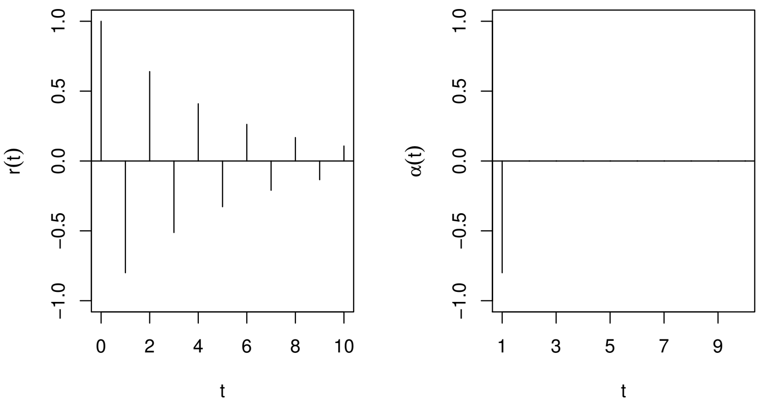

Figure 8: Autocorrelation (left) and partial autocorrelation function (right) of the AR(1) sequence $X_t = -0{,}8\, X_{t-1} + Y_t$

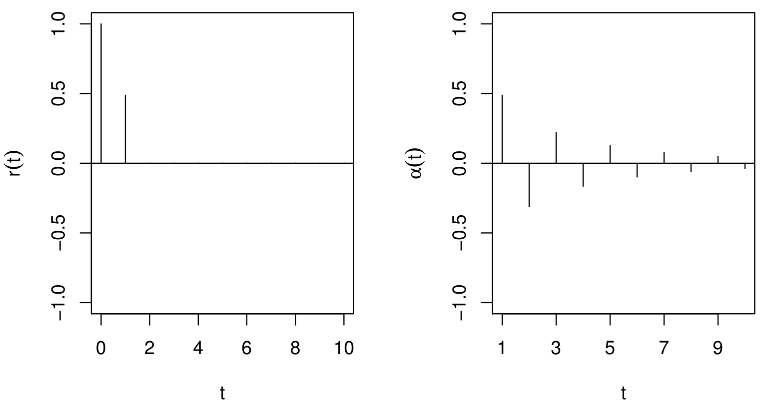

Figure 9: Autocorrelation function (left) and partial autocorrelation function (right) of the MA(1) sequence $X_t = Y_t + 0{,}8\, Y_{t-1}$

A common estimator of the mean value is the *sample mean* defined by

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

We know that $\overline{X}_n$ is an unbiased estimator of the expected value, since $\mathsf{E}\overline{X}_n = \mu$. We also know that if the sequence $\{X_t,\, t \in \mathbb{Z}\}$ is mean square ergodic, then $\overline{X}_n \to \mu$ 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 of a stationary sequence is

$$\mathrm{var}\overline{X}_n = \frac{1}{n} \sum_{k=-n+1}^{n-1} R(k) \left(1 - \frac{|k|}{n}\right),$$

and if $\sum_{-\infty}^{\infty} |R(k)| < \infty$, then $n\,\mathrm{var}\overline{X}_n \to \sum_{-\infty}^{\infty} R(k) = 2\pi f(0)$ where $f(\lambda)$ is the spectral density of the sequence $\{X_t,\, t \in \mathbb{Z}\}$, see Theorem 40. We have also proved a few central limit theorems for selected strictly stationary sequences, saying that $\overline{X}_n$ has asymptotically distribution $\mathcal{N}(\mu, \frac{\Delta^2}{n})$, where Δ^2 is an asymptotic variance (see Theorems 47, 48 and 49).

However, the sample mean $\overline{X}_n$ 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

$$X_t = \mu + \widetilde{X}_t, \quad t = 1, \ldots, n, \tag{111}$$

where $\widetilde{X}_t,\ t = 1, \ldots, n$, 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 μ is statistic

$$\widehat{\mu}_n = (\mathbf{1}_n' \mathbf{\Gamma}_n^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n' \mathbf{\Gamma}_n^{-1} \mathbf{X}_n, \tag{112}$$

where

$$\mathbf{\Gamma}_n = \mathrm{var}\,\mathbf{X}_n = \begin{pmatrix} R(0) & R(1) & \ldots & R(n-1) \\ R(1) & R(0) & \ldots & R(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(n-1) & R(n-2) & \ldots & R(0) \end{pmatrix},$$

is regular matrix according to Theorem 53, $\mathbf{1}_n = (1, \ldots, 1)'$ and $\mathbf{X}_n = (X_1, \ldots, X_n)'$.

The variance of this estimator is

$$\mathrm{var}\,\widehat{\mu}_n = (\mathbf{1}_n' \mathbf{\Gamma}_n^{-1} \mathbf{1}_n)^{-1}. \tag{113}$$

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 problems. For estimators we usually work with *the sample autocovariance function*

$$\widehat{R}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right), \qquad k = 0, 1, \dots, n-1 \tag{114}$$

and $\widehat{R}(k) = \widehat{R}(-k)$ pro $k < 0$. Let us remark that the sample autocovatriance function is not an unbiased estimator of the autocovariance function, i.e., $\mathsf{E}\widehat{R}(k) \neq R(k)$.

The matrix

$$\widehat{\mathbf{\Gamma}}_n = \begin{pmatrix} \widehat{R}(0) & \widehat{R}(1) & \dots & \widehat{R}(n-1) \\ \widehat{R}(1) & \widehat{R}(0) & \dots & \widehat{R}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{R}(n-1) & \widehat{R}(n-2) & \dots & \widehat{R}(0) \end{pmatrix}$$

will be regular, if $\widehat{R}(0) > 0$. For given $X_1, \dots, X_n$, function

$$\widehat{R}(k) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-|k|}(X_t - \overline{X}_n)(X_{t+k} - \overline{X}_n), & |k| < n, \\ 0, & |k| \geq n \end{cases} \tag{115}$$

can be wieved to be the autocovariance function of an MA$(n-1)$ sequence and thus, the regularity of matrix $\widehat{\mathbf{\Gamma}}_n$ follows from Theorem 53.

If we dispose only n observations $X_1, \dots, X_n$, we can estimate $R(k)$, $k = 0, \dots, n-1$. 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

$$\widehat{r}(k) = \frac{\widehat{R}(k)}{\widehat{R}(0)} = \frac{\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right)}{\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2},$$

if $\widehat{R}(0) = \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2 > 0$.

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*

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \alpha_j Y_{t-j},$$

where Y_t, $t \in \mathbb{Z}$, are independent identically distributed random variables with zero mean and finite positive variance σ^2, and let $\mathsf{E}|Y_t|^4 < \infty$ and $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$.

Let $r(k)$, $k \in \mathbb{Z}$, be the autocorrelation function of the sequence $\{X_t,\, t \in \mathbb{Z}\}$ and $\widehat{r}(k)$ be the sample autocorrelation at lag k, based on $X_1, \dots, X_n$.

Then for each $h = 1, 2, \dots,$ as $n \to \infty$, the random vector $\sqrt{n}\big(\widehat{\mathbf{r}}(h) - \mathbf{r}(h)\big)$ converges in distribution to a random vector with normal distribution $\mathcal{N}_h(\mathbf{0}, \mathbf{W})$, where

$$\widehat{\mathbf{r}}(h) = (\widehat{r}(1), \dots, \widehat{r}(h))', \;\; \mathbf{r}(h) = (r(1), \dots, r(h))',$$

and $\mathbf{W}$ is an $h \times h$ matrix elements of which are

$$w_{ij} = \sum_{k=1}^{\infty} \big[r(k+i) + r(k-i) - 2r(i)r(k)\big]\big[r(k+j) + r(k-j) - 2r(j)r(k)\big] \qquad (116)$$

$i, j = 1, \dots, h.$

Proof. Brockwell and Davis (1991), Theorem 7.2.1. □

Remark 22. Formula (116) is called *the Bartlett formula.* From the assertion of the theorem we especially get for any i

$$\sqrt{n}\big(\widehat{r}(i) - r(i)\big) \xrightarrow{\text{D}} \mathcal{N}(0, w_{ii}), \quad n \to \infty,$$

i.e., for large n,

$$\widehat{r}(i) \sim \mathcal{N}\left(r(i), \frac{w_{ii}}{n}\right).$$

Example 50. Consider the AR(1) sequence

$$X_t = \varphi X_{t-1} + Y_t,\; t \in \mathbb{Z},$$

where $|\varphi| < 1$ and $Y_t, t \in \mathbb{Z}$ are i. i. d. with zero mean, finite non-zero variance σ^2 and with finite moments $\mathsf{E}|Y_t|^4$. Then $r(k) = \varphi^{|k|}$, thus $r(1) = \varphi$ and according to Theorem 58,

$$\sqrt{n}\big(\widehat{r}(1) - \varphi)\big) \xrightarrow{\text{D}} \mathcal{N}(0, w_{11}), \quad n \to \infty,$$

where

$$\begin{aligned} w_{11} &= \sum_{k=1}^{\infty} \big[r(k+1) + r(k-1) - 2r(1)r(k)\big]^2 = \sum_{k=1}^{\infty} \big(\varphi^{k-1} - \varphi^{k+1}\big)^2 \\ &= (1-\varphi^2)^2 \sum_{k=1}^{\infty} \varphi^{2(k-1)} = 1 - \varphi^2. \end{aligned}$$

If we denote $\widehat{r}(1) := \widehat{\varphi}$, we can write

$$\sqrt{n}\big(\widehat{\varphi} - \varphi\big) \xrightarrow{\text{D}} \mathcal{N}(0, 1 - \varphi^2), \quad n \to \infty$$

or

$$\sqrt{n}\frac{\widehat{\varphi}-\varphi}{\sqrt{1-\varphi^2}} \xrightarrow{\text{D}} \mathcal{N}(0,1), \quad n\to\infty.$$

From here it follows that $\widehat{\varphi} \xrightarrow{\text{P}} \varphi$ (see, e.g. Brockwell and Davis, 1991, Chap. 6) and also

$$\sqrt{n}\frac{\widehat{\varphi}-\varphi}{\sqrt{1-\widehat{\varphi}^2}} \xrightarrow{\text{D}} \mathcal{N}(0,1), \quad n\to\infty.$$

The asymptotic 95% confidence interval for φ is

$$\left(\widehat{\varphi}-1,96\sqrt{\frac{1-\widehat{\varphi}^2}{n}},\ \widehat{\varphi}+1,96\sqrt{\frac{1-\widehat{\varphi}^2}{n}}\right).$$

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 $\mathbf{W}$ 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., $\mathbf{W}=\mathbf{I}$ is the identity matrix. It means that for large n, the vector $\widehat{\mathbf{r}}(h)=\big(\widehat{r}(1),\dots,\widehat{r}(h)\big)'$ has approximately normal distribution $\mathcal{N}_h(\mathbf{0},\frac{1}{n}\mathbf{I})$. For large n, therefore the random variables $\widehat{r}(1),\dots,\widehat{r}(h)$ are approximately independent and identically distributed with zero mean and variance $\frac{1}{n}$. In the plot of sample autocorrelations $\widehat{r}(k)$ for $k=1,\dots,$ approximately 95% of them should be in the interval $\big(-1,96\frac{1}{\sqrt{n}},\ 1,96\frac{1}{\sqrt{n}}\big)$.

The sample partial autocorrelation function is defined to be $\widehat{\alpha}(k)=\widehat{\varphi}_k$, where $\widehat{\varphi}_k$ 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 $\frac{1}{n}\sum_{t=1}^{n}(X_t-\overline{X}_n)^2>0$.

Example 52. In Figure 12 the plot of the Wolf index of the annual number of the Sunspots (1700-1987)[3] is displayed. In Figures 13 and 14 we can see the sample autocovariance 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

$$\begin{aligned} a(z)=1&-1.182z+0.4248z^2+0.1619z^3-0.1687z^4\\ &+0.1156z^5-0.02689z^6-0.005769z^7\\ &+0.02251z^8-0.2062z^9 \end{aligned}$$

$Y_t\sim \text{WN}(0,\sigma^2), \sigma^2=219.58.$

[3]Source: WDC-SILSO, Royal Observatory of Belgium, Brussels

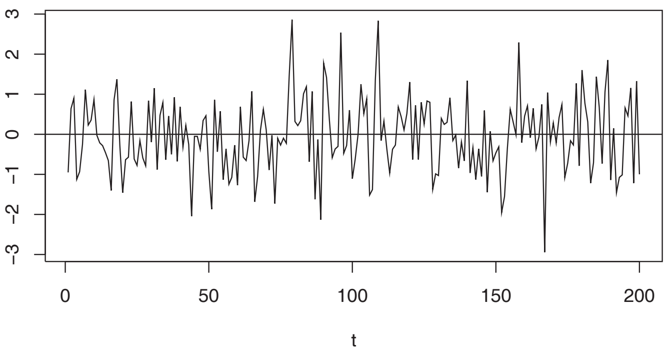

Figure 10: Trajectory of a strict white noise process

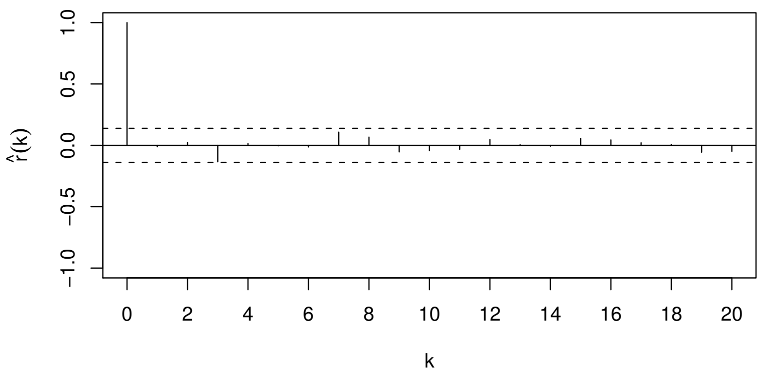

Figure 11: Sample autocorrelation function of a strict white noise

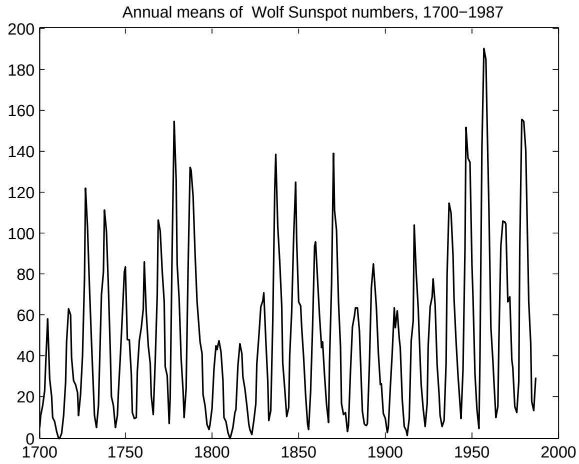

Figure 12: Number of Sunspots, the Wolf index

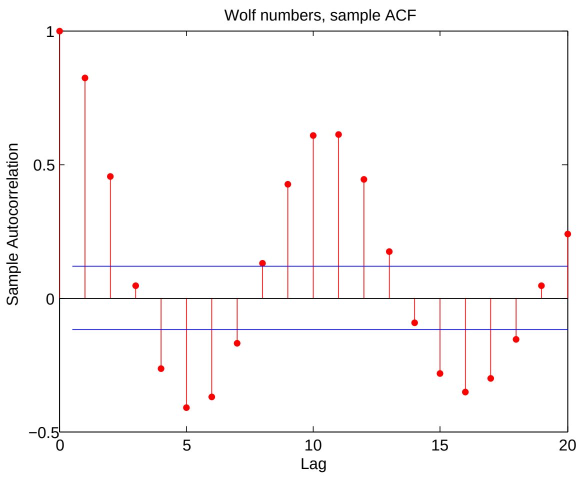

Figure 13: Wolf index, estimated autocorrelation function

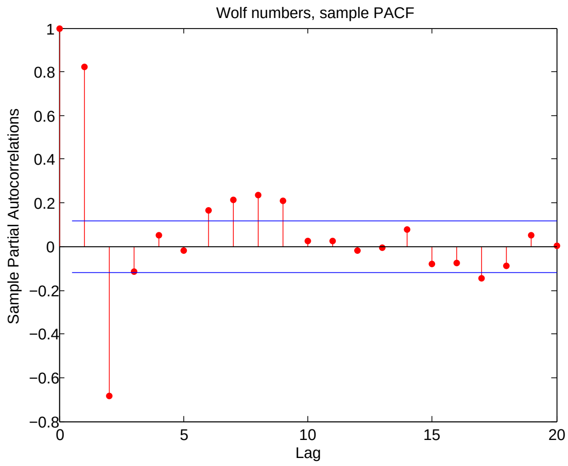

Figure 14: Wolf index, estimated partial autocorrelation function

15 Estimation of parameters in ARMA models

15.1 Estimation in AR sequences

Let us consider a real-valued stationary causal AR(p) sequence of known order p,

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t,\ \ t \in \mathbb{Z}, \tag{117}$$

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

$$R(0) = \varphi_1 R(1) + \cdots + \varphi_p R(p) + \sigma^2, \tag{118}$$

$$R(k) = \varphi_1 R(k-1) + \cdots + \varphi_p R(k-p), \qquad k \geq 1. \tag{119}$$

The system of equations for $k = 1, \ldots, p$ can be written in the matrix form

$$\boldsymbol{\Gamma}\boldsymbol{\varphi} = \boldsymbol{\gamma}, \tag{120}$$

where

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} R(0) & \ldots & R(p-1) \\ \vdots & \ddots & \vdots \\ R(p-1) & \ldots & R(0) \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} R(1) \\ \vdots \\ R(p) \end{pmatrix}.$$

If we replace the values of $R(k)$ in $\mathbf{\Gamma}$ and $\boldsymbol{\gamma}$ by their sample counterparts

$$\widehat{R}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right),$$

we get the matrix $\widehat{\mathbf{\Gamma}}$ and the vector $\widehat{\boldsymbol{\gamma}}$. If we plug these estimators into equation (120), we obtain moment estimators of $\varphi_1, \ldots, \varphi_p$ by solving

$$\widehat{\boldsymbol{\varphi}} = (\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p)' = \widehat{\mathbf{\Gamma}}^{-1}\widehat{\boldsymbol{\gamma}}, \tag{121}$$

provided $\widehat{\mathbf{\Gamma}}^{-1}$ exists. From subsection 14.2 we know that a sufficient condition for $\widehat{\mathbf{\Gamma}}$ to be regular is

$$\widehat{R}(0) = \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2 > 0.$$

The moment estimate of σ^2 is obtained from (118) as

$$\widehat{\sigma}^2 = \widehat{R}(0) - \widehat{\varphi}_1\widehat{R}(1) - \cdots - \widehat{\varphi}_p\widehat{R}(p) = \widehat{R}(0) - \widehat{\boldsymbol{\varphi}}'\widehat{\boldsymbol{\gamma}}. \tag{122}$$

Remark 23. 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 $\mathrm{WN}(0, \sigma^2)$. Moment estimators of parameters φ and σ^2 are

$$\widehat{\varphi} = \frac{\widehat{R}(1)}{\widehat{R}(0)} = \widehat{r}(1), \quad \widehat{\sigma}^2 = \widehat{R}(0) - \widehat{\varphi}\widehat{R}(1) = \widehat{R}(0)\left(1 - \widehat{\varphi}^2\right).$$

Moment estimator of the parameter φ is in this case the same as the sample autocorrelation coefficient $\widehat{r}(1)$ (compare with Example 50.)

Asymptotic properties of the 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 i. i. d. random variables with zero mean and finite non-zero variance σ^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 $\widehat{\boldsymbol{\varphi}} = (\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p)'$ and $\widehat{\sigma}^2$ be the moment estimators of $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)'$ and σ^2, repectively, computed from $X_1, \ldots, X_n$.*

Then

$$\sqrt{n}\left(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\right) \xrightarrow{\mathrm{D}} \mathcal{N}_p\left(\mathbf{0}, \sigma^2\mathbf{\Gamma}^{-1}\right), \quad n \to \infty,$$

where $\mathbf{\Gamma}$ is the matrix with elements $\Gamma_{ij} = R(i-j), 1 \le i, j \le p$, R is the autocovariance function of $\{X_t,\ t \in \mathbb{Z}\}$.

Further, it holds

$$\widehat{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma^2, \quad n \to \infty.$$

Proof. Brockwell and Davis (1991), Theorem 8.1.1. □

Least squares method

Consider again sequence (117) and suppose $X_1, \dots, X_n$ to be known. The least square estimators of parameters $\varphi_1, \dots, \varphi_p$ are obtained by minimizing the sum of squares

$$\min_{\varphi_1,\dots,\varphi_p} \sum_{t=p+1}^{n} \left(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}\right)^2 .$$

The problem leads to the solution of the system of equations

$$\sum_{t=p+1}^{n} \left(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}\right) X_{t-j} = 0, \qquad j = 1, \dots, p,$$

i. e., to the system

$$\begin{aligned}
\varphi_1 \sum_{t=p+1}^{n} X_{t-1}^2 + \cdots + \varphi_p \sum_{t=p+1}^{n} X_{t-1} X_{t-p} &= \sum_{t=p+1}^{n} X_t X_{t-1}, \\
&\vdots \\
\varphi_1 \sum_{t=p+1}^{n} X_{t-1} X_{t-p} + \cdots + \varphi_p \sum_{t=p+1}^{n} X_{t-p}^2 &= \sum_{t=p+1}^{n} X_t X_{t-p}.
\end{aligned}$$

If we write (117) in commonly used form

$$X_t = \boldsymbol{\varphi}' \boldsymbol{X}_{t-1} + Y_t,$$

where $\boldsymbol{X}_{t-1} = (X_{t-1}, \dots, X_{t-p})'$, then the solution is of the form

$$\widetilde{\boldsymbol{\varphi}} = (\widetilde{\varphi}_1, \dots, \widetilde{\varphi}_p)' = \left(\sum_{t=p+1}^{n} \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}' \right)^{-1} \sum_{t=p+1}^{n} \boldsymbol{X}_{t-1} X_t. \tag{123}$$

The least squares estimator of σ^2 is

$$\widetilde{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} \left(X_t - \widetilde{\boldsymbol{\varphi}}' \boldsymbol{X}_{t-1}\right)^2 . \tag{124}$$

It can be shown that estimators $\widetilde{\boldsymbol{\varphi}}$ and $\widetilde{\sigma}^2$ have the same asymptotic properties as the moment estimators. In particular, as $n \to \infty$

$$\sqrt{n}\,(\widetilde{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \xrightarrow{\mathrm{D}} \mathcal{N}_p\left(0, \sigma^2 \boldsymbol{\Gamma}^{-1}\right)$$

and

$$\widetilde{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma^2,$$

where $\mathbf{\Gamma}$ is the same matrix as in Theorem 59 (Brockwell and 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 i. i. d. 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

$$f(x_1, y_2, \ldots, y_n) = f_1(x_1) f_2(y_2, \ldots, y_n) = f_1(x_1) \left(2\pi\sigma^2\right)^{-(n-1)/2} \exp\Big\{-\frac{1}{2\sigma^2} \sum_{t=2}^{n} y_t^2\Big\}.$$

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 the transformation density theorem we easily obtain that the joint density of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n) = \left(2\pi\sigma^2\right)^{-n/2} \sqrt{1-\varphi^2} \exp\Big\{-\frac{1}{2\sigma^2}\Big((1-\varphi^2)x_1^2 + \sum_{t=2}^{n} \left(x_t - \varphi x_{t-1}\right)^2\Big)\Big\}. \quad (125)$$

The likelihood function $L(\varphi, \sigma^2)$ is of the same form as (125). Maximum likelihood estimates are then $\bar{\varphi}, \bar{\sigma}^2$, 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 the *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

$$f(x_2, \ldots, x_n | x_1) = \left(2\pi\sigma^2\right)^{-(n-1)/2} \exp\Big\{-\frac{1}{2\sigma^2} \sum_{t=2}^{n} \left(x_t - \varphi x_{t-1}\right)^2\Big\}. \quad (126)$$

The conditional maximum likelihood estimators are obtained by maximizing function (126) with respect to φ and σ^2.

Similarly, if we consider a general causal AR(p) sequence (117), where Y_t are i. i. d. 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

$$f(x_{p+1}, \ldots, x_n | x_1, \ldots, x_p) = \left(2\pi\sigma^2\right)^{-(n-p)/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} \left(x_t - \boldsymbol{\varphi}' \boldsymbol{x}_{t-1}\right)^2\right\},$$

where $\boldsymbol{x}_{t-1} = (x_{t-1}, \ldots, x_{t-p})'$, and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)'$.

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 the least squares estimators and the 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 the 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

$$R_X(k) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|k|} \theta_j \theta_{j+|k|}, & |k| \le q, \\ 0 & \text{elsewhere} \end{cases} \tag{127}$$

(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 $R_X(k)$ by the sample autocovariances $\widehat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} \left(X_t - \overline{X}_n\right) \left(X_{t+k} - \overline{X}_n\right)$. We get the system of $q+1$ equations for $\theta_1, \ldots, \theta_q, \sigma^2$

$$\begin{aligned} \widehat{R}(0) &= \sigma^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right), \\ \widehat{R}(1) &= \sigma^2 \left(\theta_1 + \theta_1\theta_2 + \cdots + \theta_{q-1}\theta_q\right), \\ &\vdots \\ \widehat{R}(q) &= \sigma^2 \theta_q. \end{aligned} \tag{128}$$

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, $R_X(0) = \sigma^2(1 + \theta^2)$ a $R_X(1) = \sigma^2\theta$, thus

$$r(1) = \frac{R_X(1)}{R_X(0)} = \frac{\theta}{1 + \theta^2}.$$

It can be shown that in this case, $|r(1)| \le \frac{1}{2}$ for all real values of θ. Consequently, solving the last equation with respect to θ, 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 those with the negative sign is in absolute value less than 1, which corresponds to the invertible process.

The moment estimators of θ a σ^2 now can be obtained from equations (128) that can be rewritten into the form

$$\begin{aligned} \widehat{R}(0) &= \sigma^2(1+\theta^2), \\ \widehat{r}(1) &= \frac{\theta}{1+\theta^2}. \end{aligned}$$

For θ we have two solutions

$$\widehat{\theta}_{1,2} = \frac{1 \pm \sqrt{1 - 4\widehat{r}^2(1)}}{2\widehat{r}(1)},$$

that take real values if $|\widehat{r}(1)| \le \frac{1}{2}$.

Provided that the process is invertible and $|\widehat{r}(1)| < \frac{1}{2}$, the moment estimators are

$$\begin{aligned} \widehat{\theta} &= \frac{1 - \sqrt{1 - 4\widehat{r}^2(1)}}{2\widehat{r}(1)}, \\ \widehat{\sigma}^2 &= \frac{\widehat{R}(0)}{1+\widehat{\theta}^2}. \end{aligned}$$

If $|\widehat{r}(1)| = \frac{1}{2}$, we take

$$\begin{aligned} \widehat{\theta} &= \frac{1}{2\widehat{r}(1)} = \frac{\widehat{r}(1)}{|\widehat{r}(1)|}, \\ \widehat{\sigma}^2 &= \frac{1}{2}\widehat{R}(0). \end{aligned}$$

For $|\widehat{r}(1)| > \frac{1}{2}$ the real-valued solution of (128) does not exist. In such a case we use the same estimates as given for $|\widehat{r}(1)| = \frac{1}{2}$.

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 $Y_t \sim \text{WN}(0, \sigma^2)$, $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2$ are unknown parameters and $X_1, \dots, X_n$ are given observations, we can proceed as follows:

First we use an analogy of the Yule-Walker equations for the autocovariances $R_X(k)$, $k = q+1, \dots, q+p$. We get equations

$$R_X(k) = \varphi_1 R_X(k-1) + \dots + \varphi_p R_X(k-p)$$

for unknown parameters $\varphi_1, \dots, \varphi_p$. If we replace the theoretical values R_X by their estimates $\widehat{R}_X(k) = \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right)$, we obtain estimates of parameters $\widehat{\varphi}_1, \dots, \widehat{\varphi}_p$.

Further, we put $Z_t = X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p}$ and want to estimate $\theta_1, \dots \theta_q$ and σ^2 in the MA(q) model

$$Z_t = Y_t + \theta_1 Y_{t-1} + \dots + \theta_q Y_{t-q}.$$

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 $\beta_0 = 1, \beta_j = -\varphi_j, j = 1, \dots, p$, 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 $\theta_1, \dots, \theta_q$ a σ^2 are obtained from (128) replacing $\widehat{R}(k)$ by estimates

$$\widehat{R}_Z(k) = \sum_{j=0}^{p}\sum_{l=0}^{p} \widehat{\beta}_j \widehat{\beta}_l \widehat{R}_X(k+j-l),$$

where $\widehat{\beta}_j = -\widehat{\varphi}_j$ and $\widehat{R}_X(k)$ are sample autocovariances computed from $X_1, \dots, X_n$.

The moment estimators are under some assumptions consistent and asymptotically normal, but they are not too stable 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

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \dots + \theta_q Y_{t-q},$$

where $Y_t \sim \text{WN}(0, \sigma^2)$, $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2$ are unknown parameters and $X_1, \dots, X_n$ 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 $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_m$.

- Estimate residuals $\widehat{\tilde{Y}}_t,\ t = m+1, \ldots, n$ and use them as known regressors in the regression model

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \theta_1 \widehat{\tilde{Y}}_{t-1} + \cdots + \theta_q \widehat{\tilde{Y}}_{t-q} + Y_t,\ t = \max(p, q, m) + 1, \ldots, n$$

  and estimate $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2$ from this regression model with regressors $X_{t-1}, \ldots, X_{t-p}, \widehat{\tilde{Y}}_{t-1}, \widehat{\tilde{Y}}_{t-q}$.

For other estimating methods see, e. g., Prášková, 2016.

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

$$I_n(\lambda) = \frac{1}{2\pi n} \Big| \sum_{t=1}^{n} X_t e^{-it\lambda} \Big|^2, \quad \lambda \in [-\pi, \pi]. \tag{129}$$

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

$$A(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \cos t\lambda, \quad B(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \sin t\lambda. \tag{131}$$

For a real-valued sequence, the periodogram can be also expressed by

$$I_n(\lambda) = \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n} X_t X_s e^{-i(t-s)\lambda} = \frac{1}{2\pi n}\sum_{k=-n+1}^{n-1}\sum_{s=\max(1,1-k)}^{\min(n,n-k)} X_s X_{s+k} e^{-ik\lambda}$$

$$= \frac{1}{2\pi}\sum_{k=-n+1}^{n-1} e^{-ik\lambda} C_k, \qquad (132)$$

where

$$C_k = \frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k}, \quad k \geq 0 \qquad (133)$$
$$= C_{-k}, \quad k < 0.$$

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*

$$EI_n(\lambda) \to f(\lambda), \quad \lambda \in [-\pi, \pi], \qquad (134)$$

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

$$\mathsf{E}I_n(\lambda) = \frac{1}{2\pi n}\sum_{k=-n+1}^{n-1} e^{it\lambda} R(k)(n - |k|).$$

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 $n \to \infty$,

$$|f(\lambda) - EI_n(\lambda)| \leq \frac{1}{2\pi}\sum_{|k|\geq n} |R(k)| + \frac{1}{2\pi n}\sum_{k=-n+1}^{n-1} |R(k)||k| \to 0. \qquad (135)$$

□

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 i. i. d. random variables with zero mean and variance σ^2, $0 < \sigma^2 < \infty$. Put $n = 2m + 1$ and consider the periodogram I_n computed from $X_1, \dots, X_n$ at frequencies $\lambda_j = \frac{2\pi j}{n}, j = 1, \dots, m$. Then random variables $I_n(\lambda_1), \dots, I_n(\lambda_m)$ are independent and identically distributed as $\frac{\sigma^2}{4\pi}\chi^2(2)$, where $\chi^2(2)$ denotes the χ^2 distribution with two degrees of freedom.*

Proof. Consider random vector $\boldsymbol{J} = (A(\lambda_1), \dots, A(\lambda_m), B(\lambda_1), \dots, B(\lambda_m))'$, 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 the random vector $(X_1, \dots, X_n)'$. Further we prove that all the components of the vector $\boldsymbol{J}$ are mutually uncorrelated (and thus, independent), and identically distributed with zero mean and variance σ^2. For this we use the following identities for trigonometric functions

$$\sum_{t=1}^{n} \cos^2(t\lambda_r) = \frac{n}{2}, \quad r = 1, \dots, m,$$

$$\sum_{t=1}^{n} \sin^2(t\lambda_r) = \frac{n}{2}, \quad r = 1, \dots, m,$$

$$\sum_{t=1}^{n} \sin(t\lambda_r)\cos(t\lambda_s) = 0, \quad r, s = 1, \dots, m,$$

$$\sum_{t=1}^{n} \sin(t\lambda_r)\sin(t\lambda_s) = 0, \quad r, s = 1, \dots, m, r \neq s,$$

$$\sum_{t=1}^{n} \cos(t\lambda_r)\cos(t\lambda_s) = 0, \quad r, s = 1, \dots, m, r \neq s$$

from that the result follows using simple computations. Particularly, we get for any $r = 1, \dots, m$ that $A(\lambda_r) \sim \mathcal{N}(0, \sigma^2), B(\lambda_r) \sim \mathcal{N}(0, \sigma^2)$, thus

$$\frac{A^2(\lambda_r) + B^2(\lambda_r)}{\sigma^2} = \frac{4\pi}{\sigma^2} I_n(\lambda_r) \sim \chi^2(2).$$

□

Remark 24. From the assumption that $\{X_t, t \in \mathbb{Z}\}$ is a Gaussian random sequence of i. i. d. random variables with zero mean and variance σ^2 we can easily conclude that the spectral density of this sequence is

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

since $\{X_t, t \in \mathbb{Z}\}$ is the white noise. From Theorem 61 and properties of the χ^2 distribution we have for $r = 1, \dots, m$,

$$\mathsf{E} I_n(\lambda_r) = 2\frac{\sigma^2}{4\pi} = \frac{\sigma^2}{2\pi} = f(\lambda_r),$$

$$\operatorname{var} I_n(\lambda_r) = 4\frac{\sigma^4}{16\pi^2} = f^2(\lambda_r).$$

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

$$\begin{aligned}\lim_{n\to\infty} \operatorname{var} I_n(\lambda) &= f^2(\lambda), \quad \lambda \neq 0,\ \lambda \in (-\pi, \pi)\\ &= 2f^2(\lambda), \quad \lambda = 0,\ \lambda = \pm\pi\end{aligned}$$

(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

$$X_t = \alpha e^{it\lambda_0} + Y_t,\ Y_t \sim \mathrm{WN}(0, \sigma^2)$$

where α is a nonzero constant and $\lambda_0 \in [-\pi, \pi]$. Then

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t e^{-it\lambda} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t e^{-it\lambda} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e^{-it(\lambda-\lambda_0)} \tag{136}$$

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 ∞ as $n \to \infty$ while for $\lambda \neq \lambda_0$ is negligible. It means that if there is a single periodic component at frequency λ_0 the periodogram takes in it the largest value. Since usually the frequency λ_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 i. i. d. random variables with zero mean and variance σ^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 statistic*

$$W = \frac{\max_{1\le r\le m} I_n(\lambda_r)}{I_n(\lambda_1) + \cdots + I_n(\lambda_m)} \tag{137}$$

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

$$P(W > x) = 1 - \sum_{k=0}^{[1/x]} (-1)^k \binom{m}{k}(1-kx)^{m-1}, \quad 0 < x < 1. \tag{138}$$

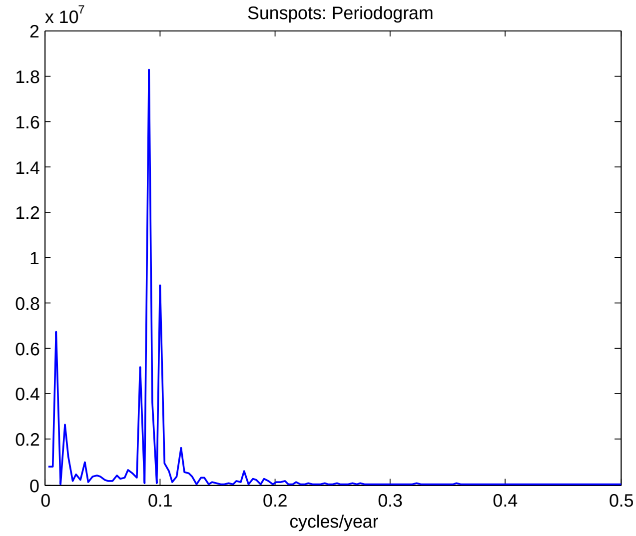

Figure 15: Periodogram of the Sunspots, Wolf index. The maximum corresponds to the cycle with period 11.0769 years

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 i. i. d. with distribution $\mathcal{N}(0, \sigma^2)$ " against the alternative that the null hypothesis is violated. The test statistic is based on Theorem 62 and reject the null hypothesis at level α if $W > c_\alpha$ where c_α is a critical value that can be computed from (138).

Estimators of spectral density

We have seen in Theorem 60 that the periodogram is the 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

$$K(\lambda) \geq 0, \;\; K(\lambda) = K(-\lambda), \int_{-\pi}^{\pi} K(\lambda) d\lambda = 1, \; \int_{-\pi}^{\pi} K^2(\lambda) d\lambda < \infty,$$

is an asymptotically unbiased and consistent estimator of

$$\int_{/\pi}^{\pi} f(\lambda)K(\lambda)d\lambda.$$

Then a consistent estimator of $f(\lambda_0)$ is considered to be

$$\widehat{f}_n(\lambda_0) = \int_{-\pi}^{\pi} K(\lambda - \lambda_0)I_n(\lambda)d\lambda.$$

If we expand function K into the Fourier series with the Fourier coefficients w_k and express the periodogram by using formulas (132) and (133), we get

$$\begin{aligned}
\widehat{f}_n(\lambda_0) &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} e^{-ik\lambda} C_k K(\lambda - \lambda_0) d\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \int_{-\pi}^{\pi} e^{-ik\lambda} K(\lambda - \lambda_0) d\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \int_{-\pi}^{\pi} e^{-ik\lambda} \sum_{j=-\infty}^{\infty} e^{ij(\lambda-\lambda_0)} w_j d\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \sum_{j=-\infty}^{\infty} e^{-ij\lambda_0} w_j \int_{-\pi}^{\pi} e^{(ij\lambda - ik\lambda)} d\lambda \\
&= \sum_{k=-n+1}^{n-1} e^{-ik\lambda_0} C_k w_k = C_0 w_0 + 2\sum_{k=1}^{n-1} C_k w_k \cos(k\lambda). \qquad (139)
\end{aligned}$$

One of the commonly used kernel function is so-called *Parzen window*, which is usually presented by coefficients

$$w_k = \begin{cases} 1 - 6\left(\frac{k}{M}\right)^2 - 6\left(\frac{|k|}{M}\right)^3, & |k| < \frac{M}{2} \\ 2\left(1 - \frac{|k|}{M}\right)^3, & \frac{M}{2} < |k| \le M \\ 0, & |k| > M \end{cases}$$

where M is a truncation point that depends on n ($\frac{n}{6} < M < \frac{n}{5}$). For more information on the choice of K, respectively of w_k, see, e.g., Anděl, 1976, or Brockwell and Davis, Chap. 10.

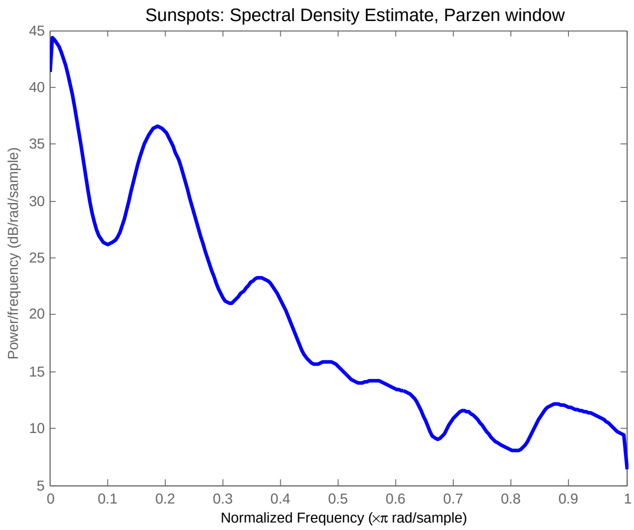

Figure 16: Sunspots, Wolf index. Estimator of the spectral density, Parzen kernel

References

[1] Anděl, J.: Statistická analýza časových řad. SNTL, Praha 1976 (In Czech)

[2] Anděl, J.: Základy matematické statistiky. MATFYZPRESS, Praha 2007 (In Czech)

[3] Bosq, D., Nguyen, Hung T.: A Course in Stochastic Processes. Stochastic Models and Statistical Inference. Kluwer, Dordrecht 1996

[4] Brockwell, P. J., Davis, R. A.: Time Series: Theory and Methods. Springer-Verlag, New York 1991

[5] Loève, M.: Probability Theory. Van Nostrand, Princeton 1960

[6] Prášková, Z.: Základy náhodných procesů II, Karolinum, Praha 2016. 2nd, extended edition (In Czech)

[7] Prášková, Z., Lachout, P.: Základy náhodných procesů I, Matfyzpress, Praha 2012. 2nd, extended edition (In Czech)

[8] Priestley, M. B.: Spectral Analysis and Time Series, Vol. I. Academic Press, London 1981

[9] Rao, R. C. : Lineární metody statistické indukce a jejich aplikace. Academia, Praha 1978. Czech translation from Rao, R. C. : Linear Statistical Inference and its Applications, Wiley, New York, 1973

[10] Rozanov, Yu. A. : Stacionarnyje slučajnyje processy. Gosudarstvennoje izdatelstvo fiziko-matematičeskoj literatury, Moskva 1963 (In Russian)

[11] Rudin, W. : Analýza v reálném a komplexním oboru. Academia, Praha. 2003. Czech translation from Rudin, W. : Real and Complex Analysis., 3rd Ed. McGraw Hill, New York, 1987

[12] Shumway, R. H., Stoffer, D. S. : Time Series Analysis and Its Applications. Springer-Verlag, New York 2001

[13] Štěpán, J. : Teorie pravděpodobnosti. Matematické základy. Academia, Praha 1987. In Czech.

[14] WDC-SILSO, Royal Observatory of Belgium, Brussels, http://www.sidc.be/silso/datafiles

List of symbols

$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	set of nonnegative integers
$\mathbb{Z}$	set of integers
$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\boldsymbol{x}$	column vector
$\boldsymbol{A}$	matrix
$\boldsymbol{I}$	identity matrix
$\|\| \cdot \|\|$	norm in a Hilbert space
$\mathcal{B}$	Borel $\sigma-$algebra
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with parameters μ, σ^2
$X \sim \mathcal{N}(\mu, \sigma^2)$	random variable with distribution $\mathcal{N}(\mu, \sigma^2)$
$\{X_t,\, t \in T\}$	stochastic process indexed by set T
$\mathcal{M}\{X_t,\, t \in T\}$	linear span of $\{X_t,\, t \in T\}$
$\mathcal{H}\{X_t,\, t \in T\}$	Hilbert space generated by the stochastic process $\{X_t,\, t \in T\}$
AR(p)	autoregressive sequence of order p
MA(q0	moving average sequence of order q
ARMA(p, q)	mixed ARMA sequence of orders p and q
WN($0, \sigma^2$)	white noise with zero mean and variance σ^2
$X \perp Y$	orthogonal (perpendicular) random variables
$\overline{\lim}$	limes superior
$\xrightarrow{\text{P}}$	convergence in probability
$\xrightarrow{\text{D}}$	convergence in distribution
l. i. m.	convergence in mean square (limit in the mean)