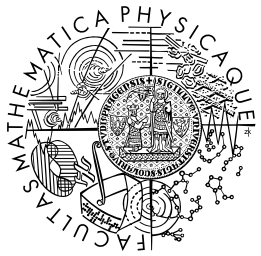

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

STOCHASTIC PROCESSES 2

Lecture notes

Jiří Dvořák

Department of Probabilty and Mathematical Statistics

Faculty of Mathematics and Physics

Charles University, Prague

Version 26/09/2026

Contents

1 Background on stochastic processes

1.1 Background and basics

Definition 1.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(S, \mathcal{E})$ a measurable space, and $T \subset \mathbb{R}, T \neq \emptyset$. A family of random variables $\{X_t, t \in T\}$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(S, \mathcal{E})$ is called* a stochastic process *or* a random process. *The space $(S, \mathcal{E})$ is called* the state space *of the process $\{X_t, t \in T\}$. For a fixed $\omega \in \Omega$, the function $X_\cdot(\omega) : t \mapsto X_t(\omega), t \in T$, is called* the trajectory *of $\{X_t, t \in T\}$.*

Remark: When $T \subseteq \mathbb{Z}$, we call $\{X_t, t \in T\}$ a discrete-time stochastic process (random sequence, time series). When T is an interval with endpoints a, b with $-\infty \leq a < b \leq \infty$, we call $\{X_t, t \in T\}$ a continuous-time stochastic process.

Remark: In the following we will consider mostly real-valued processes, i.e. we set $(S, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. However, it will often be useful to consider complex-valued processes, too.

Remark: Finite-dimensional distributions of a real-valued stochastic process $\{X_t, t \in T\}$ are given by the system of distribution functions

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n),\ n \in \mathbb{N}, t_1, \ldots, t_n \in T, x_1, \ldots, x_n \in \mathbb{R}.$$

Definition 1.2. *Let $\{F_{t_1,\ldots,t_n}, n \in \mathbb{N}, t_1, \ldots, t_n \in T\}$ be a system of functions such that for each $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in T$, $F_{t_1,\ldots,t_n} : \mathbb{R}^n \to [0, 1]$ is a distribution function. The system is said to be* consistent *if the following properties are fulfilled for each $n \geq 2, t_1, \ldots, t_n \in T, x_1, \ldots, x_n \in \mathbb{R}$:*

1. *$F_{t_{\pi(1)},\ldots,t_{\pi(n)}}(x_{\pi(1)}, \ldots, x_{\pi(n)}) = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$ for each permutation π on the set $\{1, \ldots, n\}$,*
2. *$\lim_{x_n \to \infty} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = F_{t_1,\ldots,t_{n-1}}(x_1, \ldots, x_{n-1})$.*

Remark: Any real-valued stochastic process determines a consistent system of distribution functions. The converse is discussed in the following theorem.

Theorem 1.1 (Daniell-Kolmogorov)**.** *Let $\{F_{t_1,\ldots,t_n}, n \in \mathbb{N}, t_1, \ldots, t_n \in T\}$ be a consistent system of distribution functions. Then there is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a stochastic process $\{X_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that for each $n \in \mathbb{N}, t_1, \ldots, t_n \in T$ and $x_1, \ldots, x_n \in \mathbb{R}$ it holds that*

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

Proof. See Štěpán (1987, Theorem I.9.4) or Kallenberg (2002, Theorem 6.16) for a proof of a more general version of the theorem. □

Example: Using the Daniell-Kolmogorov theorem, we can show the existence of the following random sequences or processes:

- the sequence $\{X_t, t \in \mathbb{Z}\}$ of independent, identically distributed random variables with standard normal distribution,
- the process $\{X_t, t \in \mathbb{R}\}$ of independent, identically distributed random variables with standard normal distribution,
- the Wiener process $\{W_t, t \geq 0\}$, also called the Brownian motion, see the Definition 1.8 below. Note that in this case, in addition to the Daniell-Kolmogorov theorem, we need the Kolmogorov continuity theorem to construct a continuous modification of the process.

Definition 1.3. *A real-valued 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* $(X_{t_1}, \dots, X_{t_n})^T$ *has the multivariate normal distribution* $\mathcal{N}_n(\mathbf{m_t}, \mathbf{V_t})$, *where* $\mathbf{m_t} = (\mathbb{E}X_{t_1}, \dots, \mathbb{E}X_{t_n})^T$ *and*

$$\mathbf{V_t} = \operatorname{var}(X_{t_1}, \dots, X_{t_n})^T = \begin{pmatrix} \operatorname{var} X_{t_1} & \operatorname{cov}(X_{t_1}, X_{t_2}) & \cdots & \operatorname{cov}(X_{t_1}, X_{t_n}) \\ \operatorname{cov}(X_{t_2}, X_{t_1}) & \operatorname{var} X_{t_2} & \cdots & \operatorname{cov}(X_{t_2}, X_{t_n}) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_{t_n}, X_{t_1}) & \operatorname{cov}(X_{t_n}, X_{t_2}) & \cdots & \operatorname{var} X_{t_n} \end{pmatrix}.$$

Remark: A random vector $\mathbb{Y} = (Y_1, \dots, Y_n)^T$ has the multivariate normal distribution $\mathcal{N}_n(\mu, \Sigma)$ if there is a vector $\mu \in \mathbb{R}^n$, an $n \times k$ matrix A and independent random variables $Z_1, \dots, Z_k$ with the univariate $\mathcal{N}(0,1)$ distribution such that $\Sigma = AA^T$ and $\mathbb{Y} = \mu + A\mathbb{Z}$, where $\mathbb{Z} = (Z_1, \dots, Z_k)^T$. Equivalently, $\mathbb{Y} = (Y_1, \dots, Y_n)^T$ has a multivariate normal distribution if for any $a_1, \dots, a_n \in \mathbb{R}$ the linear combination $a_1 Y_1 + \dots + a_n Y_n$ has a univariate normal distribution.

Complex-valued random variables

Complex-valued signals are useful e.g. in the fields of communications, optics, acoustics, and more. They may contain information about the amplitude and phase of a wave in a single object, making the analysis and manipulation of such signals more efficient and mathematically convenient.

A complex-valued random variable X is defined as $X = Y + iZ$, where Y, Z are real-valued random variables and $i = \sqrt{-1}$. Assuming that the expectations $\mathbb{E}Y, \mathbb{E}Z$ exist, we define $\mathbb{E}X = \mathbb{E}Y + i\mathbb{E}Z$.

If the random variables Y and Z have finite second moments, we define

$$\operatorname{var} X = \mathbb{E}(X - \mathbb{E}X)\overline{(X - \mathbb{E}X)} = \mathbb{E}|X - \mathbb{E}X|^2.$$

It follows that $\operatorname{var} X \geq 0$ and hence the variance of a complex variable is a (non-negative) real number. Similarly, the second moment of X is $\mathbb{E}|X|^2$.

Let X_1, X_2 be complex-valued random variables with finite second moments. Their covariance is defined as

$$\operatorname{cov}(X_1, X_2) = \mathbb{E}(X_1 - \mathbb{E}X_1)\overline{(X_2 - \mathbb{E}X_2)}.$$

Daniell-Kolmogorov theorem for complex-valued stochastic processes*

For complex-valued stochastic processes, we need to work with probability measures instead of distribution functions. There is a version of the Daniell-Kolmogorov theorem which covers this case, see below. Note that even more general versions of the theorem are also available.

Let $\{X_t, t \in T\}$ be a real-valued or complex-valued process. Its finite-dimensional distributions $\{\mathsf{P}_{t_1,\dots,t_n}, n \in \mathbb{N}, t_1, \dots, t_n \in T\}$ are defined by

$$\mathsf{P}_{t_1,\dots,t_n}(B_1 \times \dots \times B_n) = \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n)$$

for $n \in \mathbb{N}, t_1, \dots, t_n \in T$ and Borel sets $B_1, \dots, B_n$ (in $\mathbb{R}$ or $\mathbb{C}$). Naturally, this system fulfills the following properties for each $n \geq 2, t_1, \dots, t_n \in T$ and Borel sets $B_1, \dots, B_n$:

1. $\mathsf{P}_{t_{\pi(1)},\dots,t_{\pi(n)}}(B_{\pi(1)} \times \dots \times B_{\pi(n)}) = \mathsf{P}_{t_1,\dots,t_n}(B_1 \times \dots \times B_n)$ holds for each permutation π on the set $\{1, \dots, n\}$,
2. $\mathsf{P}_{t_1,\dots,t_n}(B_1 \times \dots \times B_{n-1} \times \mathbb{R}) = \mathsf{P}_{t_1,\dots,t_{n-1}}(B_1 \times \dots \times B_{n-1})$.

A system of probability measures $\{\mathsf{Q}_{t_1,\ldots,t_n}, n \in \mathbb{N}, t_1, \ldots, t_n \in T\}$, defined on appropriate spaces, is said to be *consistent* if it has the two properties given above.

Theorem 1.2 (Daniell-Kolmogorov)**.** *Let $\{\mathsf{Q}_{t_1,\ldots,t_n}, n \in \mathbb{N}, t_1 \ldots, t_n \in T\}$ be a consistent system of probability measures. Then there is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a stochastic process $\{X_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that for each $n \in \mathbb{N}, t_1, \ldots, t_n \in T$ and Borel sets $B_1, \ldots, B_n$ it holds that*

$$\mathbb{P}(X_{t_1} \in B_1 \ldots, X_{t_n} \in B_n) = \mathsf{Q}_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n).$$

Proof. See Štěpán (1987, Theorem I.9.4) or Kallenberg (2002, Theorem 6.16) for a proof of a more general version of the theorem. □

1.2 Important examples of stochastic processes

Markov processes

Definition 1.4. *Let $S \subset \mathbb{Z}$ be a discrete state space. A stochastic process $\{X_t, t \in T\}$ with the state space $(S, \mathcal{E})$ is* Markov *if for each $n \in \mathbb{N}, i_0, \ldots, i_n \in S$ and $t_0, \ldots, t_n \in T$ such that $t_0 < t_1 < \ldots < t_n$ it holds that*

$$\mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_0} = i_0) = \mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}).$$

Independent increment processes

Definition 1.5. *A (real-valued or complex-valued) stochastic process $\{X_t, t \in T\}$ has* independent increments *if for any $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in T$ such that $t_1 < t_2 < \ldots < t_n$, the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*

Definition 1.6. *A (real-valued or complex-valued) stochastic process $\{X_t, t \in T\}$ has* stationary increments *if for any $s, t \in T, s < t$, the distribution of $X_t - X_s$ depends only on $t - s$. Formally, the random variables $X_t - X_s$ and $X_{t+h} - X_{s+h}$ have the same distribution for any $s, t \in T$ and $h \in \mathbb{R}$ such that $s + h, t + h \in T$.*

Definition 1.7. *The* Poisson process *$\{N_t, t \geq 0\}$ with parameter $\lambda > 0$ is defined by the following properties:*

- *$N_0 = 0$ a.s.,*
- *the process has independent increments,*
- *for each $0 \leq s < t$ the random variable $N_t - N_s$ has the Poisson distribution $\mathrm{Po}(\lambda(t - s))$.*

Remark: The last property implies that $\{N_t, t \geq 0\}$ has stationary increments and (in combination with the first property) that for each $t \geq 0$ we have $N_t \sim \mathrm{Po}(\lambda t)$. Furthermore, the Poisson process is Markov, but it is not strictly nor weakly stationary, as indicated e.g. by the non-constant mean and variance. Sample realizations of the Poisson process are given in Figure 1.

Definition 1.8. *The* Wiener process *$\{W_t, t \geq 0\}$ with parameter $\sigma^2 > 0$ is defined by the following properties:*

- *$W_0 = 0$ a.s.,*
- *the process has continuous trajectories a.s.,*

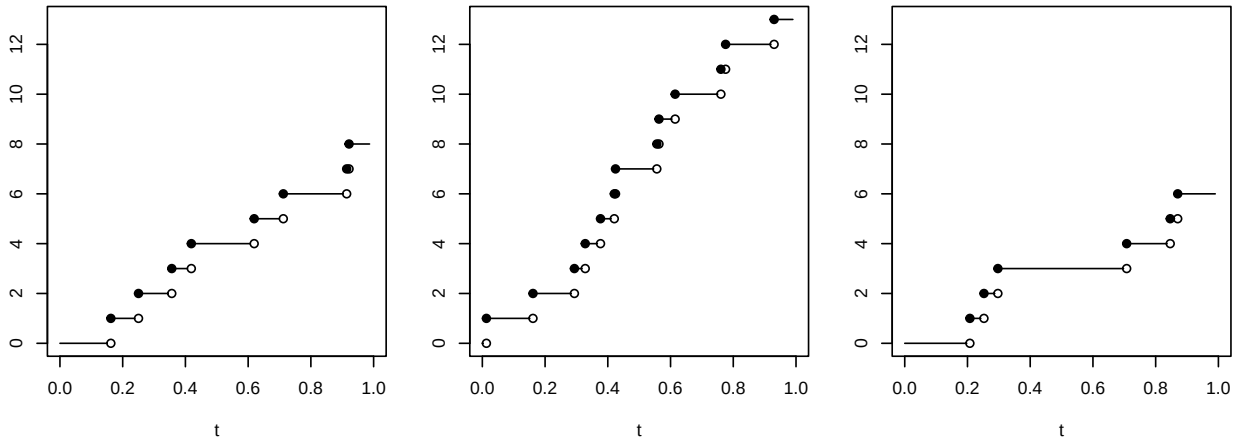

Figure 1: Sample realizations of the Poisson process with $\lambda = 10$.

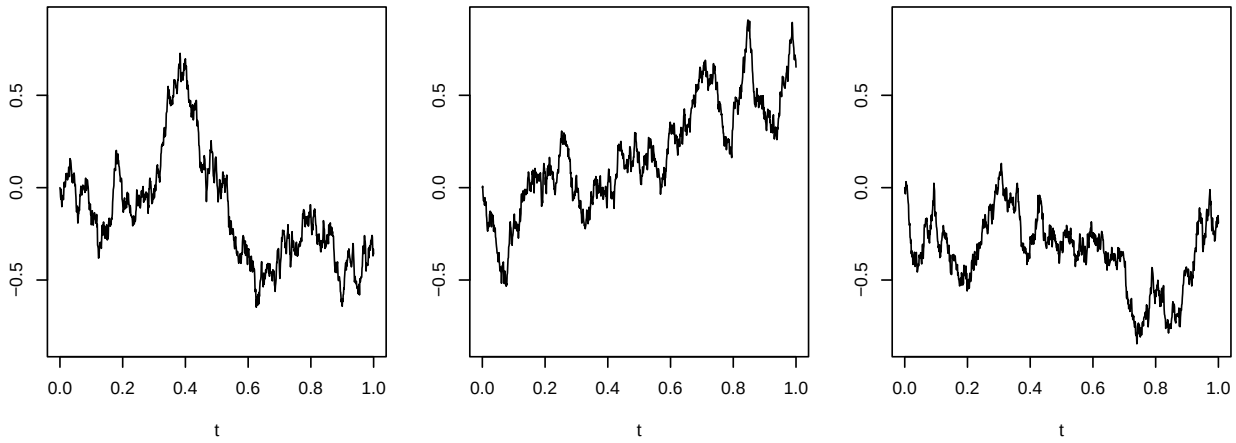

Figure 2: Sample realizations of the Wiener process with $\sigma^2 = 1$.

- *the process has independent increments,*
- *for each* $0 \leq s < t$ *the random variable* $W_t - W_s$ *has the normal distribution* $\mathcal{N}(0, \sigma^2(t-s))$.

Remark: The last property implies that $\{W_t, t \geq 0\}$ has stationary increments and (in combination with the first property) that for each $t \geq 0$ we have $W_t \sim \mathcal{N}(0, \sigma^2 t)$. The definition also implies that the Wiener process is Gaussian – this can be easily proved using the fact that a sum of two independent Gaussian random variables is also Gaussian. Furthermore, the Wiener process is Markov, but it is not strictly nor weakly stationary, as indicated e.g. by the non-constant variance. Sample realizations of the Wiener process are given in Figure 2.

Martingales

Definition 1.9. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space,* $T \subset \mathbb{R}, T \neq \emptyset$. *For any* $t \in T$, *let* $\mathcal{F}_t \subset \mathcal{A}$ *be a* σ*-algebra. A system of* σ*-algebras* $\{\mathcal{F}_t, t \in T\}$ *such that* $\mathcal{F}_s \subseteq \mathcal{F}_t$ *for each* $s, t \in T, s < t$, *is called* a filtration.

Definition 1.10. *Let* $\{X_t, t \in T\}$ *be a stochastic process defined on* $(\Omega, \mathcal{A}, \mathbb{P})$ *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 each* $t \in T$, X_t *is* $\mathcal{F}_t$*-measurable.*

Remark: In the previous definition, the filtration carries the information about the possible history of the process. $\mathcal{F}_t$ contains all the events that could possibly happen up to (and including) time $t \in T$.

Definition 1.11. *Let $\{X_t, t \in T\}$ be adapted to $\{\mathcal{F}_t, t \in T\}$ and $\mathbb{E}|X_t| < \infty$ for each $t \in T$. Then $\{X_t, t \in T\}$ is said to be* $\mathcal{F}_t$-martingale *if $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ almost surely for each $s, t \in T, s < t$.*

Remark: Martingales are useful e.g. for modeling a series of fair games or in survival analysis.

Remark: Wiener process is a martingale.

2 Autocovariance function and stationarity

2.1 Autocovariance and autocorrelation function

Definition 2.1. *Let $\{X_t, t \in T\}$ be a real-valued or complex-valued stochastic process. Let $\mu_t = \mathbb{E}X_t$ exist for each $t \in T$. The function $\{\mu_t, t \in T\}$ defined on T is* the mean value *of the process $\{X_t, t \in T\}$. If $\mu_t = 0$ for each $t \in T$, we say that the process is* centered.

Definition 2.2. *Let $\{X_t, t \in T\}$ be a real-valued or complex-valued stochastic process with finite second moments, i.e. $\mathbb{E}|X_t|^2 < \infty$ for each $t \in T$. The function $R : T \times T \to \mathbb{C}$ defined as*

$$R(s,t) = \operatorname{cov}(X_s, X_t) = \mathbb{E}(X_s - \mu_s)\overline{(X_t - \mu_t)}, \ s,t \in T,$$

is called the autocovariance function *of the process $\{X_t, t \in T\}$.*

Remark: As a special case we have $R(t,t) = \operatorname{cov}(X_t, X_t) = \operatorname{var} X_t = \mathbb{E}|X_t - \mu_t|^2, t \in T$, i.e. $R(t,t)$ is the variance of the process at time t.

Remark: Two different stochastic processes may have the same autocovariance functions.

Definition 2.3. *Let $\{X_t, t \in T\}$ be a real-valued or complex-valued stochastic process with finite second moments and positive variances.* The autocorrelation function *of the process $\{X_t, t \in T\}$ is defined as*

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

2.2 Strict and weak stationarity

Definition 2.4. *A real-valued (or complex-valued) stochastic process $\{X_t, t \in T\}$ is* strictly stationary *if for any $n \in \mathbb{N}$, any Borel subsets $B_1, \ldots, B_n$ of $\mathbb{R}$ (or $\mathbb{C}$), any $t_1, \ldots, t_n \in T$ and $h \in \mathbb{R}$ such that $t_1 + h, \ldots, t_n + h \in T$ it holds that*

$$\mathbb{P}(X_{t_1+h} \in B_1, \ldots, X_{t_n+h} \in B_n) = \mathbb{P}(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n).$$

Remark: For a real-valued stochastic process $\{X_t, t \in T\}$, strict stationarity is equivalently defined using the property $F_{t_1+h,\ldots,t_n+h}(x_1, \ldots, x_n) = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$ for each $n \in \mathbb{N}, x_1, \ldots, x_n \in \mathbb{R}, t_1, \ldots, t_n \in T$ and $h \in \mathbb{R}$ such that $t_1 + h, \ldots, t_n + h \in T$.

Example: Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of independent, identically distributed random variables. The sequence is strictly stationary. The proof follows easily from the definition of strict stationarity, independence, and the fact that the random variables have the same distribution.

Remark: If a process $\{X_t, t \in T\}$ is strictly stationary, all random variables X_t have the same distribution, i.e. their properties do not change over time. This applies to the expectation, variance, and covariances (provided they exist), but also to all higher-order moments or any other properties. This can be too restrictive to be useful in many settings.

Definition 2.5. *A stochastic process $\{X_t, t \in T\}$ with finite second moments is* weakly stationary *or* second-order stationary *if $\mu_t = \mu$ for each $t \in T$ and its autocovariance function fulfills $R(s + h, t + h) = R(s,t)$ for each $s,t \in T$ and $h \in \mathbb{R}$ such that $s + h, t + h \in T$. If only the latter property holds, we call the process* covariance stationary.

Remark: Weak stationarity of a stochastic process implies that the first- and second-order moment properties do not change over time. This is in contrast with strict stationarity which implies that all finite-dimensional distributions do not change over time.

Example: Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of uncorrelated random variables with zero mean and finite positive variance $\operatorname{var} X_t = \sigma^2$, the same for all $t \in \mathbb{Z}$. The sequence is weakly stationary. The sequence is called *the white noise* and denoted $WN(0, \sigma^2)$, see Definition 7.1. It will serve as an important example and building block for constructing various models of time series.

Remark: The autocovariance function of a weakly stationary process can be defined as a function of a single argument: $\widetilde{R}(t) = R(t, 0) = R(t + h, h)$ for appropriate values of t, h. Without the risk of confusion, we will write $R(t)$ in place of $\widetilde{R}(t)$. Similarly, the autocorrelation function is then $r(t) = R(t)/R(0)$ for $t \in T - T$ (the set of differences of elements of T).

Theorem 2.1. *Let $\{X_t, t \in T\}$ be a strictly stationary stochastic process with finite second moments. Then $\{X_t, t \in T\}$ is weakly stationary.*

Proof. Strict stationarity of $\{X_t, t \in T\}$ implies that X_{t_1} and X_{t_2} have the same distribution for any $t_1, t_2 \in T$, i.e. $\mathbb{P}(X_{t_1} \in B) = \mathbb{P}(X_{t_2} \in B)$ for any Borel set B. Since the sequence has finite second moments, $\mathbb{E}X_t$ exists and is finite for each $t \in T$ and it follows that $\mathbb{E}X_{t_1} = \mathbb{E}X_{t_2} = \mu$ for each $t_1, t_2 \in T$.

Similarly, strict stationarity implies that $(X_{t_1}, X_{t_2})^T$ and $(X_{t_1+h}, X_{t_2+h})^T$ have the same distribution for each $t_1, t_2 \in T$ and $h \in \mathbb{R}$ such that $t_1 + h, t_2 + h \in T$. Let R be the autocovariance function of $\{X_t, t \in T\}$. It follows that

$$R(t_1, t_2) = \operatorname{cov}(X_{t_1}, X_{t_2}) = \operatorname{cov}(X_{t_1+h}, X_{t_2+h}) = R(t_1 + h, t_2 + h) = R(t_1 - t_2).$$

□

Remark: The opposite implication does not hold without additional assumptions, see the following example.

Example: Let X be a random variable with $\mathbb{P}(X = -1/4) = 3/4$, $\mathbb{P}(X = 3/4) = 1/4$, and define $X_t = (-1)^t \cdot X, t \in \mathbb{Z}$. The sequence $\{X_t, t \in \mathbb{Z}\}$ is weakly stationary but it is not strictly stationary.

Theorem 2.2. *Let $\{X_t, t \in T\}$ be a real-valued, weakly stationary Gaussian process. Then $\{X_t, t \in T\}$ is strictly stationary.*

Proof. We fix $n \in \mathbb{N}, t_1, \ldots, t_n \in T$ and $h \in \mathbb{R}$ such that $t_1 + h, \ldots, t_n + h \in T$. Weak stationarity implies that

$$\mathbb{E}(X_{t_1}, \ldots, X_{t_n})^T = \mathbb{E}(X_{t_1+h}, \ldots, X_{t_n+h})^T = (\mu, \ldots, \mu)^T = \boldsymbol{\mu},$$
$$\operatorname{var}(X_{t_1}, \ldots, X_{t_n})^T = \operatorname{var}(X_{t_1+h}, \ldots, X_{t_n+h})^T = \Sigma,$$

where the variance matrix Σ consists of the elements $\Sigma_{ij} = R(t_i - t_j)$. Since the process $\{X_t, t \in T\}$ is Gaussian, both vectors $(X_{t_1}, \ldots, X_{t_n})^T$ and $(X_{t_1+h}, \ldots, X_{t_n+h})^T$ have n-dimensional normal distribution with vector of mean values $\boldsymbol{\mu}$ and variance matrix Σ, meaning they have the same distribution. □

Remark: Assuming strict or weak stationarity or some specific form of the mean value function is a modeling choice. In practice, we observe a realization of a random sequence or process without knowing its theoretical properties.

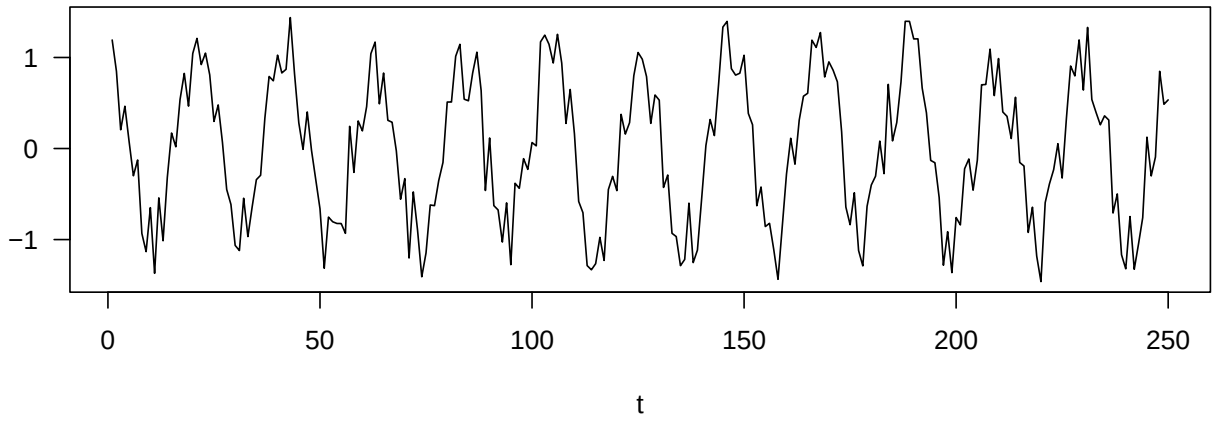

Figure 3: Observation of a simulated random sequence. Note that this is a discrete-time sequence, and the lines joining the corresponding points in the plot are used only for clarity.

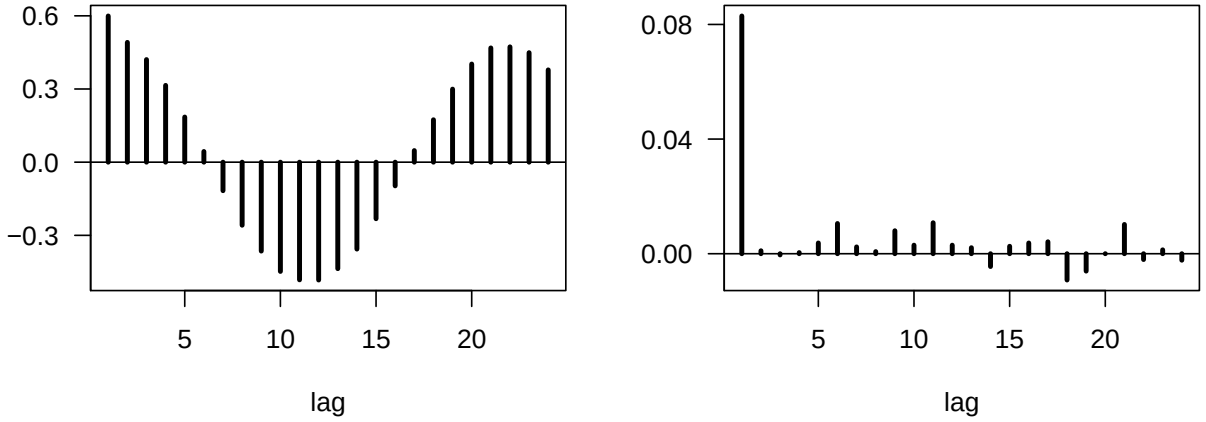

Figure 4: Estimate of the autocovariance function of the random sequence from Figure 3 assuming a zero mean (left) or assuming an appropriate form of the non-constant mean value function (right).

For illustration, consider the simulated random sequence $\{X_t, t \in \mathbb{Z}\}$ given in Figure 3. The sequence clearly contains a periodic component. In the modeling step, the periodic behavior can be attributed to either the mean value function $\{\mu_t, t \in \mathbb{Z}\}$ or the autocovariance function. We may decide to assume a zero mean, and the resulting estimate of the autocovariance function is given in the left panel of Figure 4. Alternatively, we may assume that $\mu_t = \cos(0.3t), t \in \mathbb{Z}$, and the corresponding estimate of the autocovariance function of the centered sequence $\{X_t - \mu_t, t \in \mathbb{Z}\}$ is given in the right panel of Figure 4. Of course, other modeling choices are also possible. We remark that the estimation of the autocovariance function will be discussed in Section 11.2.

2.3 Properties of autocovariance functions

Theorem 2.3. *Let $\{X_t, t \in T\}$ be a stochastic process with finite second moments. Its autocovariance function satisfies the following properties:*

- $R(t,t) \geq 0,\ t \in T,$
- $|R(s,t)| \leq \sqrt{R(s,s)}\sqrt{R(t,t)},\ s,t \in T.$

Proof. For any $t \in T$ we have $R(t,t) = \operatorname{cov}(X_t, X_t) = \operatorname{var} X_t = \mathbb{E}|X_t - \mathbb{E}X_t|^2 \geq 0$. Furthermore, it follows from the Cauchy-Schwarz inequality that for any $s, t \in T$,

$$|R(s,t)| = |\mathbb{E}(X_s - \mathbb{E}X_s)\overline{(X_t - \mathbb{E}X_t)}| \leq \left(\mathbb{E}|X_s - \mathbb{E}X_s|^2\right)^{1/2} \left(\mathbb{E}|X_t - \mathbb{E}X_t|^2\right)^{1/2} = \sqrt{R(s,s)}\sqrt{R(t,t)}.$$

□

Remark: The second property implies that $|r(s,t)| \leq 1, s, t \in T$. Furthermore, for a weakly stationary process we have $R(0) \geq 0, |R(t)| \leq R(0), t \in T$.

Definition 2.6. *Let f be a complex-valued function defined on $T \times T, T \subset \mathbb{R}$. It is* positive semidefinite *if for each $n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{C}$ and $t_1, \ldots, t_n \in T$ it holds that*

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

The inequality above implies that the double sum takes a real value (i.e. its imaginary part is 0). Furthermore, let g be a complex-valued function defined on $T, T \subset \mathbb{R}$. It is positive semidefinite *if for each $n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{C}$ and $t_1, \ldots, t_n \in T$ it holds that*

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

Definition 2.7. *Let f be a complex-valued function defined on $T \times T, T \subset \mathbb{R}$. It is* Hermitian *if it holds that $f(s,t) = \overline{f(t,s)}, s, t \in T$. For a real-valued function f this is the symmetry property: $f(s,t) = f(t,s), s, t \in T$. Furthermore, let g be a complex-valued function defined on $T, T \subset \mathbb{R}$. It is* Hermitian *if it holds that $g(-t) = \overline{g(t)}, t \in T$. For a real-valued function g this is again the symmetry property: $g(-t) = g(t), t \in T$.*

Theorem 2.4. *Let f be a positive semidefinite function defined on $T \times T, T \subset \mathbb{R}$. Then f is Hermitian.*

Proof. The Hermitian property follows from the positive semidefinite property by choosing $n = 1, c_1 = 1$; $n = 2, c_1 = 1, c_2 = 1$; and finally $n = 2, c_1 = 1, c_2 = i$. Note that for a complex number u, $u \geq 0$ implies $u \in \mathbb{R}$, i.e. the imaginary part of u is 0. □

Theorem 2.5. *Assuming $\{X_t, t \in T\}$ is a stochastic process with finite second moments, its autocovariance function is positive semidefinite on $T \times T$.*

Proof. Without loss of generality, assume that $\{X_t, t \in T\}$ is centered (changing the mean value does not change the autocovariance function). In this case $R(s,t) = \mathbb{E}X_s\overline{X}_t$, $s, t \in T$.

We choose $n \in \mathbb{N}, t_1, \ldots, t_n \in T, c_1, \ldots, c_n \in \mathbb{C}$. Then

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

□

Theorem 2.6. *Let R be a (complex-valued) positive semidefinite function on $T \times T$. There is a stochastic process $\{X_t, t \in T\}$ with finite second moments such that its autocovariance function is R.*

Proof. We show the proof for a real-valued function R. For the proof with a complex-valued function R see e.g. Loève (1963, Chapter X, §34).

Choose $n \in \mathbb{N}, t_1, \ldots, t_n \in T$. Since R is positive semidefinite, the matrix

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

is also positive semidefinite, and hence it is a valid variance matrix. This means that for $n \in \mathbb{N}, t_1, \ldots, t_n \in T$, we can consider the n-dimensional Gaussian distribution $\mathcal{N}_n(\mathbf{0}, \mathbf{V_t})$. From these we obtain a consistent system of distribution functions. It follows from Theorem 1.2 that there is a stochastic process $\{X_t, t \in T\}$ with those finite-dimensional distributions. Clearly, the autocovariance function of $\{X_t, t \in T\}$ is R. □

Example: The function $R(t) = \cos t, t \in \mathbb{R}$, is positive semidefinite and hence it is the autocovariance function of a stochastic process. To prove that, we first write R as a function of two variables: $R(t_1, t_2) = R(t_1 - t_2) = \cos(t_1 - t_2), t_1, t_2 \in \mathbb{R}$. Then, for a given $n \in \mathbb{N}, t_1, \ldots, t_n \in \mathbb{R}, c_1, \ldots, c_n \in \mathbb{C}$,

$$\begin{aligned} \sum_{j=1}^{n}\sum_{k=1}^{n} c_j\overline{c_k}\cos(t_j - t_k) &= \sum_{j=1}^{n}\sum_{k=1}^{n} c_j\overline{c_k}(\cos t_j \cdot \cos t_k + \sin t_j \cdot \sin t_k) \\ &= \left(\sum_{j=1}^{n} c_j \cos t_j\right)\overline{\left(\sum_{k=1}^{n} c_k \cos t_k\right)} + \left(\sum_{j=1}^{n} c_j \sin t_j\right)\overline{\left(\sum_{k=1}^{n} c_k \sin t_k\right)} \\ &= \left|\sum_{j=1}^{n} c_j \cos t_j\right|^2 + \left|\sum_{j=1}^{n} c_j \sin t_j\right|^2 \geq 0. \end{aligned}$$

Theorem 2.7. *Let f, g be positive semidefinite functions on $T \times T$. Then, the sum $f + g$ is also positive semidefinite.*

Proof. Denote $h = f + g$ and choose $n \in \mathbb{N}, t_1, \ldots, t_n \in T, c_1, \ldots, c_n \in \mathbb{C}$. Then

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

□

Remark: The previous theorems imply that a sum of two autocovariance functions is the autocovariance function of some stochastic process with finite second moments.

3 Space $L_2(\Omega, \mathcal{A}, \mathbb{P})$

3.1 Hilbert spaces

Definition 3.1. *Let H be a complex vector space. Assume that the mapping $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ fulfills the following properties for each $x, y, z \in H$ and each $\alpha \in \mathbb{C}$:*

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$,
- $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$,
- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$,
- $\langle x, x \rangle \geq 0$,
- $\langle x, x \rangle = 0 \iff x = o \in H$.

The mapping is called an inner product *and H is called* an inner product space.

Definition 3.2. *Let H be an inner product space and $\langle \cdot, \cdot \rangle$ be the corresponding inner product. The* norm *of $x \in H$ is defined as $\|x\| = \sqrt{\langle x, x \rangle}$.*

Theorem 3.1. *Let H be an inner product space and $\| \cdot \|$ be the corresponding norm. Then the following properties hold for each $x, y \in H$ and each $\alpha \in \mathbb{C}$:*

- $\|x\| \geq 0$,
- $\|x\| = 0 \iff x = o \in H$,
- $\|\alpha x\| = |\alpha| \cdot \|x\|$,
- $\|x + y\| \leq \|x\| + \|y\|$ *... triangle inequality,*
- $|\langle x, y \rangle| \leq \|x\| \cdot \|y\| = \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle}$ *... Cauchy-Schwarz inequality.*

Proof. See any textbook on functional analysis, e.g. Rudin (2003, Chapter 4). □

Definition 3.3. *Let $\{x_n, n \in \mathbb{N}\}$ be a sequence of elements of an inner product space H. The sequence* converges in norm *to $x \in H$ if $\|x_n - x\| \to 0, n \to \infty$.*

Definition 3.4. *Let $\{x_n, n \in \mathbb{N}\}$ be a sequence of elements of an inner product space H. The sequence is* Cauchy *if $\|x_n - x_m\| \to 0, n, m \to \infty$.*

Remark: Note that no limit element is needed for the definition of the Cauchy property.

Definition 3.5. *An inner product space H is* complete *if every Cauchy sequence of its elements converges in norm to some element of H. A complete inner product space is called* a Hilbert space.

Theorem 3.2 (Continuity of the inner product)**.** *Let $\{x_n, n \in \mathbb{N}\}$ and $\{y_n, n \in \mathbb{N}\}$ be sequences of elements of an inner product space H. Let $x, y \in H$ be such that $x_n \to x, y_n \to y, n \to \infty$, in norm. Then*

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

Proof. From the triangle inequality we get

$$\|x\| = \|(x - x_n) + x_n\| \leq \|x - x_n\| + \|x_n\|,$$
$$\|x_n\| = \|(x_n - x) + x\| \leq \|x_n - x\| + \|x\|,$$

and in combination

$$\big|\|x_n\| - \|x\|\big| \leq \|x - x_n\| \to 0,\ n \to \infty.$$

Concerning the second claim, it is proved using the triangle inequality and the Cauchy-Schwarz inequality:

$$\begin{aligned}
|\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n - x + x, y_n - y + y\rangle - \langle x, y\rangle| \\
&= |\langle x_n - x, y_n - y\rangle + \langle x_n - x, y\rangle + \langle x, y_n - y\rangle + \langle x, y\rangle - \langle x, y\rangle| \\
&\leq |\langle x_n - x, y_n - y\rangle| + |\langle x_n - x, y\rangle| + |\langle x, y_n - y\rangle| \\
&\leq \|x_n - x\| \cdot \|y_n - y\| + \|x_n - x\| \cdot \|y\| + \|x\| \cdot \|y_n - y\| \to 0,\ n \to \infty.
\end{aligned}$$

□

3.2 Construction and properties of $L_2(\Omega, \mathcal{A}, \mathbb{P})$

Let $\mathcal{L}$ be the set of all complex-valued random variables with finite second moment defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. $\mathcal{L}$ is a complex vector space and the random variable $X \equiv 0$ is the zero element.

For $X, Y \in \mathcal{L}$ we would like to define $\langle X, Y\rangle = \mathbb{E}X\overline{Y}$ and use it as the inner product on $\mathcal{L}$. However, this is not possible since it does not hold that $\langle X, X\rangle = 0 \iff X \equiv 0$. In fact, any random variable Y for which $Y = 0$ almost surely fulfills $\langle Y, Y\rangle = 0$.

For the above reason, we first define equivalence classes on $\mathcal{L}$ so that $X \sim Y \iff \mathbb{P}(X = Y) = 1$. We call the space of equivalence classes $L_2(\Omega, \mathcal{A}, \mathbb{P})$.

At this point, we define the mapping $\langle X, Y\rangle = \mathbb{E}X_0\overline{Y_0}$, where $X, Y \in L_2(\Omega, \mathcal{A}, \mathbb{P})$ are equivalence classes and $X_0 \in X, Y_0 \in Y$ are random variables (representants of the equivalence classes). Clearly, the value $\langle X, Y\rangle$ does not depend on the choice of the representants X_0, Y_0. In the following we will work with the equivalence classes and stop distinguishing different representants.

The mapping $\langle \cdot, \cdot \rangle$ is an inner product on $L_2(\Omega, \mathcal{A}, \mathbb{P})$. The norm on $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is given by $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{E}X\overline{X}} = \big(\mathbb{E}|X|^2\big)^{1/2}$. The convergence on $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is defined as the convergence in this norm.

Definition 3.6. *Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of random variables such that $\mathbb{E}|X_n|^2 < \infty$. The sequence* converges in the mean square *(or* in L_2*) to a random variable X if it converges to X in $L_2(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $\|X_n - X\|^2 = \mathbb{E}|X_n - X|^2 \to 0, n \to \infty$.*

Remark: The mean square limit X of $\{X_n, n \in \mathbb{N}\}$ is sometimes denoted $X = \text{l.i.m.}\, X_n$ (limit in the mean).

Theorem 3.3. *The space $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is complete with respect to the norm defined above. Hence, $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space.*

Proof. See Brockwell and Davis (2006, Section 2.10) or Rudin (2003, Theorem 3.11). □

Remark: Mean square convergence implies convergence of the first and second moments. Formally, for $X, X_1, X_2, \ldots \in L_2(\Omega, \mathcal{A}, \mathbb{P})$ it holds that $X_n \to X, n \to \infty$, in the mean square $\Rightarrow \mathbb{E}X_n \to \mathbb{E}X, \mathbb{E}|X_n|^2 \to \mathbb{E}|X|^2, n \to \infty$. If also $Y_n \to Y, n \to \infty$, in the mean square, we have $\mathbb{E}X_n\overline{Y_n} \to \mathbb{E}X\overline{Y}, n \to \infty$. All these properties follow from the continuity of the inner product.

3.3 Hilbert space generated by a stochastic process

Definition 3.7. *Let $\{X_t, t \in T\}$ be a stochastic process with finite second moments defined on $(\Omega, \mathcal{A}, \mathbb{P})$.* The linear span *of $\{X_t, t \in T\}$ is the set of all finite linear combinations of the random variables:*

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

Remark: $\mathcal{M}\{X_t, t \in T\} \subset L_2(\Omega, \mathcal{A}, \mathbb{P})$, and the equivalence classes, inner product, and convergence are defined as above.

Definition 3.8. The closure $\overline{\mathcal{M}}\{X_t, t \in T\}$ *of $\mathcal{M}\{X_t, t \in T\}$ consists of all the elements of $\mathcal{M}\{X_t, t \in T\}$ and the mean square limits of all Cauchy sequences of elements of $\mathcal{M}\{X_t, t \in T\}$.*

Remark: $\overline{\mathcal{M}}\{X_t, t \in T\}$ is a closed subset of the complete inner product space $L_2(\Omega, \mathcal{A}, \mathbb{P})$ and thus it is a complete inner product space. It is called *the Hilbert space generated by a stochastic process* $\{X_t, t \in T\}$ and denoted $\mathcal{H}\{X_t, t \in T\}$.

3.4 Convergence of processes in $L_2(\Omega, \mathcal{A}, \mathbb{P})$

Definition 3.9. *Let $\{X_t^n, t \in T\}_{n \in \mathbb{N}}$ be a collection of stochastic processes in $L_2(\Omega, \mathcal{A}, \mathbb{P})$. The processes $\{X_t^n, t \in T\}_{n \in \mathbb{N}}$ converge in the mean square to a process $\{X_t, t \in T\}$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ as $n \to \infty$, if for each $t \in T$ it holds that $X_t^n \to X_t, n \to \infty$, in the mean square, i.e. $\mathbb{E}|X_t^n - X_t|^2 \to 0, n \to \infty$.*

Theorem 3.4. *Centered processes $\{X_t^n, t \in T\}_{n \in \mathbb{N}}$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ converge in the mean square to a centered process $\{X_t, t \in T\}$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ as $n \to \infty$ if and only if*

$$\mathbb{E} X_t^n \overline{X_t^m} \to b(t), \ n, m \to \infty,$$

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

If the processes $\{X_t^n, t \in T\}_{n \in \mathbb{N}}$ converge to a process $\{X_t, t \in T\}$ in the mean square as $n \to \infty$, the autocovariance functions of the processes $\{X_t^n, t \in T\}_{n \in \mathbb{N}}$ converge to the autocovariance function of $\{X_t, t \in T\}$ as $n \to \infty$.

Proof. 1) Assume that $\{X_t^n, t \in T\}_{n \in \mathbb{N}} \to \{X_t, t \in T\}$, $n \to \infty$, in the mean square. Then for each $t, t' \in T$, $X_t^n \to X_t$, $n \to \infty$, in the mean square, and $X_{t'}^m \to X_{t'}$, $m \to \infty$, in the mean square. Continuity of the inner product gives

$$\mathbb{E} X_t^n \overline{X_{t'}^m} = \langle X_t^n, X_{t'}^m \rangle \to \langle X_t, X_{t'} \rangle, \ n, m \to \infty.$$

We consider the following special cases:

- For $t = t'$ and $n, m \to \infty$ we have

  $$\mathbb{E} X_t^n \overline{X_t^m} \to \mathbb{E} X_t \overline{X_t} = \mathbb{E}|X_t|^2 = b(t) < \infty,$$

  since $\{X_t, t \in T\}$ is in $L_2(\Omega, \mathcal{A}, \mathbb{P})$. This proves the first part of the first claim.

- For $n = m$ and $n \to \infty$ we have

  $$\mathbb{E} X_t^n \overline{X_{t'}^n} \to \mathbb{E} X_t \overline{X_{t'}}, \ t, t' \in T.$$

  Since the processes are centered, this means that $R_n(t, t') \to R(t, t')$, $n \to \infty$, where R_n is the autocovariance function of the process $\{X_t^n, t \in T\}$ and R is the autocovariance function of the process $\{X_t, t \in T\}$. This proves the second claim.

2) Assume that $\{X_t^n, t \in T\}_{n\in\mathbb{N}}$ are centered processes in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ for which $\mathbb{E}X_t^n\overline{X_t^m} \to b(t) < \infty$, $n, m \to \infty$, for each $t \in T$. For any fixed $t \in T$, the sequence $\{X_t^n\}_{n\in\mathbb{N}}$ has the Cauchy property since

$$\|X_t^n - X_t^m\|^2 = \mathbb{E}(X_t^n - X_t^m)\overline{(X_t^n - X_t^m)} = \mathbb{E}X_t^n\overline{X_t^n} - \mathbb{E}X_t^n\overline{X_t^m} - \mathbb{E}X_t^m\overline{X_t^n} + \mathbb{E}X_t^m\overline{X_t^m} \to$$
$$\to b(t) - b(t) - b(t) + b(t) = 0,\ n, m \to \infty.$$

The space $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is complete and hence for each $t \in T$ there is a random variable $X_t \in L_2(\Omega, \mathcal{A}, \mathbb{P})$ such that $X_t^n \to X_t$, $n \to \infty$, in the mean square. Therefore, there is a limiting process $\{X_t, t \in T\}$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$.

Finally, we prove that $\{X_t, t \in T\}$ is centered. Fix $t \in T$. Since $\mathbb{E}X_t^n = 0$, continuity of the inner product (Theorem 3.2) implies that also $\mathbb{E}X_t = 0$, see the remark below Theorem 3.3. This proves the second part of the first claim. □

Remark: The proof of the previous theorem shows that $b(t) = \mathbb{E}|X_t|^2 = \operatorname{var} X_t$ is the variance of the limiting process $\{X_t, t \in T\}$.

Theorem 3.5. *Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of random variables such that $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ with $\mu_n \in \mathbb{R}, \sigma_n^2 \geq 0, n \in \mathbb{N}$. Assume that there is $X \in L_2(\Omega, \mathcal{A}, \mathbb{P})$ such that $X_n \to X, n \to \infty$, in the mean square. Then $X \sim \mathcal{N}(\mu, \sigma^2)$, where $\mu_n \to \mu, \sigma_n^2 \to \sigma^2, n \to \infty$.*

Proof. The existence of the mean square limit and the continuity of the inner product gives

$$\mu_n = \mathbb{E}X_n \to \mathbb{E}X,\ n \to \infty,$$
$$\sigma_n^2 = \operatorname{var} X_n \to \operatorname{var} X,\ n \to \infty.$$

We denote $\mu = \mathbb{E}X, \sigma^2 = \operatorname{var} X$. For a given $n \in \mathbb{N}$, the characteristic function of X_n is

$$\varphi_{X_n}(t) = \exp\left\{i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right\},\ t \in \mathbb{R}.$$

For $n \to \infty$, we have the pointwise convergence

$$\varphi_{X_n}(t) \to \varphi(t) = \exp\left\{i\mu t - \frac{1}{2}\sigma^2 t^2\right\},\ t \in \mathbb{R}.$$

The limiting function $\varphi(t)$ is continuous at the origin (in fact, it is continuous everywhere). Lévy's continuity theorem gives us that $X_n \xrightarrow{d} Z, n \to \infty$, where Z has the characteristic function φ, i.e. $Z \sim \mathcal{N}(\mu, \sigma^2)$. On the other hand, we already know that $X_n \to X, n \to \infty$, in the mean square, and hence $X_n \xrightarrow{d} X, n \to \infty$. Uniqueness of the limit implies that $X \sim \mathcal{N}(\mu, \sigma^2)$. □

Remark: The interesting part of the previous theorem is that the normal distribution is preserved by the mean square convergence.

4 Mean square continuity

Definition 4.1. *Let $\{X_t, t \in T\}$ be a stochastic process with finite second moments, $T \subset \mathbb{R}$ an open interval. The process $\{X_t, t \in T\}$ is* mean square continuous *(or L_2-*continuous*)* at point $t_0 \in T$, *if $\mathbb{E}|X_t - X_{t_0}|^2 \to 0, t \to t_0$, i.e. $X_t \to X_{t_0}, t \to t_0$, in the mean square. The process $\{X_t, t \in T\}$ is* mean square continuous *if it is mean square continuous at each point of T.*

Remark: A stochastic process $\{X_t, t \in T\}$ with finite second moments which is mean square continuous is also continuous in probability, i.e. $X_t \to X_{t_0}, t \to t_0$, in probability.

Theorem 4.1. *Let $\{X_t, t \in T\}$ be a centered stochastic process with finite second moments, $T \subset \mathbb{R}$ an open 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 $[s,t] \in T \times T$ such that $s = t$ (as a function of two variables).*

Proof. 1) Let $\{X_t, t \in T\}$ be a centered, mean square continuous process. Since $\mathbb{E}X_t = 0$ for each $t \in T$, we have $R(s,t) = \mathbb{E}X_s\overline{X_t}, s, t \in T$. For any $s_0, t_0 \in T$ and sequences $s_n, t_n \in T, n \in \mathbb{N}$, such that $s_n \to s, t_n \to t, n \to \infty$, we have

$$|R(s_n, t_n) - R(s_0, t_0)| = |\mathbb{E}X_{s_n}\overline{X_{t_n}} - \mathbb{E}X_{s_0}\overline{X_{t_0}}| = |\langle X_{s_n}, X_{t_n}\rangle - \langle X_{s_0}, X_{t_0}\rangle| \to 0,\ n \to \infty,$$

since by assumption $X_{s_n} \to X_{s_0}, X_{t_n} \to X_{t_0}, n \to \infty$, in the mean square, and we use Theorem 3.2 (continuity of the inner product).

2) Let $R(s,t)$ be continuous at points $[s,t] \in T \times T$ such that $s = t$. Then for each $t_0 \in T$ and sequence $t_n \in T, n \in \mathbb{N}$, such that $t_n \to t_0, n \to \infty$, we have

$$\begin{aligned}\mathbb{E}|X_{t_n} - X_{t_0}|^2 &= \mathbb{E}(X_{t_n} - X_{t_0})\overline{(X_{t_n} - X_{t_0})} = \mathbb{E}X_{t_n}\overline{X_{t_n}} - \mathbb{E}X_{t_n}\overline{X_{t_0}} - \mathbb{E}X_{t_0}\overline{X_{t_n}} + \mathbb{E}X_{t_0}\overline{X_{t_0}} \\ &= R(t_n, t_n) - R(t_n, t_0) - R(t_0, t_n) + R(t_0, t_0) \to 0,\ n \to \infty.\end{aligned}$$

It follows that $\{X_t, t \in T\}$ is mean square continuous at point $t_0 \in T$ for any $t_0 \in T$ and hence $\{X_t, t \in T\}$ is mean square continuous. □

Theorem 4.2. *Let $\{X_t, t \in T\}$ be a stochastic process with finite second moments, $T \subset \mathbb{R}$ an open interval, with mean value $\{\mu_t, t \in T\}$ and autocovariance function $R(s,t), s, t \in T$. Then $\{X_t, t \in T\}$ is mean square continuous if and only if $\{\mu_t, t \in T\}$ is continuous on T and $R(s,t)$ is continuous at points $[s,t] \in T \times T$ such that $s = t$.*

Proof. 1) First, assume that $\{X_t, t \in T\}$ is mean square continuous. From Theorem 3.2 we have:

$$\mu_{t_n} = \mathbb{E}X_{t_n} \to \mathbb{E}X_{t_0} = \mu_{t_0}, n \to \infty,$$

for $t_n \to t_0, n \to \infty, t_0, t_n \in T, n \in \mathbb{N}$. This means that $\{\mu_t, t \in T\}$ is continuous. Similarly, using Theorem 3.2 we get for $t_n, s_n, t_0, s_0 \in T$ such that $t_n \to t_0, s_n \to s_0, n \to \infty$:

$$\mathbb{E}X_{s_n}\overline{X_{t_n}} \to \mathbb{E}X_{s_0}\overline{X_{t_0}},\ n \to \infty.$$

It follows that

$$R(s_n, t_n) = \mathbb{E}X_{s_n}\overline{X_{t_n}} - (\mathbb{E}X_{s_n})(\overline{\mathbb{E}X_{t_n}}) \to \mathbb{E}X_{s_0}\overline{X_{t_0}} - (\mathbb{E}X_{s_0})(\overline{\mathbb{E}X_{t_0}}) = R(s_0, t_0),\ n \to \infty.$$

Note that we have, in fact, shown that $R(s,t)$ is continuous at all points of $T \times T$.

2) Assume now continuity of $\{\mu_t, t \in T\}$ and continuity of $R(s,t)$ at points $[s,t] \in T \times T$ such that $s = t$. For $t_0, t_n \in T, n \in \mathbb{N}$, such that $t_n \to t_0, n \to \infty$, it holds that

$$\mathbb{E}|X_{t_n} - X_{t_0}|^2 = \mathbb{E}|(X_{t_n} - \mu_{t_n}) - (X_{t_0} - \mu_{t_0}) + (\mu_{t_n} - \mu_{t_0})|^2.$$

Denote $Y_t = X_t - \mu_t, t \in T$. The process $\{Y_t, t \in T\}$ is a centered stochastic process with the same autocovariance function R as $\{X_t, t \in T\}$. It follows by Theorem 4.1 that the process $\{Y_t, t \in T\}$ is mean square continuous. Finally,

$$\mathbb{E}|X_{t_n} - X_{t_0}|^2 = \mathbb{E}|(Y_{t_n} - Y_{t_0}) + (\mu_{t_n} - \mu_{t_0})|^2 \leq 2\,\mathbb{E}|Y_{t_n} - Y_{t_0}|^2 + 2\,|\mu_{t_n} - \mu_{t_0}|^2 \to 0,\ n \to \infty,$$

and hence the process $\{X_t, t \in T\}$ is mean square continuous. □

Theorem 4.3. *Let $\{X_t, t \in T\}$ be a weakly stationary stochastic process with the autocovariance function $R(t)$. Then $\{X_t, t \in T\}$ is mean square continuous if and only if $R(t)$ is continuous at $t = 0$.*

Proof. The claim follows from Theorem 4.2 and the fact that for weakly stationary processes the autocovariance function $R(s,t) = R(s-t)$ is continuous at points $[s,t] \in T \times T$ such that $s = t$ if and only if $R(t)$ is continuous at 0. □

Example: Let $\{X_t, t \in \mathbb{R}\}$ be a centered, weakly stationary stochastic process with the autocovariance function $R(t) = \cos t, t \in \mathbb{R}$. This process is mean square continuous.

Example: Let $\{X_t, t \in \mathbb{R}\}$ be a process of uncorrelated random variables with zero mean and finite positive variance $\operatorname{var} X_t = \sigma^2$, the same for all $t \in \mathbb{R}$. We can call this process *the continuous-time white noise*. This process is not mean square continuous.

Example: The Wiener process and the Poisson process are both mean square continuous on the interval $(0, \infty)$.

Remark: Mean square continuity is not the same as continuity of trajectories. For example, Wiener process has continuous trajectories almost surely, Poisson process has piece-wise constant trajectories with jumps, but both have the same form of the autocovariance function and both are mean square continuous.

Remark: Related concepts of mean square differentiability and mean square integrability of continuous-time stochastic processes can be also defined using the mean square limits.

5 Spectral decomposition of autocovariance functions

Remark: This chapter is not about stochastics at all – it studies properties of positive semidefinite functions (autocovariance functions) without considering any randomness. The motivation to study these topics lies e.g. in the field of signal processing, where spectral densities are one of the key tools. They can be used to find linear filters with the desired properties (for illustration of a simple low-pass filter, see Figure 16 in Section 7.6).

5.1 Auxiliary results

Lemma 5.1. *a) Let μ, ν be finite measures on the σ-algebra of Borel subsets of the interval $[-\pi, \pi]$. If for every $t \in \mathbb{Z}$ it holds that*

$$\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} \,\mathrm{d}\mu(\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} \,\mathrm{d}\nu(\lambda),$$

then $\mu(B) = \nu(B)$ for each Borel set $B \subset (-\pi, \pi)$ and $\mu(\{-\pi\} \cup \{\pi\}) = \nu(\{-\pi\} \cup \{\pi\})$.

b) Let μ, ν be finite measure on $(\mathbb{R}, \mathcal{B})$. If for every $t \in \mathbb{R}$ it holds that

$$\int_{-\infty}^{\infty} \mathrm{e}^{it\lambda} \,\mathrm{d}\mu(\lambda) = \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda} \,\mathrm{d}\nu(\lambda),$$

then $\mu(B) = \nu(B)$ for every $B \in \mathcal{B}$.

Proof. See Anděl (1976), Section III.1, Theorems 5 and 6. □

Remark: The b) part of the previous lemma is equivalent to saying that a characteristic function determines uniquely the distribution of a random variable. The a) part says a similar thing for random variables with bounded support.

Lemma 5.2 (Helly theorem)**.** *Let $\{F_n, n \in \mathbb{N}\}$ be a sequence of non-decreasing, uniformly bounded functions. Then there is a subsequence $\{F_{n_k}, k \in \mathbb{N}\}$ such that $F_{n_k} \to F, k \to \infty, n_k \to \infty$, weakly, i.e. in the points of continuity of F, where F is a non-decreasing right-continuous function.*

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

Lemma 5.3 (Helly-Bray theorem)**.** *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 $\lim_{n\to\infty} F_n(-\infty) = F(-\infty), \lim_{n\to\infty} F_n(\infty) = F(\infty)$. Let f be a continuous bounded function. Then*

$$\int_{-\infty}^{\infty} f(x) \,\mathrm{d}F_n(x) \to \int_{-\infty}^{\infty} f(x) \,\mathrm{d}F(x), \; n \to \infty.$$

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

Remark: The previous lemmas are traditionally stated for F_n, F being distribution functions. These results are generalized by the Portmonteau theorem.

Remark: For distribution functions, the assumptions of the Helly-Bray theorem on the limits are trivially fulfilled. However, they are needed in the general case. Consider e.g., for $x \in \mathbb{R}$, $F_n(x) =$

$I(x \geq n)$, $F(x) = 0$, and $f(x) = 1$. We have $F_n(x) \to F(x), n \to \infty$, for each $x \in \mathbb{R}$, but $F_n(\infty) = 1, \lim_{n\to\infty} F_n(\infty) = 1$ and $F(\infty) = 0$. On the other hand,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}F_n(x) = F_n(\infty) - F_n(-\infty) = 1 - 0 = 1 \nrightarrow \int_{-\infty}^{\infty} f(x)\,\mathrm{d}F(x) = 0.$$

Remark: The integral in 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 consider in the following that

$$\int_a^b f(x)\,\mathrm{d}F(x) = \int_{(a,b]} f(x)\,\mathrm{d}F(x) = \lim_{\delta\to 0+} \int_{[a+\delta,b]} f(x)\,\mathrm{d}F(x),$$

i.e. possible jump of the function F at point a does not influence the value of the integral.

5.2 Spectral decomposition of autocovariance functions

Theorem 5.4. *A complex-valued function $R(t), t \in \mathbb{Z}$, is the autocovariance function of a weakly stationary random sequence if and only if for any $t \in \mathbb{Z}$,*

$$R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda), \tag{5.1}$$

where F is a right-continuous, non-decreasing bounded function on $[-\pi, \pi]$ with $F(-\pi) = 0$. The function F is determined uniquely by the formula (5.1) *(in the class of functions with the required properties).*

Proof. 1. Assume (5.1) holds for a complex-valued function R defined on $\mathbb{Z}$. Choose $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{C}, t_1, \ldots, t_n \in \mathbb{Z}$, and compute:

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

because F is a non-decreasing function on $[-\pi, \pi]$. Hence R is a positive semidefinite and it is the autocovariance function of a random sequence, see Theorem 2.6.

2. Let R be the autocovariance function of a weakly stationary random sequence. Then $R(t)$ is positive semidefinite, i.e.

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c_j\bar{c}_k R(t_j - t_k) \geq 0 \quad \forall n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{C}, t_1, \ldots, t_n \in \mathbb{Z}.$$

Consider the case $t_j = j$ and $c_j = \mathrm{e}^{-ij\lambda}$ for a given $\lambda \in [-\pi, \pi]$. It follows that for every $n \in \mathbb{N}$ and $\lambda \in [-\pi, \pi]$,

$$\varphi_n(\lambda) = \frac{1}{2\pi n}\sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{e}^{-i(j-k)\lambda} R(j-k) \geq 0.$$

By substituting $\kappa = j - k$ and changing the order of summation, we get that

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

For $n \in \mathbb{N}$ we now define

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

Clearly, $F_n(-\pi) = 0$ and F_n is a non-decreasing function on $[-\pi, \pi]$ since $\varphi_n(\lambda) \ge 0$. We compute now the value $F_n(\pi)$:

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

because for $\kappa \in \mathbb{Z}$,

$$\int_{-\pi}^{\pi} \mathrm{e}^{-i\kappa\lambda}\,\mathrm{d}\lambda = \begin{cases} 2\pi, & \kappa = 0, \\ 0, & \kappa \ne 0. \end{cases} \tag{5.2}$$

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 $x \in \mathbb{R}, n \in \mathbb{N}$. Lemma 5.2 gives the existence of a subsequence $\{F_{n_k}\} \subset \{F_n\}$, $F_{n_k} \to \widetilde{F}$ weakly (in the points of continuity of $\widetilde{F}$) as $k \to \infty, n_k \to \infty$, where $\widetilde{F}$ is a non-decreasing, bounded, right-continuous function and $\widetilde{F}(x) = 0, x < -\pi$, and $\widetilde{F}(x) = R(0), x < \pi$. Note that also $\widetilde{F}(\pi) = R(0)$ since $\widetilde{F}$ is right-hand continuous.

The function $\widetilde{F}$ may have jumps both at points $-\pi$ and π but since the integrand in (5.1) is 2π-periodic, the jumps have the same effect. For uniqueness of the representation we now define

$$F(x) = \begin{cases} \widetilde{F}(x) - \widetilde{F}(-\pi), & x \in [-\pi, \pi), \\ \widetilde{F}(x), & \text{otherwise.} \end{cases}$$

From Lemma 5.3 we get, for $f(x) = \mathrm{e}^{itx}$ and $t \in \mathbb{Z}$,

$$\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_{n_k}(\lambda) \to \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda), \quad k \to \infty, n_k \to \infty.$$

On the other hand,

$$\begin{aligned} \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_{n_k}(\lambda) &= \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} \varphi_{n_k}(\lambda)\,\mathrm{d}\lambda = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} \frac{1}{2\pi n_k} \sum_{\kappa=-n_k+1}^{n_k-1} \mathrm{e}^{-i\kappa\lambda} R(\kappa)(n_k - |\kappa|)\,\mathrm{d}\lambda \\ &= \frac{1}{2\pi n_k} \sum_{\kappa=-n_k+1}^{n_k-1} R(\kappa)(n_k - |\kappa|) \int_{-\pi}^{\pi} \mathrm{e}^{i(t-\kappa)\lambda}\,\mathrm{d}\lambda, \end{aligned}$$

and hence we get using Equation (5.2) that

$$\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_{n_k}(\lambda) = \begin{cases} R(t)\left(1 - \frac{|t|}{n_k}\right), & |t| < n_k, \\ 0, & \text{otherwise.} \end{cases}$$

Finally, we get for $t \in \mathbb{Z}$:

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

We have shown that the integrals $\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_{n_k}$ converge both to $\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda)$ and to $R(t)$. We conclude that $R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda), t \in \mathbb{Z}$.

To prove uniqueness, suppose that $R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}G(\lambda), t \in \mathbb{Z}$, where G is a right-continuous, bounded, non-decreasing function on $[-\pi, \pi]$, and that $G(-\pi) = 0$. Then

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

where μ_F, μ_G are finite measures on Borel subsets of the interval $[-\pi, \pi]$ induced by functions F and G, respectively. Uniqueness follows from Lemma 5.1: $\mu_F(B) = \mu_G(B)$ for any Borel set $B \subset (-\pi, \pi)$ and $\mu_F(\{-\pi\} \cup \{\pi\}) = \mu_G(\{-\pi\} \cup \{\pi\})$. In our case we have $\mu_F(\{-\pi\}) = \mu_G(\{-\pi\}) = 0$ and hence $\mu_F = \mu_G$ and $F = G$. □

Remark: The formula (5.1) is called *the spectral decomposition* of an autocovariance function of a weakly stationary random sequence and F is called *the spectral distribution function.* If there is a function $f(\lambda) \geq 0, \lambda \in [-\pi, \pi]$, such that $F(\lambda) = \int_{-\pi}^{\lambda} f(x)\,\mathrm{d}x$ (i.e. if F is absolutely continuous), then f is called *the spectral density.* Clearly, $f = F'$. If the spectral density exists, we write

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

Remark: If F is piece-wise constant with jumps at points λ_j of size $a_j = F(\lambda_j) - F(\lambda_j-)$, we have $R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda) = \sum_j a_j \mathrm{e}^{it\lambda_j}$, $t \in \mathbb{Z}$.

Remark: If F has both an absolutely continuous part F_1 and a jump part F_2, $F = F_1 + F_2$, we have $R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_1(\lambda) + \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_2(\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} f_1(\lambda)\,\mathrm{d}\lambda + \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F_2(\lambda)$.

Remark: For weakly stationary processes with continuous time we can find a similar representation, but we need to restrict to autocovariance functions continuous at 0. For simplicity of formulation, this requirement is traditionally formulated as *weakly stationary, centered, mean square continuous process*, see below. However, the process being centered does not play any role here.

Theorem 5.5. *A complex-valued function $R(t), t \in \mathbb{R}$, is the autocovariance function of a weakly stationary, centered, mean square continuous random process if and only if for any $t \in \mathbb{R}$,*

$$R(t) = \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda), \tag{5.3}$$

where F is a non-decreasing, right-continuous function on $\mathbb{R}$ such that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = R(0) < \infty$. The function F is determined uniquely by the formula (5.3) *(in the class of functions with the required properties).*

Proof. (sketch of proof only)

1) Let R be a complex-valued function on $\mathbb{R}$ fulfilling Equation (5.3), where F is a non-decreasing, right-continuous function with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = R(0) < \infty$. Then R is positive semidefinite (similarly as in the proof of Theorem 5.4) and continuous (similarly as in the proof of the continuity of the characteristic function of a random variable).

According to Theorem 2.6, there is a centered, weakly stationary random process with the autocovariance function R. Since R is continuous, specifically continuous at point 0, this process is mean square continuous by Theorem 4.3.

2) Let R be the autocovariance function of a centered, mean square continuous process. It is positive semidefinite and continuous at point 0. The function F can be constructed analogously to the proof of Theorem 5.4. For the proof that R satisfies Equation (5.3), see e.g. Anděl (1976), Section IV.2, Theorem 2. □

Remark: Compare Theorems 5.4 and 5.5: for the representation with discrete time it was enough to consider F on a bounded interval, with continuous time it is necessary to consider F on $\mathbb{R}$.

Remark: The function F from Equation (5.3) is called *the spectral distribution function* of a weakly stationary, mean square continuous stochastic process. If F is absolutely continuous, its derivative f is called *the spectral density* and we can write $R(t) = \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda} f(\lambda)\,\mathrm{d}\lambda$, $t \in \mathbb{R}$.

Remark: Two different stochastic processes may have the same spectral distribution functions.

5.3 Existence and computation of spectral density

Remark: Below we establish a sufficient (but not necessary) condition for the existence of the spectral density and an algorithmic way of determining it. We start with an auxiliary lemma.

Lemma 5.6. *Let K be a complex-valued function on $\mathbb{Z}$ such that $\sum_{t=-\infty}^{\infty} |K(t)| < \infty$. Then*

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

where

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

Proof. Let K be such that $\sum_{t=-\infty}^{\infty} |K(t)| < \infty$ and define $f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \mathrm{e}^{-it\lambda} K(t), \lambda \in [-\pi, \pi]$. Since

$$\int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} \left| \frac{1}{2\pi} \mathrm{e}^{it\lambda} \mathrm{e}^{-ik\lambda} K(k) \right| \mathrm{d}\lambda < \infty$$

from the summability of K, Fubini theorem gives us that the following "double integral" exists and we can interchange the order of integration and summation: for a given $t \in \mathbb{Z}$,

$$\begin{aligned}
\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} f(\lambda)\,\mathrm{d}\lambda &= \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-ik\lambda} K(k)\,\mathrm{d}\lambda \\
&= \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} \frac{1}{2\pi} \mathrm{e}^{i(t-k)\lambda} K(k)\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} K(k) \int_{-\pi}^{\pi} \mathrm{e}^{i(t-k)\lambda}\,\mathrm{d}\lambda \\
&= K(t),
\end{aligned}$$

where we used (5.2) in the last equality. □

Theorem 5.7. *Let $\{X_t, t \in \mathbb{Z}\}$ be a weakly stationary sequence such that its autocovariance function R fulfills $\sum_{k=-\infty}^{\infty} |R(k)| < \infty$. Then the spectral density of $\{X_t, t \in \mathbb{Z}\}$ exists and we have*

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-ik\lambda} R(k),\ \lambda \in [-\pi, \pi]. \tag{5.4}$$

Proof. Since $\sum_{k=-\infty}^{\infty} |R(k)| < \infty$, it follows from Lemma 5.6 that

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

where

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

To prove that f is the spectral density, it is enough to show that $f(\lambda) \geq 0, \lambda \in [-\pi, \pi]$, and recall the uniqueness of the spectral decomposition.

We know from the proof of Theorem 5.4 that for every $n \in \mathbb{N}$ and $\lambda \in [-\pi, \pi]$,

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

We will now show that $f(\lambda) = \lim_{n\to\infty} \varphi_n(\lambda), \lambda \in [-\pi, \pi]$. For a given $n \in \mathbb{N}$ and $\lambda \in [-\pi, \pi]$, we have

$$f(\lambda) - \varphi_n(\lambda) = \frac{1}{2\pi} \sum_{|k|\geq n} \mathrm{e}^{-ik\lambda} R(k) + \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} \mathrm{e}^{-ik\lambda} R(k)|k|,$$

and

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

We see that $|f(\lambda) - \varphi_n(\lambda)| \to 0, n \to \infty$, as the first sum forms the remainder of a convergent series and the second sum converges to 0 from the Kronecker lemma. As a pointwise limit of non-negative functions, f is a non-negative function. □

Remark: The formula (5.4) is called *the inverse formula* for computing the spectral density of a weakly stationary random sequence. It gives the discrete-time Fourier transform of the autocovariance function R.

Remark: If $\sum_{k=-\infty}^{\infty} |R(k)| = \infty$, we do not know if the spectral density exists or not. See the exercise classes for an example of a random sequence with a spectral density but having $\sum_{k=-\infty}^{\infty} |R(k)| = \infty$.

Theorem 5.8. *Let $\{X_t, t \in \mathbb{R}\}$ be a centered, weakly stationary, mean square continuous random process such that its autocovariance function R fulfills $\int_{-\infty}^{\infty} |R(t)|\,\mathrm{d}t < \infty$. Then the spectral density of $\{X_t, t \in \mathbb{R}\}$ exists and we have*

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-it\lambda} R(t)\,\mathrm{d}t,\ \lambda \in \mathbb{R}. \tag{5.5}$$

Proof. The proof is analogous to the computation of a probability density function using the inverse Fourier transform of the characteristic function, see e.g. Štěpán (1987, Theorem IV.5.3). □

Remark: Equation (5.5) gives the Fourier transform of the autocovariance function R.

Remark: If $\{X_t, t \in \mathbb{R}\}$ is real-valued, we have $R(t) = R(-t), t \in \mathbb{R}$, and from Equation (5.5) we get $f(\lambda) = f(-\lambda), \lambda \in \mathbb{R}$. A similar property holds in the discrete time case.

Example: Let $\{X_t, t \in \mathbb{Z}\}$ be the white noise sequence $WN(0, \sigma^2)$. The sequence is weakly stationary, centered, with the autocovariance function $R(t) = \sigma^2 \cdot \delta(t), t \in \mathbb{Z}$, where $\delta(0) = 1$ and $\delta(t) = 0$ for $t \neq 0$. The summability condition $\sum_{t=-\infty}^{\infty} |R(t)| = \sigma^2 < \infty$ is fulfilled, and according to Theorem 5.7, the spectral density of $\{X_t, t \in \mathbb{Z}\}$ exists. Using the inverse formula (5.4) we get

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

This means that all frequencies contribute equally; hence the name *white noise*. By integration of $f(\lambda)$ we obtain

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

Example: In contrast with the previous example, consider the continuous-time white noise process $\{X_t, t \in \mathbb{R}\}$, i.e. the process of uncorrelated random variables with zero mean and the same (finite, positive) variance. It follows that its autocovariance function is $R(t) = \sigma^2 \cdot \delta(t), t \in \mathbb{R}$. The process $\{X_t, t \in \mathbb{R}\}$ is not mean square continuous and therefore no spectral representation of its autocovariance function exists.

Example: Consider a weakly stationary sequence with the autocovariance function $R(t) = a^{|t|}, t \in \mathbb{Z}$, for some $|a| < 1$. Later we will see that this is the autocovariance function of a causal autoregressive sequence of order 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 we can use the inverse formula (5.4) to get, for $\lambda \in [-\pi, \pi]$,

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

Example: Let $\{X_t, t \in \mathbb{R}\}$ be a centered, weakly stationary, mean square continuous process with the spectral distribution function

$$F(\lambda) = \begin{cases} 0, & \lambda < -1, \\ 1/2, & -1 \leq \lambda < 1, \\ 1, & \lambda \geq 1. \end{cases}$$

The spectral distribution function F is not absolutely continuous, therefore the spectral density does not exist. Using the formula (5.3), we get

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

The process is said to have a discrete spectrum with non-zero values at frequencies $\lambda_1 = -1, \lambda_2 = 1$.

6 Spectral representation of a stochastic process

The tools that we develop in this chapter will be useful in various proofs later in the course. However, the representation of a random sequence or random process in terms of stochastic integrals can be used e.g. for making predictions in the spectral domain or filtration of signal and noise.

In general, stochastic integration is a key tool for working with stochastic differential equations which describe systems influenced by randomness. Examples of such systems include financial markets (modeling of asset prices or interest rates; see e.g. the Black-Scholes model for option pricing), signals with noise, biological populations, and more.

6.1 Orthogonal increment processes

Definition 6.1. *Let $T \subset \mathbb{R}$ be an interval and $\{X_t, t \in T\}$ be a stochastic process with finite second moments. The process $\{X_t, t \in T\}$ is* an orthogonal increment process*, if for any $t_1, \ldots, t_4 \in T$ such that $(t_1, t_2] \cap (t_3, t_4] = \emptyset$ we have*

$$\mathbb{E}(X_{t_2} - X_{t_1})(\overline{X_{t_4}} - \overline{X_{t_3}}) = 0. \tag{6.1}$$

Remark: The formula (6.1) means that the random variables $X_{t_2} - X_{t_1}$ and $X_{t_4} - X_{t_3}$ are orthogonal in the space $L_2(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $\langle X_{t_2} - X_{t_1}, X_{t_4} - X_{t_3} \rangle = 0$.

Remark: In the following we will consider only centered, right-mean square continuous processes on $T = [a, b]$, i.e. processes such that $\mathbb{E}|X_t - X_{t_0}|^2 \to 0, t \to t_0+$, for any $t_0 \in [a, b)$.

Theorem 6.1. *Let $\{Z_\lambda, \lambda \in [a, b]\}$ be a centered, orthogonal increment, right-mean square continuous process, $[a, b]$ a bounded interval. Then there is a unique non-decreasing, right-continuous function F such that $F(\lambda) = 0$ for $\lambda \le a$, $F(\lambda) = F(b)$ for $\lambda \ge b$ and*

$$F(\lambda_2) - F(\lambda_1) = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2, \ a \le \lambda_1 < \lambda_2 \le b. \tag{6.2}$$

Proof. We start by defining the function $F(\lambda) = \mathbb{E}|Z_\lambda - Z_a|^2, \lambda \in [a, b]$, $F(\lambda) = 0, \lambda \le a$, and $F(\lambda) = F(b), \lambda \ge b$. We will show that this function is non-decreasing, right-continuous, and satisfies the conditions of the theorem. Clearly, it is enough to consider $\lambda \in [a, b]$ only.

Let $a < \lambda_1 < \lambda_2 \le b$. We have

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

since the increments $Z_{\lambda_2} - Z_{\lambda_1}$ and $Z_{\lambda_1} - Z_a$ are orthogonal. We have shown that (6.2) holds. Furthermore, we have $F(\lambda_2) - F(\lambda_1) = \mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 \ge 0$, meaning that F is non-decreasing.

The function F is also right-continuous, since $\mathbb{E}|Z_{\lambda_2} - Z_{\lambda_1}|^2 \to 0, \lambda_2 \to \lambda_1+$, due to the right-mean square continuity of the process $\{Z_\lambda, \lambda \in [a, b]\}$.

Concerning uniqueness, let G be a non-decreasing, right-continuous function that satisfies the assumptions of the theorem. Then $G(a) = 0 = F(a)$, and for $\lambda \in (a, b]$ it holds that $G(\lambda) = G(\lambda) - G(a) = \mathbb{E}|Z_{\lambda_2} - Z_a|^2 = F(\lambda) - F(a) = F(\lambda)$, meaning that $G = F$. □

Definition 6.2. *The bounded, non-decreasing, right-continuous function F from the previous theorem is* the distribution function associated with the orthogonal increment process *(or* associated distribution function *or* orthogonal distribution function*).*

Example: The Wiener process $\{W_t, t \in [0,T]\}$ on a bounded interval $[0,T]$ has independent, stationary increments. It follows that the process has orthogonal increments. The orthogonal distribution function is

$$F(\lambda) = \begin{cases} 0, & \lambda \leq 0, \\ \sigma^2\lambda, & 0 \leq \lambda \leq T, \\ \sigma^2 T, & \lambda \geq T. \end{cases}$$

Example: Let $\{\widetilde{W}_\lambda, \lambda \in [-\pi,\pi]\}$ be the transformation of the Wiener process given by $\widetilde{W}_\lambda = W_{(\lambda+\pi)/2\pi}, \lambda \in [-\pi,\pi]$. The process $\{\widetilde{W}_\lambda, \lambda \in [-\pi,\pi]\}$ is a Gaussian process with orthogonal increments and the orthogonal distribution function

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

Note that we already know this function is the spectral distribution function of a white noise sequence. Later in this chapter, we will establish a rigorous connection between these two objects.

6.2 Integral with respect to an orthogonal increment process

Background

Let $[a,b]$ be a bounded interval and $\mathcal{B}([a,b])$ the Borel sets on $[a,b]$. Furthermore, let $\{Z_\lambda, \lambda \in [a,b]\}$ be a centered, right-mean square continuous process with orthogonal increments, defined on $(\Omega, \mathcal{A}, \mathbb{P})$, F its orthogonal distribution function and μ_F the finite measure induced by F. This means that for a set $B \in \mathcal{B}([a,b])$ we have $\mu_F(B) = \int_B 1 \,\mathrm{d}F(u)$, and as a special case we get $\mu_F((c,d]) = F(d) - F(c)$ for $(c,d] \subset [a,b]$.

Let $L_2(F) = L_2([a,b], \mathcal{B}([a,b]), \mu_F)$ be the space of measurable, complex-valued functions f on $[a,b]$ such that

$$\int_a^b |f(\lambda)|^2 \,\mathrm{d}\mu_F(\lambda) = \int_a^b |f(\lambda)|^2 \,\mathrm{d}F(\lambda) < \infty.$$

We define classes of equivalence on $L_2(F)$ by the relation $f \sim g \iff f = g$ μ_F-almost everywhere, i.e. $f \sim g \iff \int_a^b |f(\lambda) - g(\lambda)|^2 \,\mathrm{d}\mu_F(\lambda) = 0$. As before, working with the classes of equivalence allows us to define a norm and have unique limits of Cauchy sequences.

The inner product on the classes of equivalence on $L_2(F)$ is defined as

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

The norm in $L_2(F)$ is then

$$\|f\| = \left[\int_a^b |f(\lambda)|^2 \,\mathrm{d}F(\lambda)\right]^{1/2}, \ f \in L_2(F).$$

Convergence in $L_2(F)$ is the convergence in the norm: $f_n \to f$ in $L_2(F)$ as $n \to \infty$, if $\|f_n - f\|^2 \to 0$, $n \to \infty$, i.e. $\int_a^b |f_n(\lambda) - f(\lambda)|^2 \,\mathrm{d}F(\lambda) \to 0, n \to \infty$.

The space $L_2(F)$ is complete with respect to the convergence defined above.

Construction of the integral – simple functions

Let $f \in L_2(F)$ be a simple function, i.e. we assume there is $n \in \mathbb{N}$, $a = \lambda_0 < \lambda_1 < \ldots < \lambda_n = b$ and $c_1, \ldots, c_n \in \mathbb{C}$ such that

$$f(\lambda) = \sum_{k=1}^{n} c_k J_{(\lambda_{k-1}, \lambda_k]}(\lambda), \ \lambda \in [a, b],$$

where $J_A(\cdot)$ is the indicator function. For uniqueness of representation, we assume that $c_k \neq c_{k+1}$ for $k = 1, 2, \ldots, n-1$.

We define

$$\int_{(a,b]} f(\lambda) \, \mathrm{d}Z(\lambda) = \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}),$$

which is a random variable in $L_2(\Omega, \mathcal{A}, \mathbb{P})$. Instead of $\int_{(a,b]} f(\lambda) \, \mathrm{d}Z(\lambda)$ we will write $\int_a^b f(\lambda) \, \mathrm{d}Z(\lambda)$. In the following we will denote $I(f) = \int_a^b f(\lambda) \, \mathrm{d}Z(\lambda)$.

Theorem 6.2. *Let $\{Z_\lambda, \lambda \in [a, b]\}$ be a centered, right-mean square continuous process with orthogonal increments and the orthogonal distribution function F. Let f, g be simple functions in $L_2(F)$ and $\alpha, \beta \in \mathbb{C}$ constants. Then*

1. $\mathbb{E} \int_a^b f(\lambda) \, \mathrm{d}Z(\lambda) = 0$,
2. $\int_a^b [\alpha f(\lambda) + \beta g(\lambda)] \, \mathrm{d}Z(\lambda) = \alpha \int_a^b f(\lambda) \, \mathrm{d}Z(\lambda) + \beta \int_a^b g(\lambda) \, \mathrm{d}Z(\lambda)$,
3. $\mathbb{E} \left(\int_a^b f(\lambda) \, \mathrm{d}Z(\lambda) \right) \overline{\left(\int_a^b g(\lambda) \, \mathrm{d}Z(\lambda) \right)} = \int_a^b f(\lambda) \overline{g(\lambda)} \, \mathrm{d}F(\lambda)$.

Proof. 1. To set the notation, let $f(\lambda) = \sum_{k=1}^{n} c_k J_{(\lambda_{k-1}, \lambda_k]}(\lambda), \lambda \in [a, b]$. Then

$$\mathbb{E} \int_a^b f(\lambda) \, \mathrm{d}Z(\lambda) = \mathbb{E} \sum_{k=1}^{n} c_k (Z_{\lambda_k} - Z_{\lambda_{k-1}}) = \sum_{k=1}^{n} c_k \mathbb{E}(Z_{\lambda_k} - Z_{\lambda_{k-1}}) = 0,$$

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

2. Without loss of generality, we assume that f, g use the same division $a = \lambda_0 < \lambda_1 < \ldots < \lambda_n = b$, otherwise we use a common refinement of the two divisions. We use the following notation:

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

Then,

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

3. Let f, g be as above. Using orthogonality of the increments of $\{Z_\lambda, \lambda \in [a,b]\}$ we get

$$\begin{aligned}\mathbb{E}\left(\int_a^b f(\lambda)\,\mathrm{d}Z(\lambda)\right)\overline{\left(\int_a^b g(\lambda)\,\mathrm{d}Z(\lambda)\right)} &= \mathbb{E}\left(\sum_{k=1}^n c_k(Z_{\lambda_k}-Z_{\lambda_{k-1}})\right)\overline{\left(\sum_{j=1}^n d_j(Z_{\lambda_j}-Z_{\lambda_{j-1}})\right)}\\ &= \sum_{k=1}^n c_k\overline{d_k}\,\mathbb{E}|Z_{\lambda_k}-Z_{\lambda_{k-1}}|^2\\ &= \sum_{k=1}^n c_k\overline{d_k}(F(\lambda_k)-F(\lambda_{k-1}))\\ &= \int_a^b f(\lambda)\overline{g(\lambda)}\,\mathrm{d}F(\lambda).\end{aligned}$$

□

Remark: Property 3. in the theorem above is useful for computing the covariance of $I(f)$ and $I(g)$. Also, note that

$$\mathbb{E}I(f)\overline{I(g)} = \langle I(f), I(g)\rangle_{L_2(\Omega,\mathcal{A},\mathbb{P})} = \langle f, g\rangle_{L_2(F)}.$$

This means that the mapping $I : L_2(F) \to L_2(\Omega,\mathcal{A},\mathbb{P})$ is an isometry (see also the analogy of property 3. for measurable functions in the theorem below).

Construction of the integral – measurable functions

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 $L_2(F)$. Hence, there is a sequence of simple functions $f_n \in L_2(F)$ such that $f_n \to f$ in $L_2(F), n \to \infty$.

The integral $I(f_n) \in L_2(\Omega,\mathcal{A},\mathbb{P})$ has been defined above. The sequence $\{I(f_n), n \in \mathbb{N}\}$ is Cauchy in $L_2(\Omega,\mathcal{A},\mathbb{P})$ since

$$\begin{aligned}\mathbb{E}|I(f_n)-I(f_m)|^2 &= \mathbb{E}(I(f_n)-I(f_m))\overline{(I(f_n)-I(f_m))}\\ &= \mathbb{E}I(f_n-f_m)\overline{I(f_n-f_m)}\\ &= \int_a^b (f_n(\lambda)-f_m(\lambda))\overline{(f_n(\lambda)-f_m(\lambda)}\,\mathrm{d}F(\lambda)\\ &= \int_a^b |f_n(\lambda)-f_m(\lambda)|^2\,\mathrm{d}F(\lambda) \to 0,\ m,n\to\infty.\end{aligned}$$

This holds because $f_n \to f, n \to \infty$, in $L_2(F)$ and hence the sequence $\{f_n, n \in \mathbb{N}\}$ is Cauchy in $L_2(F)$.

The space $L_2(\Omega,\mathcal{A},\mathbb{P})$ is complete. Therefore, there is the mean square limit $I(f) = \text{l.i.m.}_{n\to\infty} I(f_n) = \int_a^b f(\lambda)\,\mathrm{d}Z(\lambda)$. The limit is called *the integral of f with respect to the orthogonal increment process* $\{Z_\lambda, \lambda \in [a,b]\}$, or shortly *the stochastic integral.*

Note that $I(f)$ does not depend on the choice of the approximating sequence $\{f_n, n \in \mathbb{N}\}$. To see that, consider a function $f \in L_2(F)$ and let $\{f_n, n \in \mathbb{N}\}$, $\{g_n, n \in \mathbb{N}\}$ be sequences of simple functions such that $f_n \to f, g_n \to f, n \to \infty$, in $L_2(F)$. Then $I(f_n) \to I, I(g_n) \to J$ in the mean square, as $n \to \infty$.

We define the sequence $\{h_n, n \in \mathbb{N}\} = \{f_1, g_1, f_2, g_2, f_3, \ldots\}$. This is a sequence of simple functions and $h_n \to f, n \to \infty$, in $L_2(F)$. It follows that $\{I(h_n), n \in \mathbb{N}\}$ is Cauchy and there is a limiting random variable K such that $I(h_n) \to K, n \to \infty$, in the mean square.

The mean square limit is uniquely determined (recall that we consider the classes of equivalence, not the random variables themselves). Also, each selected subsequence of a convergent sequence has the same limit. In our case, $I(f_n) \to K, I(g_n) \to K, n \to \infty$, in the means square, and hence $I = J = K$.

Theorem 6.3. *Let $\{Z_\lambda, \lambda \in [a,b]\}$ be a centered, right-mean square continuous process with orthogonal increments and the orthogonal distribution function F. Let $f, g \in L_2(F)$ and $\alpha, \beta \in \mathbb{C}$ constants. Then*

1. $\mathbb{E}I(f) = \mathbb{E}\int_a^b f(\lambda)\,\mathrm{d}Z(\lambda) = 0$,
2. $I(\alpha f + \beta g) = \int_a^b [\alpha f(\lambda) + \beta g(\lambda)]\,\mathrm{d}Z(\lambda) = \alpha \int_a^b f(\lambda)\,\mathrm{d}Z(\lambda) + \beta \int_a^b g(\lambda)\,\mathrm{d}Z(\lambda) = \alpha I(f) + \beta I(g)$,
3. $\mathbb{E}I(f)\overline{I(g)} = \mathbb{E}\left(\int_a^b f(\lambda)\,\mathrm{d}Z(\lambda)\right)\overline{\left(\int_a^b g(\lambda)\,\mathrm{d}Z(\lambda)\right)} = \int_a^b f(\lambda)\overline{g(\lambda)}\,\mathrm{d}F(\lambda)$.

Furthermore, let $\{f_n, n \in \mathbb{N}\} \subset L_2(F), f \in L_2(F)$. Then

4. $f_n \to f$ *in* $L_2(F), n \to \infty \iff I(f_n) \to I(f)$ *in* $L_2(\Omega, \mathcal{A}, \mathbb{P}), n \to \infty$.

Proof. 1. Let $f \in L_2(F)$ and $\{f_n, n \in \mathbb{N}\}$ be a sequence of simple functions in $L_2(F)$ such that $f_n \to f$ in $L_2(F)$, $n \to \infty$. Then $I(f) = \text{l.i.m.}_{n\to\infty} I(f_n)$, and $\mathbb{E}I(f) = \lim_{n\to\infty} \mathbb{E}I(f_n) = 0$.

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

Consider the simple functions $h_n = \alpha f_n + \beta g_n, n \in \mathbb{N}$. We have $h_n \to h = \alpha f + \beta g$ in $L_2(F), n \to \infty$, since

$$\int_a^b |\alpha f_n(\lambda) + \beta g_n(\lambda) - (\alpha f(\lambda) + \beta g(\lambda))|^2\,\mathrm{d}F(\lambda)$$
$$\le 2|\alpha|^2 \int_a^b |f_n(\lambda) - f(\lambda)|^2\,\mathrm{d}F(\lambda) + 2|\beta|^2 \int_a^b |g_n(\lambda) - g(\lambda)|^2\,\mathrm{d}F(\lambda) \to 0, \quad n \to \infty.$$

We have

- $h_n = \alpha f_n + \beta g_n$ is a simple function,
- $I(h_n) = I(\alpha f_n + \beta g_n) = \alpha I(f_n) + \beta I(g_n)$ by Theorem 6.2,
- $h_n \to h$ in $L_2(F) \Rightarrow I(h_n) \to I(h)$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ (by construction of the stochastic integral),
- $h = \alpha f + \beta g$ is a measurable function,
- $I(h_n) \to \alpha I(f) + \beta I(g)$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ since

$$\begin{aligned}
&\mathbb{E}|\alpha I(f_n) + \beta I(g_n) - (\alpha I(f) + \beta I(g))|^2 \\
&= \mathbb{E}|\alpha(I(f_n) - I(f)) + \beta(I(g_n) - I(g))|^2 \\
&\le 2|\alpha|^2\,\mathbb{E}|I(f_n) - I(f)|^2 + 2|\beta|^2\,\mathbb{E}|I(g_n) - I(g)|^2 \to 0, \quad n \to \infty.
\end{aligned}$$

It follows that $I(h_n) \to I(h), n \to \infty$, but at the same time $I(h_n) \to \alpha I(f) + \beta I(g), n \to \infty$. The uniqueness of the limit gives $I(h) = \alpha I(f) + \beta I(g)$.

3. Let $f, g \in L_2(F)$ and $\{f_n, n \in \mathbb{N}\}, \{g_n, n \in \mathbb{N}\}$ be sequences of simple functions in $L_2(F)$ such that $f_n \to f, g_n \to g$ in $L_2(F)$, $n \to \infty$. Again, it follows that $I(f_n) \to I(f), I(g_n) \to I(g)$ in $L_2(\Omega, \mathcal{A}, \mathbb{P})$.

Continuity of the inner product in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ gives

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

On the other hand, Theorem 6.2 and continuity of the inner product in $L_2(F)$ give

$$\mathbb{E}I(f_n)\overline{I(g_n)} = \int_a^b f_n(\lambda)\overline{g_n(\lambda)}\,\mathrm{d}F(\lambda) = \langle f_n, g_n\rangle \to \langle f, g\rangle = \int_a^b f(\lambda)\overline{g(\lambda)}\,\mathrm{d}F(\lambda), \quad n \to \infty.$$

Using the uniqueness of the limit, we see that $\mathbb{E}I(f)\overline{I(g)} = \int_a^b f(\lambda)\overline{g(\lambda)}\,\mathrm{d}F(\lambda)$.

4. Let $f \in L_2(F)$ and $\{f_n, n \in \mathbb{N}\}$ be a sequence of simple functions in $L_2(F)$ such that $f_n \to f$ in $L_2(F)$, $n \to \infty$. From points 2 and 3 above, we have

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

which proves the claim. □

Remark: The construction of the stochastic integral can be extended to unbounded intervals. Let $\{Z_\lambda, \lambda \in \mathbb{R}\}$ be a centered, right-mean square continuous process with orthogonal increments. The function F defined by

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

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 a measurable function f such that $\int_{-\infty}^{\infty} |f(\lambda)|^2\,\mathrm{d}\mu_F(\lambda) = \int_{-\infty}^{\infty} |f(\lambda)|^2\,\mathrm{d}F(\lambda) < \infty$, we can define $\int_{-\infty}^{\infty} f(\lambda)\,\mathrm{d}Z(\lambda) = \text{l.i.m.} \int_a^b f(\lambda)\,\mathrm{d}Z(\lambda)$ for $a \to -\infty, b \to \infty$.

6.3 Spectral decomposition of a stochastic process

Under certain assumptions, we can represent a random sequence or a stochastic process using stochastic integrals with respect to an orthogonal increment process.

Theorem 6.4. *Let $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ be a centered, right-mean square continuous process with orthogonal increments and the orthogonal distribution function F. Let $X_t, t \in \mathbb{Z}$, be random variables defined as*

$$X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda).$$

Then $\{X_t, t \in \mathbb{Z}\}$ is a centered, weakly stationary sequence with the spectral distribution function F.

Proof. The orthogonal distribution function F of $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ is bounded, non-decreasing, right-continuous, with $F(\lambda) = 0$ for $\lambda \le -\pi$ and $F(\lambda) = F(\pi)$ for $\lambda \ge \pi$. For $t \in \mathbb{Z}$ we define a function

$$e_t(\lambda) = \mathrm{e}^{it\lambda}, \quad \lambda \in [-\pi, \pi].$$

Note that this is simply a matter of notation to stress that we treat it as a function of λ. Then $e_t \in L_2(F)$ since

$$\int_{-\pi}^{\pi} |e_t(\lambda)|^2\,\mathrm{d}F(\lambda) = \int_{-\pi}^{\pi} 1\,\mathrm{d}F(\lambda) = F(\pi) - F(-\pi) < \infty.$$

It follows that $X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)$ is a well-defined random variable for each $t \in \mathbb{Z}$.

From Theorem 6.3 we have

$$\mathbb{E}X_t = \mathbb{E}\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda) = 0 \quad \forall t \in \mathbb{Z}$$

(the sequence is centered) and

$$\begin{aligned}\mathbb{E}|X_t|^2 &= \mathbb{E}\left|\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)\right|^2 = \mathbb{E}\left(\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)\right)\overline{\left(\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)\right)} \\ &= \int_{-\pi}^{\pi}\left|\mathrm{e}^{it\lambda}\right|^2\,\mathrm{d}F(\lambda) = \int_{-\pi}^{\pi} 1\,\mathrm{d}F(\lambda) = F(\pi) - F(-\pi) < \infty\end{aligned}$$

(constant finite variance). Also, the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$ is

$$\begin{aligned}R(t+h,t) = \mathrm{cov}(X_{t+h}, X_t) &= \mathbb{E}\left(\int_{-\pi}^{\pi} \mathrm{e}^{i(t+h)\lambda}\,\mathrm{d}Z(\lambda)\right)\overline{\left(\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)\right)} \\ &= \int_{-\pi}^{\pi} \mathrm{e}^{ih\lambda}\,\mathrm{d}F(\lambda) = R(h), \quad h \in \mathbb{Z}\end{aligned}$$

(translation invariance of the autocovariance function plus the defining property of the spectral distribution function).

We conclude that $\{X_t, t \in \mathbb{Z}\}$ is centered and weakly stationary, and F has the same properties as a spectral distribution function, see Theorem 5.4. The uniqueness of the spectral decomposition of the autocovariance function implies that F is the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$. □

Remark: The important part of the previous theorem is that the orthogonal distribution function of $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ and the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$ are the same.

Example: Let $\{\widetilde{W}_\lambda, \lambda \in [-\pi, \pi]\}$ be the transformation of the Wiener process given by $\widetilde{W}_\lambda = W_{(\lambda+\pi)/2\pi}, \lambda \in [-\pi, \pi]$. Then, the random variables

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

are centered, uncorrelated, with the same variance σ^2, i.e. they form a white-noise sequence. It follows from the fact that the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$ is the same as the orthogonal distribution function of $\{\widetilde{W}_\lambda, \lambda \in [-\pi, \pi]\}$, which is the same as the spectral distribution function of a white noise sequence, see the previous examples.

Theorem 6.5. *Let $\{X_t, t \in \mathbb{Z}\}$ be a centered, weakly stationary sequence with the spectral distribution function F. Then there is a centered, orthogonal increment process $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ such that*

$$X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda),\ t \in \mathbb{Z}, \tag{6.3}$$

and

$$\mathbb{E}|Z(\lambda) - Z(-\pi)|^2 = F(\lambda),\ -\pi \le \lambda \le \pi.$$

Proof. See Brockwell and Davis (2006, Theorem 4.8.2). □

Remark: The formula (6.3) is called *the spectral decomposition of a weakly stationary random sequence*. Note that once again, the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$ is the same as the orthogonal distribution function of $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$.

Remark: Theorem 6.5 states that all random variables X_t from a centered, weakly stationary random sequence $\{X_t, t \in \mathbb{Z}\}$ can be approximated, in the mean square limit, by the sums $\sum_{j=1}^{n} \mathrm{e}^{it\lambda_j} Y_j$ of uncorrelated random variables Y_j, with $\operatorname{var} Y_j = F(\lambda_j) - F(\lambda_{j-1})$, where F is the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$. To see this, recall the construction of the stochastic integral as a mean square limit and set $Y_j = Z_{\lambda_j} - Z_{\lambda_{j-1}}$. Then we get $\operatorname{var} Y_j = \mathbb{E}|Z_{\lambda_j} - Z_{\lambda_{j-1}}|^2 = F(\lambda_j) - F(\lambda_{j-1})$, and from the orthogonal increment property of $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ we get that the random variables Y_j are uncorrelated. Note that the same random variables Y_j are used for each $t \in \mathbb{Z}$.

Theorem 6.6. *Let $\{X_t, t \in \mathbb{R}\}$ be a centered, weakly stationary, mean square continuous stochastic process with the spectral distribution function F. Then there is a centered, orthogonal increment process $\{Z_\lambda, \lambda \in \mathbb{R}\}$ such that*

$$X_t = \int_{-\infty}^{\infty} \mathrm{e}^{it\lambda} \, \mathrm{d}Z(\lambda), \; t \in \mathbb{R}, \tag{6.4}$$

and the orthogonal distribution function of $\{Z_\lambda, \lambda \in \mathbb{R}\}$ is F.

Proof. See Priestley (1981, Chapter 4.11). □

Remark: The formula (6.4) is called *the spectral decomposition of a weakly stationary, mean square continuous stochastic process.*

7 Linear models

7.1 White noise

Definition 7.1. *Let $\{Y_t, t \in \mathbb{Z}\}$ be a sequence of uncorrelated, centered random variables with finite positive variance σ^2. We call this sequence* the white noise *and denote this model by WN(0,σ^2).*

Remark: The white noise sequence $\{Y_t, t \in \mathbb{Z}\}$ is weakly stationary, $R(t) = \sigma^2\delta(t)$, $t \in \mathbb{Z}$, i.e. $R(0) = \sigma^2$ and $R(t) = 0$ for $t \neq 0$. The spectral density exists and is constant: $f(\lambda) = \frac{\sigma^2}{2\pi}$, $\lambda \in [-\pi, \pi]$, and the spectral distribution function is $F(\lambda) = \frac{\sigma^2}{2\pi}(\lambda + \pi)$, $\lambda \in [-\pi, \pi]$. The spectral decomposition of the sequence is given by $Y_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda)$, where $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ is an orthogonal increment process. Its orthogonal distribution function is F, i.e. it coincides with the spectral distribution function of the sequence $\{Y_t, t \in \mathbb{Z}\}$. Sample realizations of the white noise sequence are given in Figure 5, and the autocovariance function and the spectral density are shown in Figure 6.

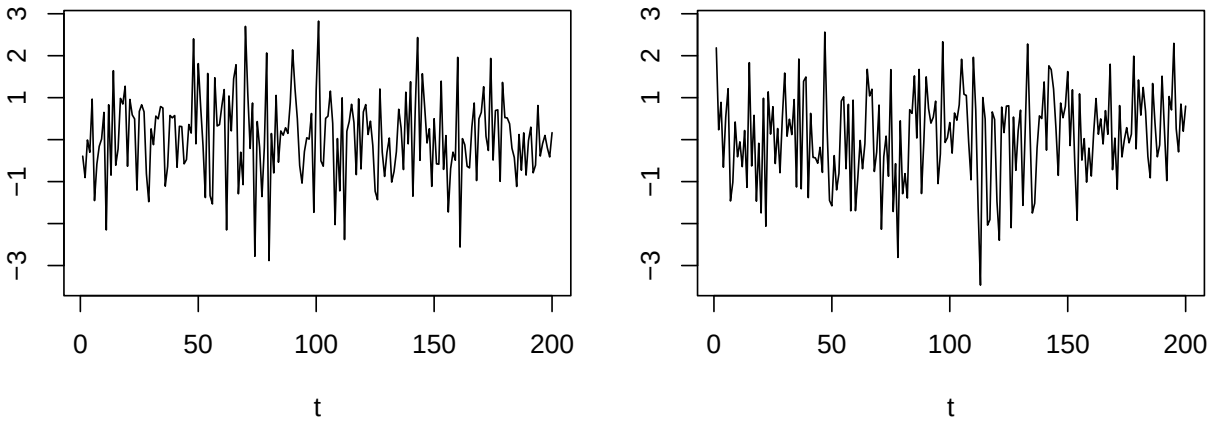

Figure 5: Sample realizations of the white noise sequence with $\sigma^2 = 1$. Note that these are discrete-time sequences, and the lines joining the corresponding points in the plots are used only for clarity.

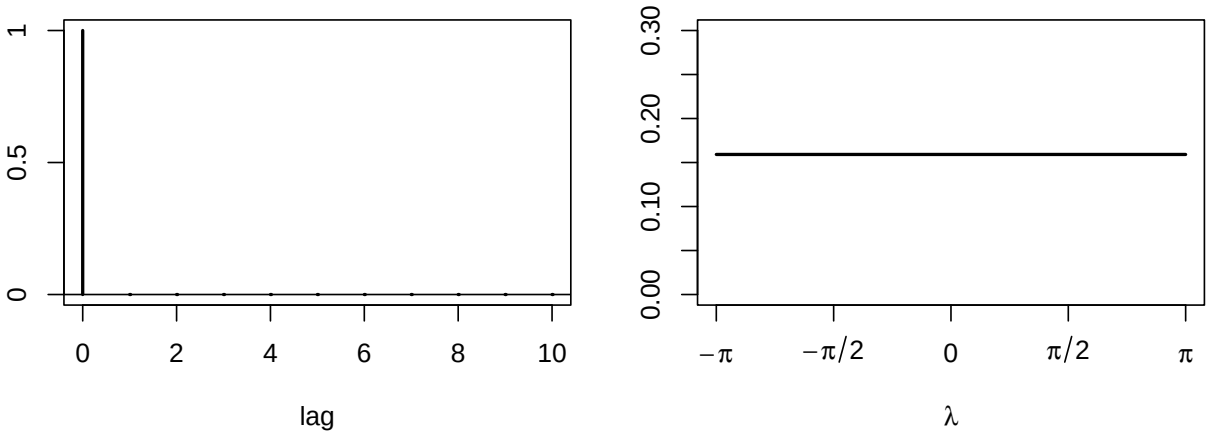

Figure 6: Autocovariance function of the white noise sequence with $\sigma^2 = 1$ (left) and the corresponding spectral density (right).

7.2 Moving average sequences

Definition 7.2. *Let* $\{Y_t, t \in \mathbb{Z}\}$ *be a white noise sequence WN(0,*σ^2*) and* $b_0, \ldots, b_n \in \mathbb{C}$*, where* $b_0 \neq 0, b_n \neq 0$*. The sequence* $\{X_t, t \in \mathbb{Z}\}$ *given by*

$$X_t = b_0 Y_t + b_1 Y_{t-1} + \ldots + b_n Y_{t-n},\ t \in \mathbb{Z}, \tag{7.1}$$

is called a moving average sequence of order n*. We write simply MA(n).*

Remark: The name *moving average sequence* is related to the special case with $b_0 = \ldots = b_n = \frac{1}{n+1}$ which gives the usual arithmetic mean.

Remark: Sample realizations of simple MA(1) models are given in Figure 7. The corresponding autocovariance functions and spectral densities are shown in Figures 8 and 9, respectively.

Theorem 7.1. *Let* $\{X_t, t \in \mathbb{Z}\}$ *be a moving average sequence of order* n *defined by* (7.1)*. The sequence is centered, weakly stationary, with the autocovariance function*

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

The spectral density of $\{X_t, t \in \mathbb{Z}\}$ *exists and is given by*

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

Proof. 1. $\mathbb{E}X_t = 0, t \in \mathbb{Z}$, from the linearity of expectation.

2. We compute the autocovariance function in the time domain: for $t, s \in \mathbb{Z}, t \geq 0$, we have

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

Note that $\mathbb{E}Y_{s+t-j}\overline{Y_{s-k}} = \sigma^2$ if $s + t - j = s - k$, i.e. if $j = t - k$, and $\mathbb{E}Y_{s+t-j}\overline{Y_{s-k}} = 0$ otherwise. It follows that

$$\mathrm{cov}(X_{s+t}, X_s) = \begin{cases} \sigma^2 \sum_{k=0}^{n-t} b_{t+k}\overline{b_k}, & 0 \leq t \leq n, \\ 0, & t > n. \end{cases}$$

For $t < 0$, we proceed analogously. Since $\mathrm{cov}(X_{s+t}, X_s)$ depends on t only, weak stationarity follows.

3. Now we compute the autocovariance function in the spectral domain, but first we consider the spectral decomposition of the white noise sequence $\{Y_t, t \in \mathbb{Z}\}$:

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

where $\{Z_Y(\lambda), \lambda \in [-\pi, \pi]\}$ is a centered, orthogonal increment process with the orthogonal distribution function F_Y, which is also the spectral distribution function of $\{Y_t, t \in \mathbb{Z}\}$. We compute for $t \in \mathbb{Z}$:

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

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

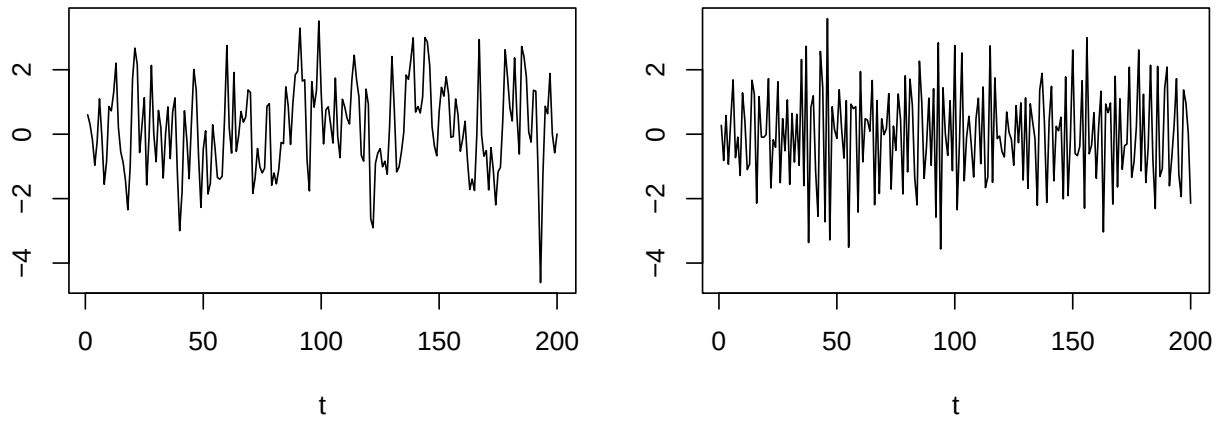

Figure 7: Sample realizations of the MA(1) sequence with $b_0 = 1, b_1 = 1$ (left) and $b_0 = 1, b_1 = -1$ (right). In both cases $\sigma^2 = 1$. Note that these are discrete-time sequences, and the lines joining the corresponding points in the plots are used only for clarity.

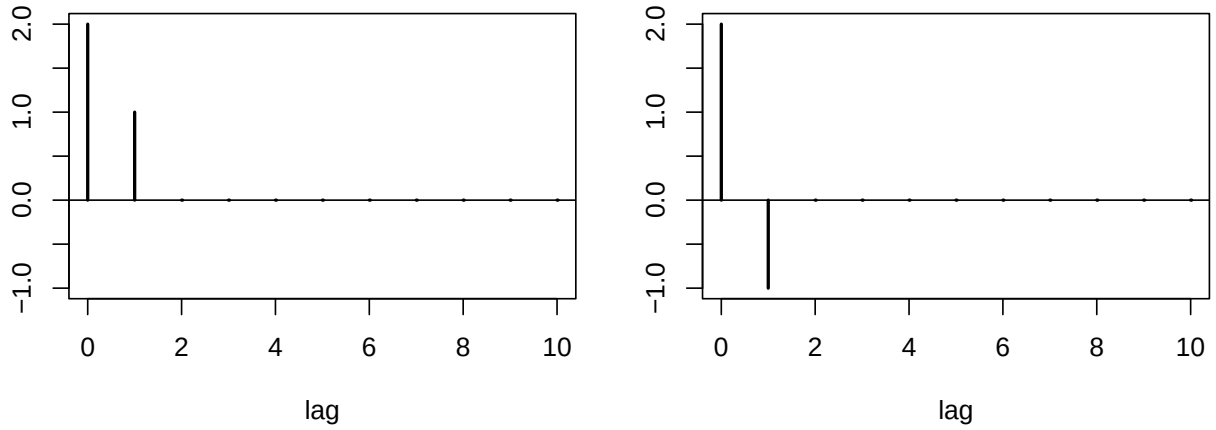

Figure 8: Autocovariance functions of the MA(1) sequence with $b_0 = 1, b_1 = 1$ (left) and $b_0 = 1, b_1 = -1$ (right). In both cases $\sigma^2 = 1$.

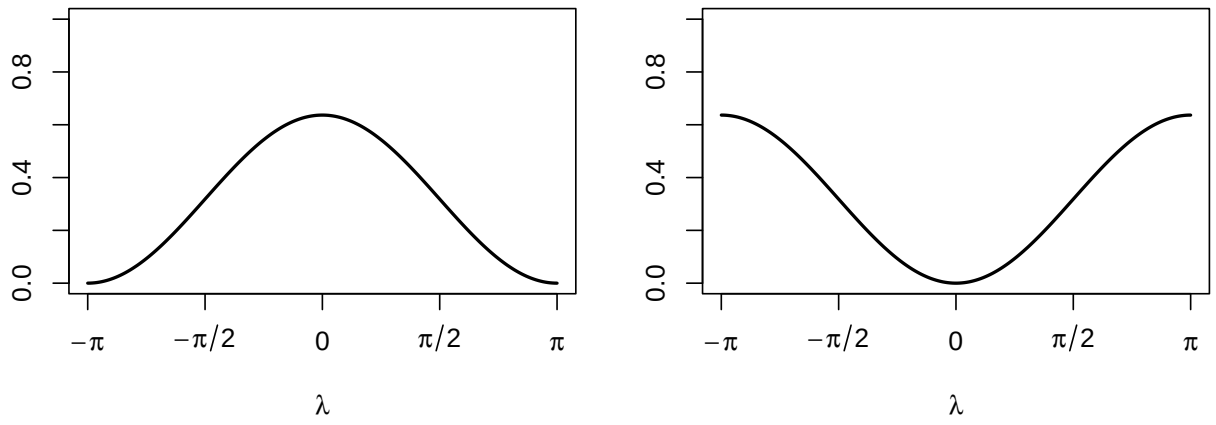

Figure 9: Spectral densities of the MA(1) sequence with $b_0 = 1, b_1 = 1$ (left) and $b_0 = 1, b_1 = -1$ (right). In both cases $\sigma^2 = 1$.

where $g(\lambda) = \sum_{j=0}^{n} b_j \mathrm{e}^{-ij\lambda}, \lambda \in [-\pi, \pi]$. Clearly, $g \in L_2(F_Y)$. From the properties of the stochastic integral we again get $\mathbb{E}X_t = 0, t \in \mathbb{Z}$, and we can finally compute the autocovariance function. For $s, t \in \mathbb{Z}$ we get

$$\begin{aligned}
\mathbb{E}X_{s+t}\overline{X_s} &= \mathbb{E}\left(\int_{-\pi}^{\pi} \mathrm{e}^{i(s+t)\lambda} g(\lambda)\,\mathrm{d}Z_Y(\lambda)\right)\overline{\left(\int_{-\pi}^{\pi} \mathrm{e}^{is\lambda} g(\lambda)\,\mathrm{d}Z_Y(\lambda)\right)} \\
&= \int_{-\pi}^{\pi} \mathrm{e}^{i(s+t)\lambda} g(\lambda)\mathrm{e}^{-is\lambda}\overline{g(\lambda)}\,\mathrm{d}F_Y(\lambda) \\
&= \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}|g(\lambda)|^2 f_Y(\lambda)\,\mathrm{d}\lambda \\
&= \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}|g(\lambda)|^2 \frac{\sigma^2}{2\pi}\,\mathrm{d}\lambda = R_X(t).
\end{aligned}$$

Uniqueness of the spectral decomposition of the autocovariance function (Theorem 5.4) proves that $\frac{\sigma^2}{2\pi}|g(\lambda)|^2, \lambda \in [-\pi, \pi]$, is the spectral density of $\{X_t, t \in \mathbb{Z}\}$. □

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

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

7.3 Linear processes

In this section, we generalize the MA(n) models to MA(∞) models. To do that, we first discuss the summability of white noise sequences, and more generally of weakly stationary sequences (this will be useful later when discussing linear filters and the AR(∞) representation).

Theorem 7.2. *Let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise sequence WN(0,σ^2) and $\{c_j, j \in \mathbb{N}_0\}$ be a sequence of complex-valued constants.*

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

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

2. *If $\sum_{j=0}^{\infty} |c_j| < \infty$, the series $\sum_{j=0}^{\infty} c_j Y_{t-j}$ converges for each $t \in \mathbb{Z}$ absolutely with probability 1.*

Proof. 1. We will show that the sequence $\{\sum_{j=0}^{n} c_j Y_{t-j}, n \in \mathbb{N}\}$ is a Cauchy sequence in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ for every $t \in \mathbb{Z}$. Without loss of generality, we assume $m < n$. Random variables $\{Y_t, t \in \mathbb{Z}\}$ are uncorrelated, with constant variance σ^2, and hence

$$\mathbb{E}\left|\sum_{j=0}^{n} c_j Y_{t-j} - \sum_{j=0}^{m} c_j Y_{t-j}\right|^2 = \mathbb{E}\left|\sum_{j=m+1}^{n} c_j Y_{t-j}\right|^2 = \sum_{j=m+1}^{n} |c_j|^2\,\mathbb{E}|Y_{t-j}|^2 = \sigma^2 \sum_{j=m+1}^{n} |c_j|^2.$$

The right-hand side converges to 0 with $m, n \to \infty$, and it follows that there is the mean square limit of $\{\sum_{j=0}^{n} c_j Y_{t-j}, n \in \mathbb{N}\}$. We denote it $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$.

2. Since $\mathbb{E}|Y_{t-j}| \le \left(\mathbb{E}|Y_{t-j}|^2\right)^{1/2} = \sqrt{\sigma^2} < \infty$ by Cauchy-Schwarz inequality, we get

$$\sum_{j=0}^{\infty} \mathbb{E}|c_j Y_{t-j}| = \sum_{j=0}^{\infty} |c_j|\, \mathbb{E}|Y_{t-j}| \le \sqrt{\sigma^2} \sum_{j=0}^{\infty} |c_j| < \infty,$$

and thus $\sum_{j=0}^{\infty} |c_j Y_{t-j}| < \infty$ almost surely, see Rudin (2003, Theorem 1.38). □

Theorem 7.3. *Let $\{X_t, t \in \mathbb{Z}\}$ be a centered, weakly stationary random sequence with the 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 each $t \in \mathbb{Z}$, the series $\sum_{j=0}^{\infty} c_j X_{t-j}$ converges in the mean square and also absolutely with probability 1.*

Proof. 1. For $m < n$ we have

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

From the weak stationarity and Cauchy-Schwarz inequality, we get

$$\mathbb{E}|X_{t-j}|\,|X_{t-k}| \le \left(\mathbb{E}|X_{t-j}|^2\right)^{1/2} \left(\mathbb{E}|X_{t-k}|^2\right)^{1/2} = R(0),$$

and thus

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

It follows that $\left\{\sum_{j=0}^{n} c_j X_{t-j}, n \in \mathbb{N}\right\}$ is Cauchy in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ and the mean square limit exists.

2. From the weak stationarity we also get

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

and the rest of the proof follows as in the previous theorem. □

Remark: Theorems 7.2 and 7.3 can be extended to $\sum_{j=-\infty}^{\infty} c_j Y_{t-j}$ and $\sum_{j=-\infty}^{\infty} c_j X_{t-j}$, $t \in \mathbb{Z}$, respectively.

Definition 7.3. *Let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise sequence WN(0,σ^2) and $\{c_j, j \in \mathbb{N}_0\}$ be a sequence of complex-valued constants such that $\sum_{j=0}^{\infty} |c_j| < \infty$. The sequence $\{X_t, t \in \mathbb{Z}\}$ given by*

$$X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, \ t \in \mathbb{Z}, \tag{7.2}$$

is called a causal linear process. *We write MA(∞).*

Remark: The causality of $\{X_t, t \in \mathbb{Z}\}$ with respect to $\{Y_t, t \in \mathbb{Z}\}$ means that X_t depends on Y_s for $s \le t$ (present and past values of the white noise), and that X_t does not depend on Y_s for $s > t$ (future values of the white noise, from the perspective of X_t).

Remark: A general linear process can be defined as $X_t = \sum_{j=-\infty}^{\infty} c_j Y_{t-j}, t \in \mathbb{Z}$, for $\sum_{j=-\infty}^{\infty} |c_j| < \infty$.

Remark: The weaker condition $\sum_{j=0}^{\infty} |c_j|^2 < \infty$ implies only the mean square convergence. We impose the stronger condition $\sum_{j=0}^{\infty} |c_j| < \infty$ so that we are able to prove Theorem 7.5 below.

Theorem 7.4. *Let $\{X_t, t \in \mathbb{Z}\}$ be a causal linear process defined by (7.2), where $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2) and $\{c_j, j \in \mathbb{N}_0\}$ is a sequence of complex-valued constants such that $\sum_{j=0}^{\infty} |c_j| < \infty$. The sequence is centered, weakly stationary, with the autocovariance function*

$$R_X(t) = \begin{cases} \sigma^2 \sum_{k=0}^{\infty} c_{k+t}\overline{c_k}, & t \geq 0, \\ \overline{R_X(-t)}, & t \leq 0. \end{cases}$$

The spectral density of $\{X_t, t \in \mathbb{Z}\}$ exists and is given by

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

Proof. Note that:

- $\{X_t^{(n)}, t \in \mathbb{Z}\}$, where $X_t^{(n)} = \sum_{j=0}^{n} c_j Y_{t-j}$, $t \in \mathbb{Z}$, is a MA(n) sequence for each $n \in \mathbb{N}$;
- $\sum_{j=0}^{\infty} |c_j| < \infty$ implies $X_t^{(n)} \to X_t$, $n \to \infty$, in the mean square, for each $t \in \mathbb{Z}$;
- $\{X_t^{(n)}, t \in \mathbb{Z}\}$ is a centered, weakly stationary random sequence with the autocovariance function given in Theorem 7.1;
- $\mathbb{E}X_t = \lim_{n\to\infty} \mathbb{E}X_t^{(n)} = 0$ for each $t \in \mathbb{Z}$, since the mean square convergence preserves the first two moments;
- Theorem 3.4 implies that the autocovariance functions of $\{X_t^{(n)}, t \in \mathbb{Z}\}$ converge to the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$, and the weak stationarity of $\{X_t, t \in \mathbb{Z}\}$ follows, together with the formula for its autocovariance function given in this theorem;
- the spectral decomposition of $\{X_t^{(n)}, t \in \mathbb{Z}\}$ is

  $$X_t^{(n)} = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} g_n(\lambda) \, \mathrm{d}Z_Y(\lambda), \quad t \in \mathbb{Z}, \qquad g_n(\lambda) = \sum_{j=0}^{n} c_j \mathrm{e}^{-ij\lambda}, \quad \lambda \in [-\pi, \pi],$$

  where $g_n \in L_2(F_Y)$ and Z_Y is the orthogonal increment process from the spectral decomposition of the white noise sequence $\{Y_t, t \in \mathbb{Z}\}$ and F_Y is its orthogonal distribution function;
- denote $g(\lambda) = \sum_{j=0}^{\infty} c_j \mathrm{e}^{-ij\lambda}, \lambda \in [-\pi, \pi]$, and compute:

  $$\begin{aligned} \int_{-\pi}^{\pi} |g_n(\lambda) - g(\lambda)|^2 \, \mathrm{d}F_Y(\lambda) &= \int_{-\pi}^{\pi} \left| \sum_{j=n+1}^{\infty} c_j \mathrm{e}^{-ij\lambda} \right|^2 \mathrm{d}F_Y(\lambda) \leq \int_{-\pi}^{\pi} \left(\sum_{j=n+1}^{\infty} |c_j| \right)^2 f_Y(\lambda) \, \mathrm{d}\lambda \\ &= \int_{-\pi}^{\pi} \left(\sum_{j=n+1}^{\infty} |c_j| \right)^2 \frac{\sigma^2}{2\pi} \, \mathrm{d}\lambda = \sigma^2 \left(\sum_{j=n+1}^{\infty} |c_j| \right)^2 \to 0, \quad n \to \infty. \end{aligned}$$

  This means that $g_n \to g, n \to \infty$, in $L_2(F_Y)$. Theorem 6.3 implies that

  $$X_t^{(n)} = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} g_n(\lambda) \, \mathrm{d}Z_Y(\lambda) \to \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} g(\lambda) \, \mathrm{d}Z_Y(\lambda), \quad n \to \infty,$$

  in the mean square. At the same time, $X_t^{(n)} \to X_t, n \to \infty$, in the mean square. Uniqueness of the limit gives that $X_t = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda} g(\lambda) \, \mathrm{d}Z_Y(\lambda), t \in \mathbb{Z}$.

- for $s, t \in \mathbb{Z}$:

$$\mathbb{E}X_{s+t}\overline{X_s} = R_X(s+t,s) = \mathbb{E}\int_{-\pi}^{\pi} \mathrm{e}^{i(s+t)\lambda} g(\lambda)\,\mathrm{d}Z_Y(\lambda) \cdot \overline{\int_{-\pi}^{\pi} \mathrm{e}^{is\lambda} g(\lambda)\,\mathrm{d}Z_Y(\lambda)}$$
$$= \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}|g(\lambda)|^2\,\mathrm{d}F_Y(\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}|g(\lambda)|^2 \frac{\sigma^2}{2\pi}\,\mathrm{d}\lambda.$$

This provides the spectral decomposition of the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$ – its spectral density is $f_X(\lambda) = \frac{\sigma^2}{2\pi}|g(\lambda)|^2$, $\lambda \in [-\pi, \pi]$.

□

Example: Let us consider a causal linear process with $c_j = \varphi^j, j = 0, 1, \ldots$, for some $\varphi \in \mathbb{R}$ with $|\varphi| < 1$. We define $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, t \in \mathbb{Z}$, where $\{Y_t, t \in \mathbb{Z}\}$ is WN(0,σ^2). The sequence $\{X_t, t \in \mathbb{Z}\}$ is a centered, weakly stationary random sequence. Its autocovariance function is

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

Its spectral density is

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

Note that we can write

$$X_t = \sum_{j=0}^{\infty} \varphi^j Y_{t-j} = Y_t + \sum_{j=1}^{\infty} \varphi^j Y_{t-j} = Y_t + \varphi \sum_{j=1}^{\infty} \varphi^{j-1} Y_{t-j}$$
$$= Y_t + \varphi \sum_{k=0}^{\infty} \varphi^k Y_{t-1-k} = Y_t + \varphi X_{t-1},\ t \in \mathbb{Z}.$$

This means that X_t depends linearly on the previous value X_{t-1} and an error term Y_t. Hence, we can call $\{X_t, t \in \mathbb{Z}\}$ the autoregressive sequence of order 1. This motivates the following definition.

7.4 Autoregressive sequences

Definition 7.4. *Let $\varphi_1, \ldots, \varphi_n \in \mathbb{R}, \varphi_n \neq 0$, be constants and let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise sequence WN(0,σ^2). A random sequence $\{X_t, t \in \mathbb{Z}\}$ which satisfies*

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_n X_{t-n} + Y_t,\ t \in \mathbb{Z},$$

is called an autoregressive sequence of order n. *We write AR(n).*

Remark: The assumption $\varphi_n \neq 0$ is needed to ensure that the order of the autoregressive sequence is captured correctly. This will be important when studying the stationarity of autoregressive sequences in the theorem below.

Remark: Equivalently, X_t can be defined by

$$X_t + a_1 X_{t-1} + \ldots + a_n X_{t-n} = \sum_{j=0}^{n} a_j X_{t-j} = Y_t,\ t \in \mathbb{Z}, \tag{7.3}$$

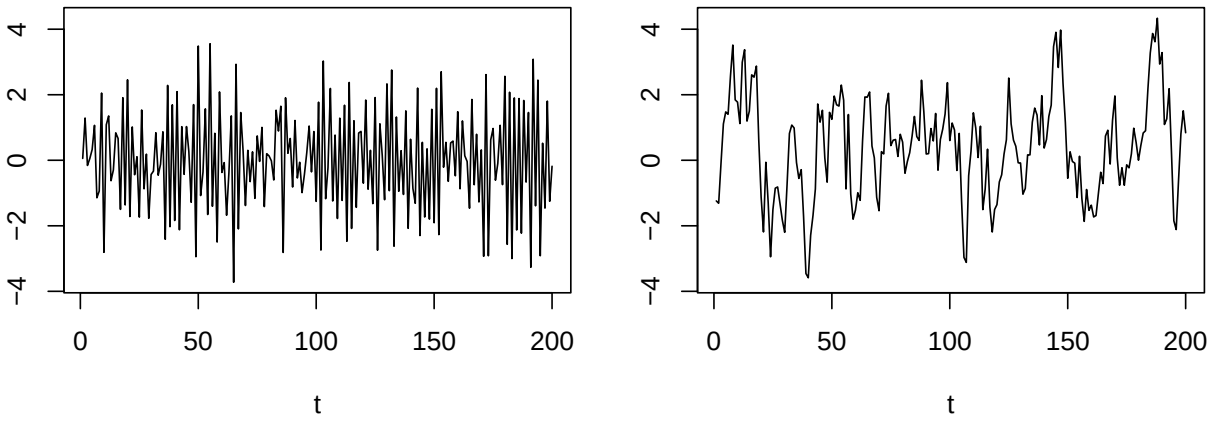

Figure 10: Sample realizations of the AR(1) sequence with $a_1 = 0.8$ (left) and with $a_1 = -0.8$ (right). In both cases $\sigma^2 = 1$. Note that these are discrete-time sequences, and the lines joining the corresponding points in the plots are used only for clarity.

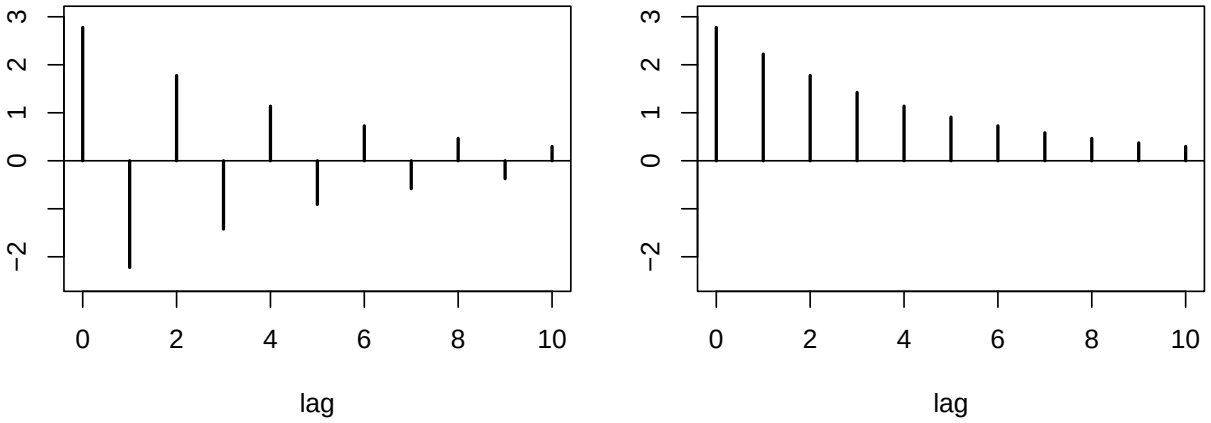

Figure 11: Autocovariance functions of the AR(1) sequence with $a_1 = 0.8$ (left) and with $a_1 = -0.8$ (right). In both cases $\sigma^2 = 1$.

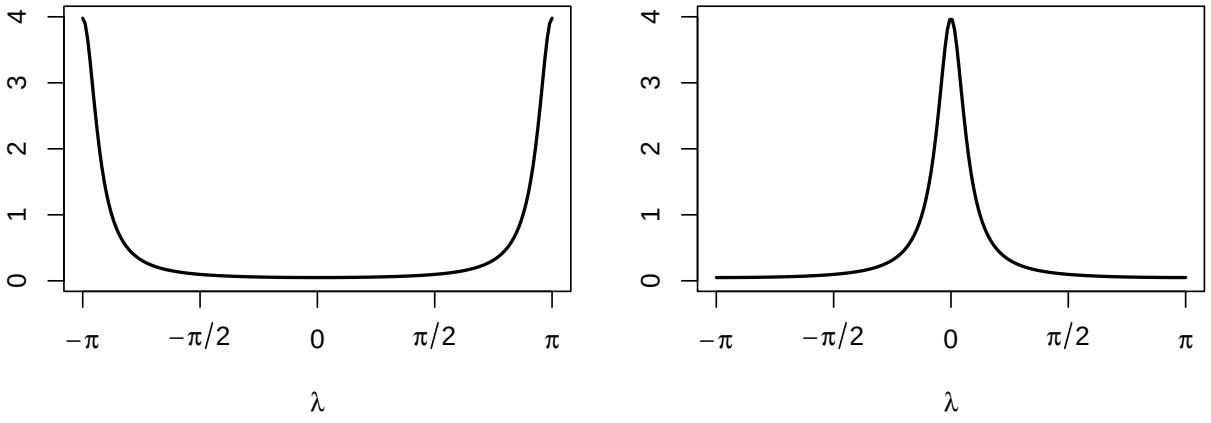

Figure 12: Spectral densities of the AR(1) sequence with $a_1 = 0.8$ (left) and with $a_1 = -0.8$ (right). In both cases $\sigma^2 = 1$.

where $a_0 = 1$. Sample realizations of AR(1) sequences are given in Figure 10. The corresponding autocovariance functions and spectral densities are shown in Figures 11 and 12, respectively.

Remark: Now we look for conditions under which we can express $\{X_t, t \in \mathbb{Z}\}$ as a causal linear process. Such a representation would, among others, imply weak stationarity of the sequence. The key tool will be the *lag operator* B defined below.

Let $\mathcal{S}$ be the vector space of all random sequences $\{X_t, t \in \mathbb{Z}\}$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with the addition defined as $\{X_t, t \in \mathbb{Z}\} + \{Y_t, t \in \mathbb{Z}\} = \{X_t + Y_t, t \in \mathbb{Z}\}$ and multiplication by $c \in \mathbb{C}$ defined as $c \cdot \{X_t, t \in \mathbb{Z}\} = \{cX_t, t \in \mathbb{Z}\}$. The lag operator $B : \mathcal{S} \to \mathcal{S}$ is defined as

$$B\{X_t, t \in \mathbb{Z}\} = \{X_{t-1}, t \in \mathbb{Z}\}.$$

More explicitly, we can write $B\{X_t, t \in \mathbb{Z}\} = \{\widetilde{X}_t, t \in \mathbb{Z}\}$, where $\widetilde{X}_t = X_{t-1}, t \in \mathbb{Z}$. Also, we define B^0 to be the identity operator, and $B^k, k \in \mathbb{N}$, is defined iteratively as $B^k\{X_t, t \in \mathbb{Z}\} = B\left(B^{k-1}\{X_t, t \in \mathbb{Z}\}\right) = \{X_{t-k}, t \in \mathbb{Z}\}$. Finally, B is invertible and we can write $B^{-1}\{X_t, t \in \mathbb{Z}\} = \{X_{t+1}, t \in \mathbb{Z}\}$.

The lag operator B is clearly linear and we can construct polynomial operators such as $a(B) = B^0 + a_1 B + \ldots a_n B^n$, formally identical to the algebraic operator $a(z) = 1 + a_1 z + \ldots + a_n z^n, z \in \mathbb{C}$.

Remark: Using the lag operator, we may rewrite the model equation (7.3) as $a(B)\{X_t, t \in \mathbb{Z}\} = \{Y_t, t \in \mathbb{Z}\}$.

Theorem 7.5. *Let $\{X_t, t \in \mathbb{Z}\}$ be an autoregressive sequence of order n defined by (7.3). $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process if and only if all the roots of the polynomial $a(z) = 1 + a_1 z + \ldots + a_n z^n$ lie outside of the unit circle in $\mathbb{C}$, i.e. $a(z) \neq 0$ for $|z| \leq 1$. If $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process, the coefficients $\{c_j, j \in \mathbb{N}_0\}$ in the representation $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$, $t \in \mathbb{Z}$, are given by*

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

The autocovariance function is

$$R_X(t) = \begin{cases} \sigma^2 \sum_{k=0}^{\infty} c_{k+t}\overline{c_k}, & t \geq 0, \\ \overline{R_X(-t)}, & t \leq 0, \end{cases}$$

and the spectral density is

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

where $a_0 = 1$.

Proof. 1. We assume that all the roots of the polynomial $a(z)$ lie outside of the unit circle in $\mathbb{C}$. Let $z_1, \ldots, z_n \in \mathbb{C}$ be the roots of $a(z)$. Let $\delta > 0$ be such that $\min_{i=1,\ldots,n} |z_i| \geq 1 + \delta > 1$. Denote $M = \{z \in \mathbb{C} : |z| < 1 + \delta\}$, then $a(z) \neq 0$ for $z \in M$. Furthermore, $c(z) = \frac{1}{a(z)}$ is holomorphic on M, i.e. it has derivative in each point of M, see Rudin (2003, p. 221). It follows that $c(z)$ can be expressed as a power series (Rudin, 2003, Theorem 10.16):

$$c(z) = \sum_{j=0}^{\infty} c_j z^j, \ z \in M.$$

This series converges absolutely in every closed circle with radius $r < 1 + \delta$, implying $\sum_{j=0}^{\infty} |c_j| < \infty$ and $c(z)a(z) = 1, |z| \leq 1$, i.e. $c(z)$ and $a(z)$ are inverse operators. Thus,

$$\{X_t, t \in \mathbb{Z}\} = c(B)a(B)\{X_t, t \in \mathbb{Z}\} = c(B)\{Y_t, t \in \mathbb{Z}\}.$$

It follows that $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, t \in \mathbb{Z}$, and we have proved that $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process that satisfies the assumptions of Theorem 7.4. It follows that $\{X_t, t \in \mathbb{Z}\}$ is centered, weakly stationary, with the autocovariance function of the given form, and the spectral density

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} c_j \mathrm{e}^{-ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| c(\mathrm{e}^{-i\lambda}) \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|a(\mathrm{e}^{-i\lambda})|^2} = \frac{\sigma^2}{2\pi} \frac{1}{|\sum_{j=0}^{n} a_j \mathrm{e}^{-ij\lambda}|^2}, \ \lambda \in [-\pi, \pi].$$

2. Now we assume that $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process, i.e. there is a sequence of constants $\{c_j, j \in \mathbb{N}_0\}$ with $\sum_{j=0}^{\infty} |c_j| < \infty$ such that $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$, $t \in \mathbb{Z}$, in other words, $\{X_t, t \in \mathbb{Z}\} = c(B)\{Y_t, t \in \mathbb{Z}\}$.

From the model equation (7.3), we obtain

$$\{Y_t, t \in \mathbb{Z}\} = a(B)\{X_t, t \in \mathbb{Z}\} = a(B)c(B)\{Y_t, t \in \mathbb{Z}\}.$$

We define

$$\eta(z) = a(z)c(z) = \sum_{j=0}^{\infty} \eta_j z^j, \ |z| \le 1,$$

and see that

$$\{Y_t, t \in \mathbb{Z}\} = a(B)c(B)\{Y_t, t \in \mathbb{Z}\} = \eta(B)\{Y_t, t \in \mathbb{Z}\},$$

and hence $Y_t = \sum_{j=0}^{\infty} \eta_j Y_{t-j}, t \in \mathbb{Z}$. Now we compare the coefficients on both sides (formally speaking, on both sides of the equation, we take the inner product with $Y_{t-k}, k = 0, 1, 2, \ldots$) to obtain $\eta_0 = 1, \eta_k = 0, k = 1, 2, \ldots$ It follows that $\eta(z) = 1, |z| \le 1$. Since $|c(z)| < \infty$ for $|z| \le 1$, we have $a(z) \neq 0$ for $|z| \le 1$ and all roots of $a(z)$ lie outside of the unit circle. □

Remark: The coefficients $\{c_j, j \in \mathbb{N}_0\}$ can be obtained by decomposing $c(z) = \frac{1}{a(z)}$ using partial fractions. Assuming for simplicity that all the roots of the polynomial $a(z)$ are simple, we denote them $z_1, \ldots, z_n$. We also assume that the roots all lie outside of the unit circle, i.e. $|z_j| \ge 1+\delta, j = 1, \ldots, n$, for some $\delta > 0$. For $|z| \le 1$ we get

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

for some constants $A_1, \ldots, A_n$. For $|z| \le 1$ and $|z_j| > 1$ we have, for $j = 1, \ldots, n$,

$$\frac{A_j}{z_j - z} = \frac{A_j}{z_j\,(1 - z/z_j)} = \frac{A_j}{z_j} \sum_{k=0}^{\infty} \left(\frac{z}{z_k} \right)^k,$$

and hence

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

where

$$c_k = \sum_{j=1}^{n} \frac{A_j}{z_j^{k+1}}.$$

Since

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

we have $\sum_{k=0}^{\infty} |c_k| < \infty$. Also, for $|z| \leq 1$ it holds that $c(z)a(z) = 1$ and we can write

$$\{X_t, t \in \mathbb{Z}\} = c(B)a(B)\{X_t, t \in \mathbb{Z}\} = c(B)\{Y_t, t \in \mathbb{Z}\}.$$

It follows that $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}, t \in \mathbb{Z}$.

Remark: Another way of obtaining the values of $\{c_k, k \in \mathbb{N}_0\}$ is to compare the coefficients with the same powers of z on both sides of the equation $a(z)c(z) = 1$, where $a_0 = 1$ and $a_1, \ldots, a_n$ are known while $c_0, c_1, \ldots$ are to be determined:

$$\begin{aligned} & c_0 = 1, \\ & c_1 + a_1 c_0 = 0, \\ & c_2 + a_1 c_1 + a_2 c_0 = 0, \\ & \ldots \\ & c_p + a_1 c_{p-1} + \ldots + a_n c_{p-n} = 0, \ p = n, n+1, \ldots \end{aligned} \tag{7.4}$$

Formally, this procedure is equivalent to plugging $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$ into Equation (7.3) and comparing the coefficients with the same Y_{t-k}. To find the coefficients, we need to solve the homogeneous difference equation (7.4) with constant coefficients and initial conditions for $c_0, c_1, \ldots, c_{n-1}$. The initial conditions can be determined from the first n equations above.

Remark: If $z_1, \ldots, z_n$ are the roots of the polynomial $a(z) = 1 + a_1 z + \ldots + a_n z^n$, then $\lambda_i = 1/z_i, i = 1, \ldots, n$, are the roots of the polynomial $L(z) = z^n + a_1 z^{n-1} + \ldots + a_n$. It is easy to prove this directly. Hence, an AR(n) sequence is a causal linear process if and only if all the roots of the polynomial $L(z)$ lie inside of the unit circle in $\mathbb{C}$, i.e. $|\lambda_i| < 1, i = 1, \ldots, n$.

Remark: In general, it is difficult to find the roots of the polynomials $a(z)$ or $L(z)$. There are interesting results concerning the localization of the roots of polynomials inside/outside of the unit circle in $\mathbb{C}$ without determining the actual roots, see e.g. the *Schur-Cohn criterion.*

Yule-Walker equations

While Theorem 7.5 provides an expression for the autocovariance function of the autoregressive sequence, it relies on the knowledge of the coefficients $\{c_j, j \in \mathbb{N}_0\}$. A different way of computing the autocovariance function is provided by the so-called *Yule-Walker equations.* For clarity, we write in the following simply R in place of R_X for the autocovariance function of the autoregressive sequence.

Let the sequence $\{X_t, t \in \mathbb{Z}\}$ satisfy

$$X_t + a_1 X_{t-1} + \ldots + a_n X_{t-n} = Y_t, \ t \in \mathbb{Z}, \tag{7.5}$$

this time with $a_1, \ldots, a_n \in \mathbb{R}$ and $\{Y_t, t \in \mathbb{Z}\}$ being a real-valued white noise sequence WN($0,\sigma^2$). Assume that the sequence $\{X_t, t \in \mathbb{Z}\}$ satisfies the conditions of Theorem 7.5. Hence, $\{X_t, t \in \mathbb{Z}\}$ is a *real-valued* causal linear process, meaning the sequence is centered, weakly stationary and its autocovariance function is symmetric: $R(-t) = R(t), t \in \mathbb{Z}$. We also have the representation $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$, $t \in \mathbb{Z}$.

Since the sequence $\{Y_t, t \in \mathbb{Z}\}$ consists of uncorrelated random variables, continuity of the inner product implies for $s < t, s, t \in \mathbb{Z}$, that

$$\mathbb{E} X_s Y_t = \langle X_s, Y_t \rangle = \lim_{n \to \infty} \left\langle \sum_{j=0}^{n} c_j Y_{s-j}, Y_t \right\rangle = 0.$$

Now we multiply both sides of Equation (7.5) by Y_t and take expectation:

$$\mathbb{E}X_tY_t + a_1\mathbb{E}X_{t-1}Y_t + \ldots + \mathbb{E}X_{t-n}Y_t = \mathbb{E}Y_t^2 = \sigma^2.$$

This means that $\mathbb{E}X_tY_t = \sigma^2, t \in \mathbb{Z}$. Similarly, we multiply both sides of Equation (7.5) by X_{t-k}, $k \in \mathbb{N}_0$, and take expectation:

$$\mathbb{E}X_tX_{t-k} + a_1\mathbb{E}X_{t-1}X_{t-k} + \ldots + \mathbb{E}X_{t-n}X_{t-k} = \mathbb{E}Y_tX_{t-k},\ k = 0, 1, 2, \ldots$$

The sequence $\{X_t, t \in \mathbb{Z}\}$ is centered and real-valued, and we can write $R(s,t) = \mathbb{E}X_sX_t, s, t \in \mathbb{Z}$. Also, weak stationarity means that we can work with the function of a single argument $R(t)$ only. It follows that

$$R(0) + a_1R(1) + \ldots + a_nR(n) = \sigma^2,\ (k = 0) \tag{7.6}$$

$$R(k) + a_1R(k-1) + \ldots + a_nR(k-n) = 0,\ k = 1, 2, \ldots \tag{7.7}$$

These are called the Yule-Walker equations. Note that the coefficients $a_1, \ldots, a_n$ are known here, and the values of R are to be computed. This means that we need to solve the difference equation (7.7) with the initial conditions obtained by solving the set of equations for $k = 0, 1, \ldots, n-1$.

For autoregressive sequences, we can use a trick to reduce the size of the set of equations needed to find the initial conditions for the difference equation. Instead of working with the autocovariance function R, we work with the autocorrelation function r, using the relation $r(t) = R(t)/R(0), t \in \mathbb{Z}$. We divide both sides of Equation (7.7) by $R(0)$ and we get, using $r(0) = 1$ and the symmetry $r(-t) = r(t), t \in \mathbb{Z}$,

$$\begin{aligned} r(1) + a_1 + a_2r(1) + a_3r(2) + \ldots + a_nr(n-1) &= 0, \\ r(2) + a_1r(1) + a_2 + a_3r(1) + \ldots + a_nr(n-2) &= 0, \\ &\ldots \\ r(n-1) + a_1r(n-2) + \ldots + a_nr(1) &= 0. \end{aligned}$$

We solve this system of equations to find the values of $r(1), \ldots, r(n-1)$. This is enough since we already know that $r(0) = 1$. Together, these values serve as the initial conditions for the difference equation

$$r(k) + a_1r(k-1) + \ldots + r(k-n) = 0,\ k \geq n.$$

The characteristic polynomial corresponding to this difference equation is

$$\lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n = L(\lambda),\ \lambda \in \mathbb{C}.$$

Solving the difference equation, we find the values $r(k), k \in \mathbb{N}_0$.

Finally, we plug in $R(k) = r(k)R(0), k \in \mathbb{Z}$, into Equation (7.6) to get

$$R(0)\left(1 + a_1r(1) + \ldots + a_nr(n)\right) = \sigma^2,$$

and hence

$$R(0) = \frac{\sigma^2}{1 + a_1r(1) + \ldots + a_nr(n)}.$$

In this way, all the values of $R(k), k \in \mathbb{Z}$, are now determined.

Example: Let $\{X_t, t \in \mathbb{Z}\}$ be an AR(1) sequence following the model $X_t + aX_{t-1} = Y_t, t \in \mathbb{Z}$, for some $a \in \mathbb{R}$ with $|a| < 1$, where $\{Y_t, t \in \mathbb{Z}\}$ is a real-valued white noise sequence WN(0,σ^2).

The polynomial $a(z) = 1 + az$ has root $-1/a$ which lies outside of the unit circle. It follows that $\{X_t, t \in \mathbb{Z}\}$ is a centered, weakly stationary, causal linear process. Using the procedure described above, we obtain the equations

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

In this case, the initial condition is very simple: $r(0) = 1$. It follows that

$$\begin{aligned} r(1) &= -a, \\ r(2) &= -a\, r(1) = (-a)^2, \\ r(k) &= -a\, r(k-1) = (-a)^k,\ k = 0, 1, 2, \ldots \end{aligned}$$

The value of the variance $R(0)$ is determined from the only equation containing σ^2,

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

In conclusion, $R(k) = \sigma^2 \frac{(-a)^k}{1-a^2}, k = 0, 1, 2, \ldots$ Note that this is the same result as in the example at the end of Section 7.3.

Remark: Consider an AR(1) sequence $X_t + aX_{t-1} = Y_t, t \in \mathbb{Z}$, with $|a| > 1$. This sequence is not causal but we can still investigate its properties. To do that, we rewrite the model as $X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, where $\varphi = -a$. We can write $X_{t+1} = \varphi X_t + Y_{t+1}$ and hence $X_t = \varphi^{-1} X_{t+1} - \varphi^{-1} Y_{t+1}$. Iteratively plugging in the corresponding formulas we obtain

$$X_t = -\varphi^{-1} Y_{t+1} + \varphi^{-1} \left(\varphi^{-1} X_{t+2} - \varphi^{-1} Y_{t+2} \right) = -\varphi^{-1} Y_{t+1} - \varphi^{-2} Y_{t+2} + \varphi^{-2} X_{t+2} = \ldots$$

In this way we obtain the representation $X_t = -\sum_{k=1}^{\infty} \varphi^{-k} Y_{t+k}, t \in \mathbb{Z}$, using the future values of the white noise sequence. The coefficients in the representation are summable in the appropriate way, i.e. $\sum_{k=1}^{\infty} \left| \varphi^k \right| < \infty$, and hence X_t is a well-defined random variable for each $t \in \mathbb{Z}$. From this representation, we get that $\{X_t, t \in \mathbb{Z}\}$ is a centered, weakly stationary sequence, see Section 7.6 on linear filters.

7.5 ARMA sequences

Trying to generalize the concept of autoregressive sequences, we might consider the case where the innovations $\{Y_t, t \in \mathbb{Z}\}$ are not uncorrelated (white noise), but instead have some correlation structure. In the following, we assume that the innovations have the structure of an MA(n) sequence, resulting in a more flexible model. The same class of models is obtained if we take moving averages from an autoregressive sequence.

Definition 7.5. *Let $a_1, \ldots, a_m \in \mathbb{R}, b_1, \ldots, b_n \in \mathbb{R}, a_m \neq 0, b_n \neq 0$, be constants and let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise sequence WN(0,σ^2). A random sequence $\{X_t, t \in \mathbb{Z}\}$ which satisfies*

$$X_t + a_1 X_{t-1} + \ldots + a_m X_{t-m} = Y_t + b_1 Y_{t-1} + \ldots + b_n Y_{t-n},\ t \in \mathbb{Z}, \tag{7.8}$$

is called an ARMA sequence of order m and n. *We write ARMA(m, n).*

Remark: Both the MA(n) and AR(m) models are special cases of ARMA models. Sample realizations of ARMA(1,1) sequences are given in Figure 13. The corresponding autocovariance functions and spectral densities are shown in Figures 14 and 15, respectively.

Remark: Let us define the following polynomials: $a(z) = 1 + a_1 z + \ldots + a_m z^m$ and $b(z) = 1 + b_1 z + \ldots + b_n z^n, z \in \mathbb{C}$. Then, we can write $a(B)\{X_t, t \in \mathbb{Z}\} = b(B)\{Y_t, t \in \mathbb{Z}\}$.

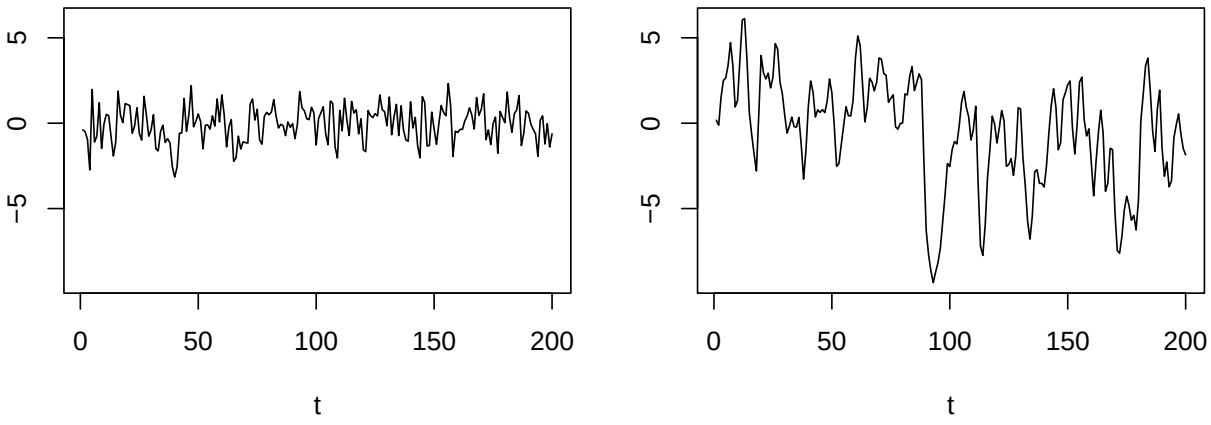

Figure 13: Sample realizations of the ARMA(1,1) sequence with $a_1 = 0.8, b_1 = 1$ (left) and with $a_1 = -0.8, b_1 = 1$ (right). In both cases $\sigma^2 = 1$. Note that these are discrete-time sequences, and the lines joining the corresponding points in the plots are used only for clarity.

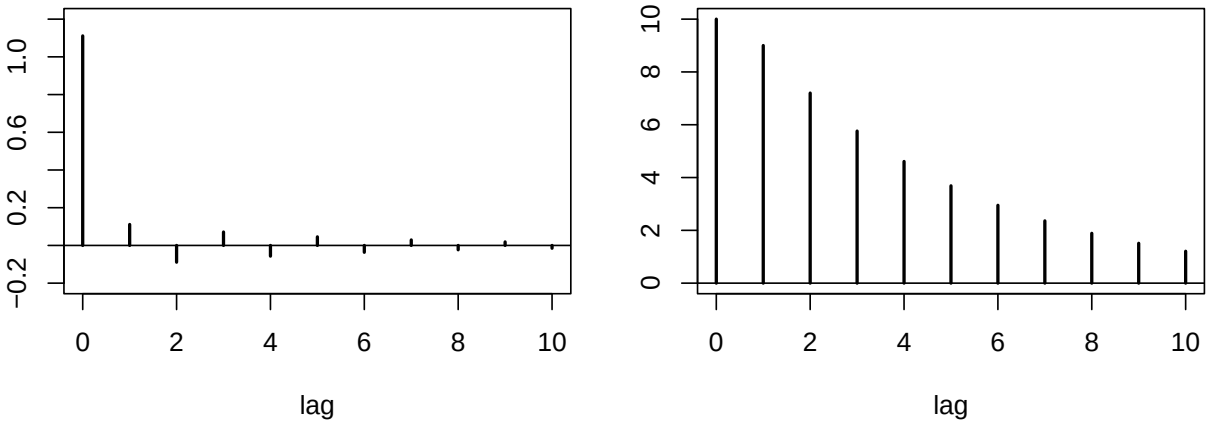

Figure 14: Autocovariance functions of the ARMA(1,1) sequence with $a_1 = 0.8, b_1 = 1$ (left) and with $a_1 = -0.8, b_1 = 1$ (right). In both cases $\sigma^2 = 1$.

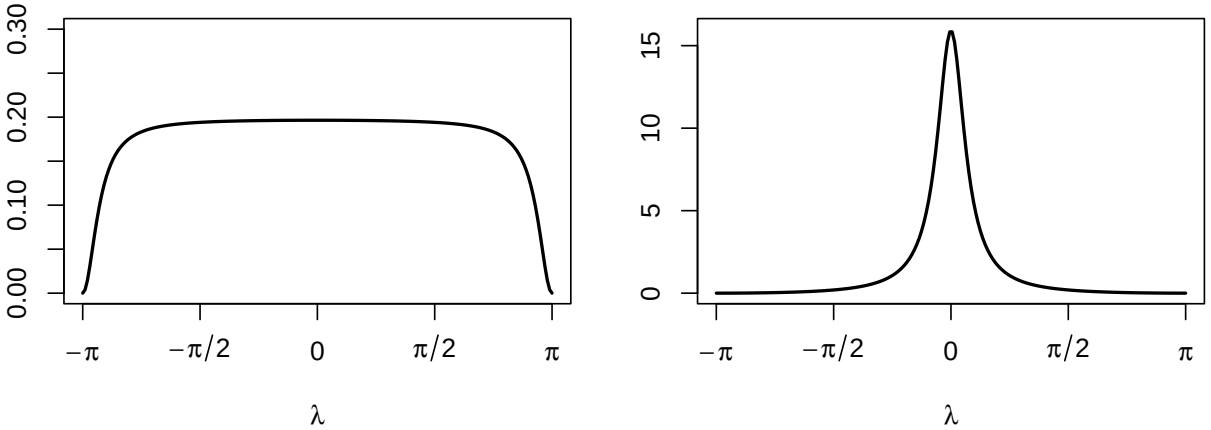

Figure 15: Spectral densities of the ARMA(1,1) sequence with $a_1 = 0.8, b_1 = 1$ (left) and with $a_1 = -0.8, b_1 = 1$ (right). In both cases $\sigma^2 = 1$.

Causality of ARMA sequences

Theorem 7.6. *Let $\{X_t, t \in \mathbb{Z}\}$ be an ARMA(m,n) sequence defined by (7.8) and assume that the polynomials $a(z) = 1 + a_1 z + \ldots + a_m z^m$ and $b(z) = 1 + b_1 z + \ldots + b_n z^n$ have no common roots. $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process if and only if all the roots of the polynomial $a(z)$ lie outside of the unit circle in $\mathbb{C}$, i.e. $a(z) \neq 0$ for $|z| \leq 1$.*

If $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process, the coefficients $\{c_j, j \in \mathbb{N}_0\}$ in the representation $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$, $t \in \mathbb{Z}$, are given by

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

the autocovariance function is

$$R_X(t) = \begin{cases} \sigma^2 \sum_{k=0}^{\infty} c_{k+t}\overline{c_k}, & t \geq 0, \\ \overline{R_X(-t)}, & t \leq 0, \end{cases}$$

and the spectral density is

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

where $a_0 = b_0 = 1$.

Proof. Similar to the proof of Theorem 7.5. □

Remark: If the polynomials $a(z), b(z)$ have common roots which lie outside of the unit circle, then the rational function $c(z) = \frac{b(z)}{a(z)}$, after canceling the terms corresponding to the common roots, defines an ARMA(p,q) model with $p < m, q < n$. If at least one of the common roots lies in the unit circle, there may be more than one weakly stationary solution (Brockwell and Davis, 2006, p. 86).

Yule-Walker equations

Assuming the ARMA sequence $\{X_t, t \in \mathbb{Z}\}$ from Theorem 7.6 is a causal linear process, we multiply the representation $X_t = \sum_{j=0}^{\infty} c_j Y_{t-j}$, $t \in \mathbb{Z}$, by $Y_{t-k}, k \in \mathbb{Z}$, and by taking expectation of both sides we obtain $\mathbb{E}X_t Y_{t-k} = c_k \sigma^2$ for $k \geq 0$, and $\mathbb{E}X_t Y_{t-k} = 0$ otherwise (causality implies that $\mathbb{E}X_s Y_t = 0$ for each $s < t$).

Now we multiply both sides of Equation (7.8) by $X_{t-k}, k \in \mathbb{N}_0$, and take expectation:

$$\mathbb{E}X_t X_{t-k} + \sum_{j=1}^{m} a_j \mathbb{E}X_{t-j} X_{t-k} = \mathbb{E}Y_t X_{t-k} + \sum_{j=1}^{n} b_j \mathbb{E}Y_{t-j} X_{t-k}, \ k = 0, 1, 2, \ldots$$

As in Section 7.4, we assume that $\{Y_t, t \in \mathbb{Z}\}$ is real-valued, meaning that $\{X_t, t \in \mathbb{Z}\}$ is real-valued and its autocovariance function is symmetric: $R(-k) = R(k), k \in \mathbb{Z}$. It follows that

$$R(k) + \sum_{j=1}^{m} a_j R(k-j) = \sigma^2 \sum_{j=k}^{n} b_j c_{j-k}, \ (k \leq n) \tag{7.9}$$

$$R(k) + \sum_{j=1}^{m} a_j R(k-j) = 0, \ (k > n) \tag{7.10}$$

where $b_0 = 1$. For $k \geq n+1$ (to have 0 on the right-hand side) and $k \geq m$ (so that on the left-hand side we avoid the restrictions such as $R(-1) = R(1)$) we solve the difference equation (7.10) with the initial conditions obtained from the system of linear equations with $k < \max(m, n+1)$.

Example: Consider an ARMA(1,1) sequence following the model $X_t + aX_{t-1} = Y_t + bY_{t-1}, t \in \mathbb{Z}$, where $a \neq b, a \neq 0, b \neq 0, |a| < 1$, and $\{Y_t, t \in \mathbb{Z}\}$ is a real-valued white noise sequence WN(0,σ^2). The sequence $\{X_t, t \in \mathbb{Z}\}$ is causal, and hence $\mathbb{E}X_tY_s = 0$ for $s > t$.

Multiplying the model equation by Y_t and taking the expectation, we obtain, for each $t \in \mathbb{Z}$,

$$\mathbb{E}X_tY_t = \sigma^2.$$

Similarly, we multiply the model equation by Y_{t-1} and take the expectation to get

$$\begin{aligned}
&\mathbb{E}X_tY_{t-1} + a\mathbb{E}X_{t-1}Y_{t-1} = b\sigma^2,\\
&\mathbb{E}X_tY_{t-1} + a\sigma^2 = b\sigma^2,\\
&\mathbb{E}X_tY_{t-1} = \sigma^2(b-a).
\end{aligned}$$

Furthermore, we multiply by X_t:

$$\begin{aligned}
&\mathbb{E}X_t^2 + a\mathbb{E}X_tX_{t-1} = \mathbb{E}X_tY_t + b\mathbb{E}X_tY_{t-1},\\
&R(0) + aR(1) = \sigma^2 + b\sigma^2(b-a) = \sigma^2(1 + b(b-a)).
\end{aligned}$$

Next, we multiply by X_{t-1}:

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

Finally, we multiply by X_{t-k} for $k \geq 2$:

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

This is a homogeneous difference equation of order 1. Its general solution is $R(k) = (-a)^{k-1}R(1)$, $k \geq 1$. The initial condition is obtained by solving the system of two linear equations above containing $R(0), R(1)$ and σ^2. The solution is

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

Invertibility of ARMA sequences

We have seen that under certain conditions an ARMA sequence is causal, i.e. it is possible to express it as $X_t = \sum_{j=0}^{\infty} c_jY_{t-j}, t \in \mathbb{Z}$. Now we will investigate if it is possible to invert such representation, i.e. to write $Y_t = \sum_{j=0}^{\infty} d_jX_{t-j}, t \in \mathbb{Z}$. This would be useful when making predictions, see Section 9.3.

Definition 7.6. *Let $\{X_t, t \in \mathbb{Z}\}$ be an ARMA(m,n) sequence defined by Equation (7.8), where $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). The sequence $\{X_t, t \in \mathbb{Z}\}$ is called* invertible, *if there is a sequence of constants $\{d_j, j \in \mathbb{N}_0\}$ such that $\sum_{j=0}^{\infty} |d_j| < \infty$ and $Y_t = \sum_{j=0}^{\infty} d_jX_{t-j}, t \in \mathbb{Z}$.*

Remark: If $\{X_t, t \in \mathbb{Z}\}$ is weakly stationary, e.g. under causality, the definition is correct since by Theorem 7.3 we obtain that $\sum_{j=0}^{\infty} d_jX_{t-j}$ converges almost surely for any $t \in \mathbb{Z}$ provided that $\sum_{j=0}^{\infty} |d_j| < \infty$.

Theorem 7.7. *Let $\{X_t, t \in \mathbb{Z}\}$ be a weakly stationary ARMA(m,n) sequence defined by (7.8). Let the polynomials $a(z)$ and $b(z)$ have no common roots. $\{X_t, t \in \mathbb{Z}\}$ is invertible if and only if all the roots of the polynomial $b(z)$ lie outside of the unit circle in $\mathbb{C}$, i.e. $b(z) \neq 0$ for $|z| \leq 1$.*

If $\{X_t, t \in \mathbb{Z}\}$ is invertible, the coefficients $\{d_j, j \in \mathbb{N}_0\}$ in the representation $Y_t = \sum_{j=0}^{\infty} d_j X_{t-j}$, $t \in \mathbb{Z}$, are given by

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

Proof. Similar to the proof of Theorem 7.5. □

Remark: By taking $z = 0$ we see that $d(0) = a(0)/b(0)$ and hence $d_0 = 1$. It follows that $X_t + \sum_{j=1}^{\infty} d_j X_{t-j} = Y_t, t \in \mathbb{Z}$. We call this *the AR$(\infty)$ representation of an invertible, weakly stationary ARMA(m,n) sequence.*

7.6 Linear filters

Definition 7.7. *Let $\{Y_t, t \in \mathbb{Z}\}$ be a centered, weakly stationary random sequence. Let $\{c_j, j \in \mathbb{Z}\}$ be a sequence of complex-valued constants 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 the sequence $\{Y_t, t \in \mathbb{Z}\}$, if*

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

The sequence $\{c_j, j \in \mathbb{Z}\}$ is called a time-invariant linear filter. *Provided that $c_j = 0$ for each $j < 0$, the filter is called* causal.

Remark: The sum in the definition above is absolutely convergent almost surely, see Theorem 7.3 and the remark below it.

Remark: Linear filters are one of the key tools in the field of signal processing. For illustration, consider a simple low-pass filter (which removes high frequencies) applied to a noisy signal, see Figure 16. While not perfectly reconstructing the original periodic signal, the filtering successfully removes the noise component of the observed sequence.

Remark: Finding the coefficients of a linear filter with the desired properties (transfer function) is not a trivial task, but established methods are available, see e.g. the Parks-McClellan algorithm.

Remark: The linear process MA(∞) is obtained by causal filtration of a white-noise sequence. This means that Theorem 7.4 is a special case of the following theorem.

Theorem 7.8. *Let $\{Y_t, t \in \mathbb{Z}\}$ be a centered, weakly stationary random sequence with the autocovariance function R_Y and the spectral density f_Y. Let $\{c_j, j \in \mathbb{Z}\}$ be a linear filter such that $\sum_{j=-\infty}^{\infty} |c_j| < \infty$. Then the sequence $\{X_t, t \in \mathbb{Z}\}$ defined as $X_t = \sum_{j=-\infty}^{\infty} c_j Y_{t-j}, t \in \mathbb{Z}$, is a centered, weakly stationary sequence with the autocovariance function*

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

and the spectral density

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

where $\Psi(\lambda) = \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{-ik\lambda}, \lambda \in [-\pi, \pi]$, is called the transfer function *of the linear filter.*

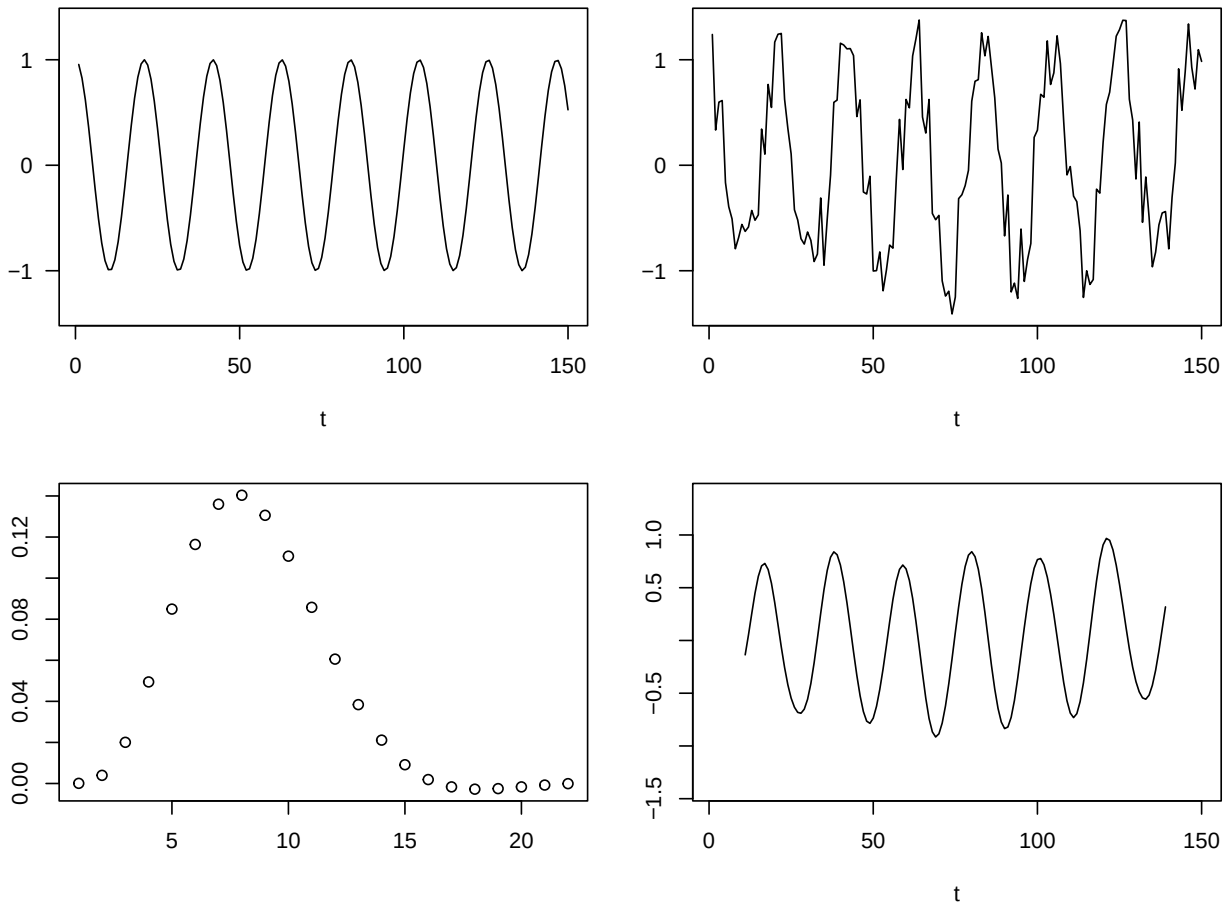

Figure 16: Top left: a periodic signal. Top right: observation $\{Y_t, t \in \mathbb{Z}\}$ of the periodic signal with additive noise. Bottom left: coefficients of the low-pass filter $\{c_j, j \in \mathbb{Z}\}$, only the non-zero values are shown. Bottom right: the resulting sequence $\{X_t, t \in \mathbb{Z}\}$. Note that these are discrete-time sequences, and the lines joining the corresponding points in the plots are used only for clarity.

Proof. The proof is analogous to the proof of Theorem 7.4, starting with $X_t = \text{l.i.m.}\, X_t^{(n)}$, where $X_t^{(n)} = \sum_{j=-n}^{n} c_j Y_{t-j}$, $t \in \mathbb{Z}$. □

Example: Let $\{Y_t, t \in \mathbb{Z}\}$ be a white noise sequence, and define $\{X_t, t \in \mathbb{Z}\}$ by $X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, for $|\varphi| > 1$. We have shown before that the sequence $\{X_t, t \in \mathbb{Z}\}$ is not causal, but we can write $X_t = -\sum_{k=1}^{\infty} \varphi^{-k} Y_{t+k}, t \in \mathbb{Z}$. This means that $\{X_t, t \in \mathbb{Z}\}$ is obtained from $\{Y_t, t \in \mathbb{Z}\}$ by linear filtration with coefficients

$$c_k = \begin{cases} 0, & k \geq 0, \\ -\varphi^k, & k < 0. \end{cases}$$

Since these coefficients are absolutely summable, it follows by Theorem 7.8 that $\{X_t, t \in \mathbb{Z}\}$ is a weakly stationary sequence, i.e. the behavior of the sequence is nice. The autocovariance function of $\{X_t, t \in \mathbb{Z}\}$ can be easily determined from Theorem 7.8, too.

Remark: Note that causality is a property of the pair $(\{X_t, t \in \mathbb{Z}\}, \{Y_t, t \in \mathbb{Z}\})$, not the sequence $\{X_t, t \in \mathbb{Z}\}$ alone. In fact, it can be shown that any non-causal AR(1) sequence ($X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, with $|\varphi| > 1$) is a causal AR(1) sequence with respect to a different white noise sequence $\{\widetilde{Y}_t, t \in \mathbb{Z}\}$.

8 Ergodicity

Definition 8.1. *Let the random sequence $\{X_t, t \in \mathbb{Z}\}$ be weakly stationary with mean value μ. The sequence is* mean square ergodic *or* satisfies the law of large numbers in $L_2(\Omega, \mathcal{A}, \mathbb{P})$, *if*

$$\frac{1}{n}\sum_{t=1}^{n} X_t \to \mu,\ n \to \infty,$$

in the mean square.

Remark: If $\{X_t, t \in \mathbb{Z}\}$ is a mean square ergodic sequence, then $\frac{1}{n}\sum_{t=1}^{n} X_t \to \mu, n \to \infty$, in probability, i.e. $\{X_t, t \in \mathbb{Z}\}$ satisfies the weak law of large numbers for weakly stationary sequences, and $\frac{1}{n}\sum_{t=1}^{n} X_t$ is a weakly consistent estimator of μ.

Theorem 8.1. *Let the random sequence $\{X_t, t \in \mathbb{Z}\}$ be weakly stationary with mean value μ and autocovariance function R. The sequence is mean square ergodic if and only if*

$$\frac{1}{n}\sum_{t=1}^{n} R(t) \to 0,\ n \to \infty.$$

Proof. We assume that $\mu = 0$, otherwise we consider the centered version of the sequence $\widetilde{X_t} = X_t - \mu, t \in \mathbb{Z}$, which has the same autocovariance function as the original sequence $\{X_t, t \in \mathbb{Z}\}$.

First, consider the spectral decomposition

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

where $\{Z_\lambda, \lambda \in [-\pi, \pi]\}$ is a centered, orthogonal increment process with the orthogonal distribution function F, which is the same as the spectral distribution function of $\{X_t, t \in \mathbb{Z}\}$, see Theorem 6.5. Then

$$\frac{1}{n}\sum_{t=1}^{n} X_t = \frac{1}{n}\sum_{t=1}^{n}\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}Z(\lambda) = \int_{-\pi}^{\pi}\left(\frac{1}{n}\sum_{t=1}^{n}\mathrm{e}^{it\lambda}\right)\mathrm{d}Z(\lambda) = \int_{-\pi}^{\pi} h_n(\lambda)\,\mathrm{d}Z(\lambda),$$

where

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

Furthermore, consider the function

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

and define the random variable $\widetilde{Z}_0 = \int_{-\pi}^{\pi} h(\lambda)\,\mathrm{d}Z(\lambda)$.

Clearly, $h_n(\lambda) \to h(\lambda), n \to \infty$, for any $\lambda \in [-\pi, \pi]$. It follows that $|h_n(\lambda) - h(\lambda)|^2 \to 0, n \to \infty$, $\lambda \in [-\pi, \pi]$. Also, $h_n \to h, n \to \infty$, in $L_2(F)$, since $|h_n(\lambda) - h(\lambda)|^2 \leq 1$, and using the Lebesgue dominated convergence theorem we get

$$\int_{-\pi}^{\pi} |h_n(\lambda) - h(\lambda)|^2\,\mathrm{d}F(\lambda) \to 0,\ n \to \infty.$$

Theorem 6.3 implies that

$$\frac{1}{n}\sum_{t=1}^{n} X_t = \int_{-\pi}^{\pi} h_n(\lambda)\,\mathrm{d}Z(\lambda) \to \int_{-\pi}^{\pi} h(\lambda)\,\mathrm{d}Z(\lambda) = \widetilde{Z}_0,\ n\to\infty,$$

in the mean square. Now, it suffices to show that

$$\widetilde{Z}_0 = 0 \text{ a.s. } \iff \frac{1}{n}\sum_{t=1}^{n} R(t) \to 0,\ n\to\infty.$$

From Theorem 6.3 we have $\mathbb{E}\widetilde{Z}_0 = 0$, and it follows that $\widetilde{Z}_0 = 0$ a.s. $\iff \mathbb{E}|\widetilde{Z}_0|^2 = 0$. From the same theorem we get

$$\mathbb{E}|\widetilde{Z}_0|^2 = \mathbb{E}\left|\int_{-\pi}^{\pi} h(\lambda)\,\mathrm{d}Z(\lambda)\right|^2 = \int_{-\pi}^{\pi} |h(\lambda)|^2 \mathrm{d}F(\lambda).$$

From the spectral decomposition of the autocovariance function and the Lebesgue dominated convergence theorem we obtain that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} R(t) &= \frac{1}{n}\sum_{t=1}^{n}\left(\int_{-\pi}^{\pi} \mathrm{e}^{it\lambda}\,\mathrm{d}F(\lambda)\right) = \int_{-\pi}^{\pi}\left(\frac{1}{n}\sum_{t=1}^{n}\mathrm{e}^{it\lambda}\right)\mathrm{d}F(\lambda) = \int_{-\pi}^{\pi} h_n(\lambda)\,\mathrm{d}F(\lambda)\\ &\to \int_{-\pi}^{\pi} h(\lambda)\,\mathrm{d}F(\lambda) = \int_{-\pi}^{\pi} |h(\lambda)|^2\,\mathrm{d}F(\lambda) = \mathbb{E}|\widetilde{Z}_0|^2,\ n\to\infty.\end{aligned}$$

□

Remark: Note that $\mathbb{E}|\widetilde{Z}_0|^2 = \int_{-\pi}^{\pi} |h(\lambda)|^2\,\mathrm{d}F(\lambda) = F(0) - F(0-)$, meaning the condition $\widetilde{Z}_0 = 0$ a.s. (and also $\frac{1}{n}\sum_{t=1}^{n} R(t) \to 0, n\to\infty$) is fulfilled if and only if the spectral distribution function of $\{X_t, t\in\mathbb{Z}\}$ is continuous at 0.

Example: Let us consider the AR(1) sequence following the model $X_t = \varphi X_{t-1} + Y_t, t\in\mathbb{Z}$, where $\{Y_t, t\in\mathbb{Z}\}$ is a white noise sequence WN(0,σ^2) and $|\varphi| < 1$. We know that $R(t) = \frac{\sigma^2}{1-\varphi^2}\varphi^{|t|}, t\in\mathbb{Z}$. Obviously,

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

implying that the sequence $\{X_t, t\in\mathbb{Z}\}$ is mean square ergodic.

Example: Let $\{X_t, t\in\mathbb{Z}\}$ be a sequence of independent, identically distributed random variables with finite second moments. Since $R(0) = \sigma^2$ and $R(k) = 0$ for each $k\neq 0$, the summability condition in Theorem 8.1 is satisfied and the sequence $\{X_t, t\in\mathbb{Z}\}$ is mean square ergodic. In this way, we recover the weak law of large numbers known from the i.i.d. setting.

Example: Let $\{X_t, t\in\mathbb{Z}\}$ be a weakly stationary, mean square ergodic random sequence with expected value μ and autocovariance function R_X. We define $\{Z_t, t\in\mathbb{Z}\}$ as $Z_t = X_t + Y, t\in\mathbb{Z}$, where $\mathbb{E}Y = 0$, $\operatorname{var} Y = \sigma^2 \in (0,\infty)$, and Y is uncorrelated with $\{X_t, t\in\mathbb{Z}\}$. It follows that $\mathbb{E}Z_t = \mu, t\in\mathbb{Z}$, and the autocovariance function of the sequence $\{Z_t, t\in\mathbb{Z}\}$ is $R_Z(t) = R_X(t) + \sigma^2, t\in\mathbb{Z}$. The sequence is weakly stationary. However, it is not mean square ergodic:

$$\frac{1}{n}\sum_{t=1}^{n} R_Z(t) = \frac{1}{n}\sum_{t=1}^{n} R_X(t) + \sigma^2 \to \sigma^2 > 0,\ n\to\infty.$$

Also,

$$\frac{1}{n}\sum_{t=1}^{n} Z_t = \frac{1}{n}\sum_{t=1}^{n} X_t + Y \to \mu + Y,\ n\to\infty,$$

in the mean square, and the sample means do not converge to any constant.

Theorem 8.2. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, weakly stationary random sequence with mean value μ and autocovariance function R such that $\sum_{t=-\infty}^{\infty} |R(t)| < \infty$. Then*

1. $\frac{1}{n}\sum_{k=1}^{n} X_k \to \mu, n \to \infty$, *in the mean square,*
2. $n \operatorname{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \to \sum_{k=-\infty}^{\infty} R(k), n \to \infty$.

Proof. 1. The assumption $\sum_{t=-\infty}^{\infty} |R(t)| < \infty$ implies that

$$\left|\frac{1}{n}\sum_{k=1}^{n} R(k)\right| \le \frac{1}{n}\sum_{k=1}^{n} |R(k)| \le \frac{1}{n}\sum_{k=1}^{\infty} |R(k)| \to 0, \ n \to \infty,$$

meaning that $\frac{1}{n}\sum_{k=1}^{n} R(k) \to 0$, $n \to \infty$. The first claim then follows from Theorem 8.1.

2. It is easy to see that

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

We assume that $\{X_t, t \in \mathbb{Z}\}$ is real-valued and hence $R(j-k) = R(k-j)$. We can write

$$\sum\sum_{1\le j\ne k\le n} R(j-k) = 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} R(j-k) = 2\sum_{l=1}^{n-1}(n-l)R(l),$$

where in the last equality we changed the order of summation and noticed that some of the terms have the same value. We now proceed with computing the variance:

$$\begin{aligned}\operatorname{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) &= \frac{1}{n^2}\left(nR(0) + 2\sum_{l=1}^{n-1}(n-l)R(l)\right) = \frac{1}{n}\left(R(0) + 2\sum_{l=1}^{n-1}\left(1-\frac{l}{n}\right)R(l)\right) \\ &= \frac{1}{n}\sum_{j=-n+1}^{n-1}\left(1-\frac{|j|}{n}\right)R(j),\end{aligned}$$

where we used again the symmetry of the autocovariance function. We conclude that

$$n \operatorname{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \sum_{j=-n+1}^{n-1} R(j) - \frac{2}{n}\sum_{j=1}^{n-1} jR(j) \to \sum_{j=-\infty}^{\infty} R(j), \ n \to \infty,$$

since the second term converges to 0 with increasing n by the Kronecker lemma. □

Remark: Theorem 8.2 also implies that $n \operatorname{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \to \sum_{k=-\infty}^{\infty} R(k) = 2\pi f(0), n \to \infty$ (recall the inverse formula for computing the spectral density in Theorem 5.7).

9 Prediction in the time domain

9.1 Projections in Hilbert spaces

Definition 9.1. *Let H be a Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| = \sqrt{\langle x, x \rangle}$, $x \in H$. We say that two elements $x, y \in H$ are* orthogonal *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 each element of M, i.e. $\langle x, y \rangle = 0$ for each $y \in M$. We write $x \perp M$.*

The set $M^\perp = \{y \in H : y \perp M\}$ is called the orthogonal complement *of the set M.*

Theorem 9.1. *Let H be a Hilbert space and $M \subset H$ be any subset of H. Then $M^\perp$ is a closed subset of H.*

Proof. Denote the null element in H by o. Since $\langle o, x \rangle = 0$ for each $x \in M$, we have $o \in M^\perp$. Linearity of the inner product implies that any linear combination of elements of $M^\perp$ is an element of $M^\perp$, and hence $M^\perp$ is a subspace of H. Continuity of the inner product implies that any limit of a sequence of elements of $M^\perp$ is an element of $M^\perp$ and hence $M^\perp$ is a closed subspace. □

Theorem 9.2 (projection theorem)**.** *Let M be a closed subspace of a Hilbert space H. Then for every element $x \in H$ there is a unique decomposition $x = \widehat{x} + (x - \widehat{x})$ such that $\widehat{x} \in M$ and $x - \widehat{x} \perp M$. Furthermore,*

$$\begin{aligned} \|x - \widehat{x}\| &= \min_{y \in M} \|x - y\|, \\ \|x\|^2 &= \|\widehat{x}\|^2 + \|x - \widehat{x}\|^2. \end{aligned} \tag{9.1}$$

Proof. See Rudin (2003, Theorem 4.11) or Brockwell and Davis (2006, Theorem 2.3.1). □

Remark: The element $\widehat{x} \in M$ is called *the orthogonal projection of* x onto the subspace M. The mapping $P_M : H \to M$ such that $\widehat{x} = P_M x \in M$ and $x - \widehat{x} = (I - P_M)x \in M^\perp$, where I is the identity mapping, is called *the projection mapping*. Obviously, for any $x \in H$ we have a unique decomposition

$$x = P_M x + (x - P_M x) = P_M x + (I - P_M)\, x. \tag{9.2}$$

Theorem 9.3 (properties of the projection mapping)**.** *Let H be a Hilbert space and let P_M be the projection mapping onto a closed subspace M. It holds that:*

1. *for every $x, y \in H$ and $\alpha, \beta \in \mathbb{C}$, $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$,*
2. *if $x \in M$, then $P_M x = x$,*
3. *if $x \in M^\perp$, then $P_M x = o$,*
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 $x_n, x \in H$ are elements of H such that $\|x_n - x\| \to 0, n \to \infty$, then $\|P_M x_n - P_M x\| \to 0$, $n \to \infty$.*

Proof. 1. We see that

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

Obviously, $\alpha P_M x + \beta P_M y \in M$ and $\alpha(x - P_M x) + \beta(y - P_M y) \in M^\perp$, since M and $M^\perp$ are linear subspaces. From this decomposition we get $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$.

2. The claim follows directly from the uniqueness of decomposition (9.2), since $x = x + o$, where $x \in M$ and $o \in M^\perp$.

3. Similarly as above, we have $x = o + x$, where $o \in M$ and $x \in M^\perp$.

4. We have $x = P_{M_2}x + (x - P_{M_2}x)$, $P_{M_2}x \in M_2$, $x - P_{M_2}x \in M_2^\perp$, and thus, using linearity from the point 1 above, $P_{M_1}x = P_{M_1}(P_{M_2}x) + P_{M_1}(x - P_{M_2}x)$. It follows from the definition of the projection mapping that $P_{M_1}(P_{M_2}x) \in M_1$. Also, from $M_2^\perp \subseteq M_1^\perp$ we get $P_{M_1}(x - P_{M_2}x) = o$ using the point 3 above. It follows that $P_{M_1}x = P_{M_1}P(_{M_2}x) + o$ for each $x \in H$.

5. Using the linearity of the projection mapping and Equation (9.1) we get

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

Since both terms on the right-hand side are non-negative, we obtain

$$\|P_M x_n - P_M x\|^2 = \|P_M(x_n - x)\|^2 \leq \|x_n - x\|^2 \to 0,\ n \to \infty.$$

□

9.2 Prediction based on finite history

Consider the following problem: we observe random variables $X_1, \ldots, X_n$ with zero mean and finite second moments, and we want to predict (forecast) the value of X_{n+h} with $h \in \mathbb{N}$.

We would like to approximate (estimate) X_{n+h} by a measurable function $g(X_1, \ldots, X_n)$ of the observations $X_1, \ldots, X_n$ which minimizes the mean square error $\mathbb{E}|X_{n+h} - g(X_1, \ldots, X_n)|^2$. It is well known that the solution is given by the conditional expectation:

$$g(X_1, \ldots, X_n) = \mathbb{E}\left[X_{n+h}|X_1, \ldots, X_n\right].$$

Indeed, let us consider for simplicity only real-valued random variables and denote $\mathbb{X}_n = (X_1, \ldots, X_n)^T$. Then,

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

The last term equals

$$\begin{aligned}&2\mathbb{E}\left[\mathbb{E}[(X_{n+h} - \mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right])(\mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right] - g(\mathbb{X}_n))|\mathbb{X}_n]\right] \\ &= 2\mathbb{E}\left[(\mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right] - g(\mathbb{X}_n))\mathbb{E}[X_{n+h} - \mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right]|\mathbb{X}_n]\right] \\ &= 0,\end{aligned}$$

since $(\mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right] - g(\mathbb{X}_n))$ is measurable with respect to the σ-algebra generated by $\mathbb{X}_n$ and $\mathbb{E}[X_{n+h} - \mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right]|\mathbb{X}_n] = 0$ a.s. It follows that

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

with equality for $g(\mathbb{X}_n) = \mathbb{E}\left[X_{n+h}|\mathbb{X}_n\right]$ a.s.

Finding the conditional expectation is usually not possible (only for Gaussian processes, where the conditional expectation is a linear combination of $X_1, \ldots, X_n$). In the following, we restrict our

attention to linear functions of $X_1, \ldots, X_n$. In this way, we will not find the best possible solution to the minimization problem, but we will find a solution achievable in practice.

The problem of finding 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} based on $X_1, \ldots, X_n$ will be denoted by $\widehat{X}_{n+h}(n)$.

Let $H = \mathcal{H}\{X_1, \ldots, X_n, \ldots, X_{n+h}\}$ be the Hilbert space generated by the centered random variables $X_1, \ldots, X_{n+h}$ and $H_1^n = \mathcal{H}\{X_1, \ldots, X_n\}$ be the Hilbert space generated by $X_1, \ldots, X_n$. Note that H_1^n is a subspace of H.

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

$$\widehat{X}_{n+h}(n) = \sum_{k=1}^{n} c_k X_k \in H_1^n, \tag{9.3}$$

such that *the prediction error* $\mathbb{E}|X_{n+h} - \widehat{X}_{n+h}(n)|^2 = \|X_{n+h} - \widehat{X}_{n+h}(n)\|^2$ takes the minimum value with respect to all linear combinations of $X_1, \ldots, X_n$. It means that

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

and the element $\widehat{X}_{n+h}(n)$ is determined uniquely, see Theorem 9.2 (projection theorem).

The space H_1^n is linear, generated by finitely many random variables $X_1, \ldots, X_n$, meaning that

$$\begin{aligned} X_{n+h} - \widehat{X}_{n+h}(n) \perp H_1^n &\iff X_{n+h} - \widehat{X}_{n+h}(n) \perp X_j,\ j = 1, \ldots, n \\ &\iff \mathbb{E}\left(X_{n+h} - \widehat{X}_{n+h}(n)\right)\overline{X}_j = 0,\ j = 1, \ldots, n. \end{aligned}$$

The constants $c_1, \ldots, c_n$ in Equation (9.3) can be determined by solving the equations

$$\mathbb{E}\left(X_{n+h} - \sum_{k=1}^{n} c_k X_k\right)\overline{X}_j = 0,\ j = 1, \ldots, n.$$

If $X_1, \ldots, X_{n+h}$ form a real-valued, centered, weakly stationary sequence with the autocovariance function R, the previous system of equations can be written as

$$\sum_{k=1}^{n} c_k R(k-j) = R(n+h-j),\ j = 1, \ldots, n,$$

or more explicitly as

$$\begin{aligned} c_1 R(0) + c_2 R(1) + \ldots + c_n R(n-1) &= R(n+h-1), \\ c_1 R(1) + c_2 R(0) + \ldots + c_n R(n-2) &= R(n+h-2), \\ &\ldots \\ c_1 R(n-1) + c_2 R(n-2) + \ldots + c_n R(0) &= R(h). \end{aligned}$$

Note that under the assumption of real-valued sequence, the autocovariance function is symmetric, i.e. $R(-t) = R(t), t \in \mathbb{Z}$. Furthermore, the system of equations can be written in the matrix form

$$\Gamma_n \mathbf{c}_n = \gamma_{nh},$$

where $\mathbf{c}_n = (c_1, \ldots, c_n)^T$, $\gamma_{nh} = (R(n+h-1), \ldots, R(h))^T$ and

$$\Gamma_n = \begin{pmatrix} R(0) & R(1) & \cdots & R(n-1) \\ R(1) & R(0) & \cdots & R(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(n-1) & R(n-2) & \cdots & R(0) \end{pmatrix}.$$

If the inverse matrix Γ_n^{-1} exists, we get $\mathbf{c}_n = \Gamma_n^{-1}\gamma_{nh}$. Denoting $\mathbf{X}_n = (X_1, \ldots, X_n)^T$, we can further write

$$\widehat{X}_{n+h}(n) = \sum_{k=1}^{n} c_k X_k = \mathbf{c}_n^T \mathbf{X}_n = \gamma_{nh}^T \Gamma_n^{-1} \mathbf{X}_n.$$

Clearly, Γ_n is the variance matrix of $(X_1, \ldots, X_n)^T$, i.e. $\Gamma_n = \operatorname{var} \mathbf{X}_n = \mathbb{E}\mathbf{X}_n \mathbf{X}_n^T$.

The prediction error is

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

From the Theorem 9.2 (projection theorem) we get

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

and hence

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

For a real-valued, centered, weakly stationary sequence such that Γ_n is regular (meaning that Γ_n^{-1} exists), we have:

$$\begin{aligned} \delta_h^2 &= \|X_{n+h}\|^2 - \left\|\widehat{X}_{n+h}(n)\right\|^2 = \mathbb{E}\,|X_{n+h}|^2 - \mathbb{E}\left|\widehat{X}_{n+h}(n)\right|^2 \\ &= R(0) - \mathbb{E}\left(\mathbf{c}_n^T \mathbf{X}_n\right)^2 = R(0) - \mathbf{c}_n^T \mathbb{E}\left(\mathbf{X}_n \mathbf{X}_n^T\right) \mathbf{c}_n \\ &= R(0) - \mathbf{c}_n^T \Gamma_n \mathbf{c}_n = R(0) - \gamma_{nh}^T \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_{nh} \\ &= R(0) - \gamma_{nh}^T \Gamma_n^{-1} \gamma_{nh}. \end{aligned}$$

Theorem 9.4. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, centered, weakly stationary sequence with the autocovariance function R such that $R(0) > 0$ and $R(k) \to 0, k \to \infty$. Then the matrix $\Gamma_n = \operatorname{var}(X_1, \ldots, X_n)$ is regular for every $n \in \mathbb{N}$.*

Proof. We prove the claim by contradiction. We suppose that Γ_n is singular for some $n \in \mathbb{N}$. Then there is a non-zero vector $\mathbf{c} = (c_1, \ldots, c_n)^T$ such that $\mathbf{c}^T \Gamma_n \mathbf{c} = 0$. Denoting $\mathbf{X}_n = (X_1, \ldots, X_n)^T$, it follows that $\mathbb{E}\mathbf{c}^T \mathbf{X}_n = 0$ (as the sequence is centered) and $\operatorname{var}(\mathbf{c}^T \mathbf{X}_n) = \mathbf{c}^T \Gamma_n \mathbf{c} = 0$. This means that $\mathbf{c}^T \mathbf{X}_n = 0$ a.s.

Since $\Gamma_1 = R(0) > 0$ is regular and Γ_n is singular, there is an integer $1 \le r < n$ such that Γ_r is regular and Γ_{r+1} is singular. Using the same arguments as above we see that there are constants $a_1, \ldots a_r$ such that $X_{r+1} = \sum_{j=1}^r a_j X_j$ a.s.

From the weak stationarity of $\{X_t, t \in \mathbb{Z}\}$ we have $\operatorname{var}(X_1, \ldots, X_{r+1}) = \ldots = \operatorname{var}(X_h, \ldots, X_{r+h}) = \Gamma_{r+1}$ and thus for each $h \ge 1$ it holds that $X_{r+h} = \sum_{j=1}^r a_j X_{j+h-1}$ a.s. By repeatedly plugging-in this formula, we see that for each $n \ge r+1$ there are constants $a_1^{(n)}, \ldots, a_r^{(n)}$ such that $X_n = \sum_{j=1}^r a_j^{(n)} X_j = \mathbf{a}^{(n)} \mathbf{X}_r$ a.s., where $\mathbf{a}^{(n)} = (a_1^{(n)}, \ldots, a_r^{(n)})^T$ and $\mathbf{X}_r = (X_1, \ldots, X_r)^T$. It follows that

$$0 < R(0) = \operatorname{var} X_n = \mathbf{a}^{(n)T} \operatorname{var} \mathbf{X}_r \, \mathbf{a}^{(n)} = \mathbf{a}^{(n)T} \Gamma_r \, \mathbf{a}^{(n)}.$$

The matrix Γ_r is positive definite (it is a variance matrix and it is regular), implying that there is a decomposition $\Gamma_r = P\Lambda P^T$, where Λ is a diagonal matrix with the eigenvalues of the matrix Γ_r on the diagonal and the product $PP^T = I$ is the identity matrix.

Since Γ_r is positive definite, all its eigenvalues are positive. Without loss of generality, we assume the eigenvalues are $0 < \lambda_1 \leq \ldots \leq \lambda_r$. Then

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

from which it follows for each $j = 1, \ldots, r$ that $\left(\mathbf{a}_j^{(n)}\right)^2 \leq R(0)/\lambda_1$. Hence $|a_j^{(n)}| \leq C$ independently of n, where C is a positive constant. We also have

$$0 < R(0) = |R(0)| = |\mathbb{E}X_n^2| = \left|\mathbb{E}X_n \left(\sum_{j=1}^{r} a_j^{(n)} X_j\right)\right| = \left|\sum_{j=1}^{r} a_j^{(n)} \mathbb{E}X_n X_j\right| = \left|\sum_{j=1}^{r} a_j^{(n)} R(n-j)\right|$$
$$\leq \sum_{j=1}^{r} \left|a_j^{(n)}\right| \cdot |R(n-j)| \leq C \sum_{j=1}^{r} |R(n-j)| \to 0,\ n \to \infty,$$

since we assumed $R(n) \to 0$, $n \to \infty$. However, this contradicts the assumption that $R(0) > 0$. We conclude that Γ_n is regular for each $n \in \mathbb{N}$. □

Remark: Let $\{X_t, t \in \mathbb{Z}\}$ be a weakly stationary sequence with mean value μ. Then

$$P_{\widetilde{H}} X_{n+h} = \mu + P_H(X_{n+h} - \mu),$$

where $\widetilde{H} = \mathcal{H}\{1, X_1, \ldots, X_n\}$, $H = \mathcal{H}\{X_1 - \mu, \ldots, X_n - \mu\}$, and 1 is a random variable equal to the number 1 almost surely. This means that without loss of generality, we may consider only centered random variables in our prediction methodology.

9.3 Prediction based on infinite history

Suppose we know the history $X_n, X_{n-1}, \ldots$, and we want to forecast (predict) $X_{n+1}, X_{n+2}, \ldots$ Again, we solve this task by using projections in Hilbert spaces.

Consider the Hilbert spaces $H = \mathcal{H}\{X_t, t \in \mathbb{Z}\}$ and $H_{-\infty}^n = \mathcal{H}\{X_n, X_{n-1}, \ldots\}$. For $h \in \mathbb{N}$, the best linear prediction $\widehat{X}_{n+h}(n)$ of X_{n+h}, based on the infinite history $X_n, X_{n-1}, \ldots$, is the projection of $X_{n+h} \in H$ onto $H_{-\infty}^n$, i.e.

$$\widehat{X}_{n+h}(n) = P_{H_{-\infty}^n} X_{n+h}.$$

For simplicity, we will denote the one-step prediction $\widehat{X}_{n+1}(n) = \widehat{X}_{n+1}$.

Causal AR(p) models

Consider the model

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + Y_t,\ t \in \mathbb{Z},$$

where $\{Y_t, t \in \mathbb{Z}\}$ is a WN(0,σ^2) sequence and all the roots of the polynomial $a(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p, z \in \mathbb{C}$, lie outside of the unit circle. It follows that $\{X_t, t \in \mathbb{Z}\}$ is a causal linear process and $Y_t \perp X_s$ for each $t, s \in \mathbb{Z}$ such that $t > s$.

One-step prediction: We want to construct the prediction $\widehat{X}_{n+1} = \widehat{X}_{n+1}(n)$ using the random variables $X_n, X_{n-1}, \ldots$ We note that

- $X_{n+1} = \varphi_1 X_n + \ldots + \varphi_p X_{n+1-p} + Y_{n+1}$,
- $\varphi_1 X_n + \ldots + \varphi_p X_{n+1-h} \in H^n_{-\infty}$,
- $Y_{n+1} \perp X_n, X_{n-1}, \ldots$, meaning that $Y_{n+1} \perp H^n_{-\infty}$ (using the linearity and continuity of the inner product).

This means that

$$\widehat{X}_{n+1} = P_{H^n_{-\infty}} X_{n+1} = \varphi_1 X_n + \ldots + \varphi_p X_{n+1-p},$$

since $X_{n+1} - \widehat{X}_{n+1} = Y_{n+1} \perp H^n_{-\infty}$.

The prediction error is $\mathbb{E}\left|X_{n+1} - \widehat{X}_{n+1}\right|^2 = \mathbb{E}|Y_{n+1}|^2 = \sigma^2$.

***h*-step prediction with $h > 1$:** Using Theorem 9.2 (projection theorem) and the formula for the one-step prediction, we see that

$$\begin{aligned}\widehat{X}_{n+h}(n) &= P_{H^n_{-\infty}} X_{n+h} = P_{H^n_{-\infty}} \left(P_{H^{n+h-1}_{-\infty}} X_{n+h}\right) = P_{H^n_{-\infty}} \widehat{X}_{n+h}(n+h-1) \\ &= P_{H^n_{-\infty}} \left(\varphi_1 X_{n+h-1} + \ldots + \varphi_p X_{n+h-p}\right) = \varphi_1 \left[X_{n+h-1}\right] + \varphi_2 \left[X_{n+h-2}\right] + \ldots + \varphi_p \left[X_{n+h-p}\right],\end{aligned}$$

where for $j \in \mathbb{Z}$ we denote

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

The prediction error can be determined as

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

see Theorem 9.2 (the projection theorem). Note that the prediction error can be expressed as a linear combination of the values of the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$.

Remark: Note that for AR(p) models, only the values $X_n, \ldots, X_{n+1-p}$ are needed for finding $\widehat{X}_{n+h}(n)$ for $h \in \mathbb{N}$. Only the last p random variables are used, and hence the prediction is in fact *not* based on infinite history, since the older random variables do not bring relevant information.

Example: Consider an AR(1) model given by $X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, with $|\varphi| < 1$ and $\{Y_t, t \in Z\}$ being a white noise sequence WN(0,σ^2). Even if we know the whole history $X_n, X_{n-1}, \ldots$, the best linear prediction of X_{n+1} is $\widehat{X}_{n+1} = \varphi X_n$. For $h > 1$ we have:

$$\widehat{X}_{n+h}(n) = \varphi[X_{n+h-1}] = \varphi \widehat{X}_{n+h-1}(n) = \varphi^2 \widehat{X}_{n+h-2}(n) = \ldots = \varphi^h X_n.$$

The prediction error is

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

Causal and invertible ARMA(p,q) models

Consider the model

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

where $\{Y_t, t \in \mathbb{Z}\}$ is a WN$(0,\sigma^2)$ sequence. Assume that $\{X_t, t \in \mathbb{Z}\}$ is causal and invertible.
Causality implies that

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

where $\sum_{j=0}^{\infty} |c_j| < \infty$. It follows that $Y_t \perp X_s$ for each $t, s \in \mathbb{Z}$ such that $t > s$.
Furthermore, invertibility implies that

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

where $\sum_{j=0}^{\infty} |d_j| < \infty$ and $d_0 = 1$. It follows that $X_t = -\sum_{j=1}^{\infty} d_j X_{t-j} + Y_t$, and specifically

$$X_{n+1} = -\sum_{j=1}^{\infty} d_j X_{n+1-j} + Y_{n+1}. \tag{9.4}$$

One-step prediction: We notice that

$$-\sum_{j=1}^{\infty} d_j X_{n+1-j} = \text{l.i.m.}_{N\to\infty} \left(-\sum_{j=1}^{N} d_j X_{n+1-j} \right) \in H^n_{-\infty},$$

and that $Y_{n+1} \perp H^n_{-\infty}$ (due to causality). It follows from (9.4) that

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

The prediction error is $\mathbb{E}\left|X_{n+1} - \widehat{X}_{n+1}\right|^2 = \mathbb{E}|Y_{n+1}|^2 = \sigma^2$.

h-step prediction with $h > 1$: Similarly as in (9.4) we get that $X_{n+h} = -\sum_{j=1}^{\infty} d_j X_{n+h-j} + Y_{n+h}$. Furthermore, causality implies that $Y_{n+h} \perp H^n_{-\infty}$. Combining these properties, it follows that

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

where the convergence of the sum on the right-hand side is secured by $\sum_{j=0}^{\infty} |d_j| < \infty$ and $[X_{n+h-j}]$ are defined as above.

The prediction error can be determined using causality. From the one-step prediction, we know that $X_t - \widehat{X}_t = Y_t, t \in \mathbb{Z}$. We can write

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

Since $P_{H^n_{-\infty}} Y_{n+h-j} = 0$ for $j < h$, we get

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

The prediction error is then

$$\mathbb{E} \left| X_{n+h} - \widehat{X}_{n+h}(n) \right|^2 = \mathbb{E} \left| \sum_{j=0}^{h-1} c_j Y_{n+h-j} \right|^2 = \sigma^2 \sum_{j=0}^{h-1} |c_j|^2.$$

Note that the previous formula for the prediction error holds also for $h = 1$.

Remark: What happens if we truncate the prediction based on infinite history because, in fact, we do not have infinitely many observations in practice? Instead of $\widehat{X}_{n+h}(n) = \sum_{k=0}^{\infty} a_k X_{n-k}$, we use $\widetilde{X}_{n+h}(n) = \sum_{k=0}^{N} a_k X_{n-k}$. Its prediction error $\mathbb{E} \left| X_{n+h} - \widetilde{X}_{n+h}(n) \right|^2$ can be expressed using the values of the autocovariance function of $\{X_t, t \in \mathbb{Z}\}$ and by choosing N, we can control how much worse this truncated prediction $\widetilde{X}_{n+h}(n)$ is compared to the truly infinite prediction $\widehat{X}_{n+h}(n)$.

Example: Consider an MA(1) = ARMA(0,1) model given by $X_t = Y_t + \theta Y_{t-1}, t \in \mathbb{Z}$, where $|\theta| < 1$ and $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). It follows that $\{X_t, t \in \mathbb{Z}\}$ is invertible and $Y_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}, t \in \mathbb{Z}$ (this is equivalent to finding MA(∞) representation of a causal AR(1) model). Since $X_{n+1} = \theta Y_n + Y_{n+1}$, $Y_n \in H^n_{-\infty}$ (from invertibility) and $Y_{n+1} \perp H^n_{-\infty}$ (from causality), we have

$$\widehat{X}_{n+1} = \theta Y_n = \theta \sum_{j=0}^{\infty} (-\theta)^j X_{n-j}.$$

The prediction error is $\mathbb{E} \left| X_{n+1} - \widehat{X}_{n+1} \right|^2 = \mathbb{E}|Y_{n+1}|^2 = \sigma^2$.

For $h > 1$ we have $\widehat{X}_{n+h}(n) = P_{H^n_{-\infty}} \left(P_{H^{n+h-1}_{-\infty}} X_{n+h} \right) = P_{H^n_{-\infty}} \widehat{X}_{n+h} = \theta \cdot P_{H^n_{-\infty}} Y_{n+h-1} = 0$, since $n+h-1 > n$, and hence $Y_{n+h-1} \perp H^n_{-\infty}$. The prediction error is $\mathbb{E} \left| X_{n+h} - \widehat{X}_{n+h}(n) \right|^2 = \mathbb{E}|X_{n+h}|^2 = R(0) = \sigma^2(1 + \theta^2)$. Note that the prediction error is higher for $h > 1$ than for $h = 1$. This is not surprising, since prediction to a more distant future is inherently a more complicated task.

9.4 Filtration of signal and noise

Consider a sequence $\{X_t, t \in \mathbb{Z}\}$ which contains the signal and a sequence $\{Y_t, t \in \mathbb{Z}\}$ which constitutes the noise (not necessarily a white noise sequence). Our observations form the sequence $\{V_t, t \in \mathbb{Z}\}$ which we assume to be the mixture of the signal and noise,

$$V_t = X_t + Y_t, \ t \in \mathbb{Z}.$$

Our goal is now to separate the signal from the noise. We assume that $\{X_t, t \in \mathbb{Z}\}$ and $\{Y_t, t \in \mathbb{Z}\}$ are real-valued, centered, weakly stationary random sequences, with autocovariance functions R_X and R_Y, respectively. We further assume that these two sequences are uncorrelated. It follows that $\{V_t, t \in \mathbb{Z}\}$ is a real-valued, centered, weakly stationary random sequence with the autocovariance function $R_V = R_X + R_Y$.

In the following, we assume that we observe the random variables $V_1, \ldots, V_n$, i.e. we have finitely many observations available. We are looking for a linear estimator of X_s in the form $\widehat{X}_s = \sum_{j=1}^{n} c_j V_j$,

with the coefficients $c_1, \dots, c_n$ minimizing the mean square error $\mathbb{E}|X_s - \widehat{X}_s|^2$. Again, we are looking for a projection in a Hilbert space.

Denote $H_1^n = \mathcal{H}\{V_1, \dots, V_n\} \subset L_2(\Omega, \mathcal{A}, \mathbb{P})$ and fix $s \in \mathbb{Z}$. For $s = n$ we call the following procedure simply *filtration*. For $s < n$ we call it *filtration with delay* and for $s > n$ *filtration and prediction*. The best linear approximation $\widehat{X}_s$ of X_s is the projection of X_s onto H_1^n, i.e. $\widehat{X}_s \in H_1^n$ and $X_s - \widehat{X}_s \perp H_1^n$. Since we have finitely many observations, $H_1^n = \mathcal{H}\{V_1, \dots, V_n\} = \mathcal{M}\{V_1, \dots, V_n\}$, and it is enough to find $c_1, \dots, c_n$ such that

$$\widehat{X}_s = \sum_{j=1}^{n} c_j V_j,$$

and

$$X_s - \widehat{X}_s \perp V_t,\ t = 1, \dots, n,$$

or equivalently

$$\mathbb{E}\left(X_s - \widehat{X}_s\right) V_t = 0,\ t = 1, \dots, n.$$

Since $V_t = X_t + Y_t$ and X_t, Y_t are uncorrelated for each $t \in \mathbb{Z}$, we have

$$\mathbb{E}X_s V_t = \mathbb{E}X_s X_t = R_X(s-t),\ s, t \in \mathbb{Z},$$

and we can write

$$\mathbb{E}\left(X_s - \sum_{j=1}^{n} c_j V_j\right) V_t = R_X(s-t) - \sum_{j=1}^{n} c_j R_V(j-t) = 0,\ t = 1, \dots, n.$$

Now it remains to solve the system of n linear equations for n unknown values $c_1, \dots, c_n$. This system of equations can be written in a matrix form. For the sufficient conditions for the regularity of the matrix with the entries $R_V(j-t)$, see Theorem 9.4.

The random variable $\widehat{X}_s$ is *the best linear filtration* of the signal X_s at time $s \in \mathbb{Z}$ from the mixture $V_1, \dots, V_n$. The filtration error is

$$\begin{aligned}\mathbb{E}\left|X_s - \widehat{X}_s\right|^2 = \left\|X_s - \widehat{X}_s\right\|^2 = \|X_s\|^2 - \left\|\widehat{X}_s\right\|^2 = R_X(0) - \mathbb{E}\left|\sum_{j=1}^{n} c_j V_j\right|^2 \\ = R_X(0) - \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k R_V(j-k).\end{aligned}$$

Remark: It is not unreasonable to assume that we know R_X, R_Y and R_V or at least their estimates. Properties of the noise $\{Y_t, t \in \mathbb{Z}\}$ can be studied in controlled experiments with zero signal and R_V can be estimated from the observations, meaning that R_X can be estimated using the formula $R_X = R_V - R_Y$. Alternatively, properties of the assumed signal (such as human speech) can be studied in repeated experiments where averaging reduces the noise component.

10 Partial autocorrelation function

We assume real-valued sequences in the following, as we will take advantage of the symmetry of the autocovariance function.

Definition 10.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, centered, weakly stationary sequence.* The partial autocorrelation function *of $\{X_t, t \in \mathbb{Z}\}$ is defined as*

$$\alpha(k) = \begin{cases} r(1) = \operatorname{corr}(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{\operatorname{var} X_1}\sqrt{\operatorname{var} X_2}}, & k = 1, \\ \operatorname{corr}\left(X_1 - \widetilde{X}_1, X_{k+1} - \widetilde{X}_{k+1}\right), & k \in \mathbb{N}, k > 1, \end{cases}$$

where $\widetilde{X}_1 = P_{H_2^k} X_1, \widetilde{X}_{k+1} = P_{H_2^k} X_{k+1}$ are linear projections onto the Hilbert space $H_2^k = \mathcal{H}\{X_2, \ldots, X_k\}$.

Remark: Note the connection to the partial correlation coefficient, measuring the correlation between X_1 and X_{k+1}, after removing the (linear) influence of $X_2, \ldots, X_k$.

The projection is a linear function of $X_2, \ldots, X_k$, and $\widetilde{X}_1 = c_2 X_2 + \ldots + c_k X_k$. From the projection theorem we get that $X_1 - \widetilde{X}_1 \perp H_2^k$, implying that $\mathbb{E}(X_1 - \widetilde{X}_1) X_j = 0, j = 2, \ldots, k$. The constants $c_2, \ldots, c_k$ are determined by this system of equations. The same applies to $\widetilde{X}_{k+1}$, too.

Remark: Weak stationarity of $\{X_t, t \in \mathbb{Z}\}$ implies that for $h \in \mathbb{N}$ and $k > 1$,

$$\alpha(k) = \operatorname{corr}\left(X_1 - \widetilde{X}_1, X_{k+1} - \widetilde{X}_{k+1}\right) = \operatorname{corr}\left(X_h - \widetilde{X}_h, X_{k+h} - \widetilde{X}_{k+h}\right),$$

where $\widetilde{X}_h = P_{H_{h+1}^{k+h-1}} X_h, \widetilde{X}_{k+h} = P_{H_{h+1}^{h+k-1}} X_{k+1}$ and $H_{h+1}^{h+k-1} = \mathcal{H}\{X_{h+1}, \ldots, X_{h+k-1}\}$.

Example: Consider a causal AR(1) model $X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, where $|\varphi| < 1$ and $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). For $k = 1$ we have $\alpha(1) = r(1) = \operatorname{corr}(X_1, X_2) = \varphi$. For $k > 1$, causality implies that $Y_{k+1} \perp H_2^k$. It follows that $\widetilde{X}_{k+1} = P_{H_2^k} X_{k+1} = P_{H_2^k}(\varphi X_k + Y_{k+1}) = \varphi X_k$. Furthermore, from causality we get $Y_{k+1} \perp X_1$ and also $Y_{k+1} \perp \widetilde{X}_1 \in H_2^k$. It follows that

$$0 = \mathbb{E}\left(X_1 - \widetilde{X}_1\right) Y_{k+1} = \mathbb{E}\left(X_1 - \widetilde{X}_1\right)\left(X_{k+1} - \widetilde{X}_{k+1}\right).$$

Since

$$\alpha(k) = \frac{\mathbb{E}\left(X_1 - \widetilde{X}_1\right)\left(X_{k+1} - \widetilde{X}_{k+1}\right)}{\sqrt{\mathbb{E}\left(X_1 - \widetilde{X}_1\right)^2}\sqrt{\mathbb{E}\left(X_{k+1} - \widetilde{X}_{k+1}\right)^2}},$$

we conclude that $\alpha(k) = 0$ for $k > 1$.

Remark: In the same way, we can show that for a causal AR(p) model it holds that $\alpha(k) = 0$ for $k > p$. The values of $X_k, X_{k-1}, \ldots, X_{k-p+1}$ contain all the information about X_{k+1} contained in $H_{-\infty}^k$, so X_1 does not provide additional information about X_{k+1}, corresponding to $\alpha(k) = 0$.

Example: Consider the MA(1) model $X_t = Y_t + bY_{t-1}$ with $b \in \mathbb{R}$ and $\{Y_t, t \in \mathbb{Z}\}$ being a white noise sequence WN(0,σ^2). In this case, the autocovariance function of the sequence $\{X_t, t \in \mathbb{Z}\}$ is $R(0) = (1 + b^2)\sigma^2, R(1) = R(-1) = b\sigma^2$ and $R(k) = 0$ for $|k| > 1$. It follows that $\alpha(1) = r(1) = \frac{b}{1+b^2}$ and $\alpha(2) = \operatorname{corr}\left(X_1 - \widetilde{X}_1, X_3 - \widetilde{X}_3\right)$. To find $\widetilde{X}_1 = P_{H_2^2} X_1 = cX_2$, we need to find c such that $\left(X_1 - \widetilde{X}_1\right) \perp H_2^2$, i.e. $\mathbb{E}(X_1 - cX_2)X_2 = 0$. This equation can be rewritten as $R(1) - cR(0) = 0$, and hence $c = R(1)/R(0) = r(1) = \frac{b}{1+b^2}$. Similarly, we get $\widetilde{X}_3 = \frac{b}{1+b^2} X_2$, too. Since the sequence

$\{X_t, t \in \mathbb{Z}\}$ is centered, it is now easy to compute the variances and the covariance of $X_1 - \widetilde{X}_1$ and $X_3 - \widetilde{X}_3$. Altogether,

$$\alpha(2) = -\frac{b^2}{1 + b^2 + b^4}.$$

In general,

$$\alpha(k) = -\frac{(-b)^k(1 - b^2)}{1 - b^{2(k+1)}},\ k \geq 1.$$

Remark: In a sense, the autocovariance function R (or the autocorrelation function r) and the partial autocorrelation function α are complementary tools: AR(p) sequences have infinitely many non-zero values of $R(k)$ and only finitely many non-zero values of $\alpha(k)$. For MA(n) models, the opposite holds.

Remark: Plots of the empirical autocovariance or autocorrelation function and the empirical partial autocorrelation function can be useful when choosing an appropriate model for a given dataset. For illustration, consider e.g. the yearly mean total sunspot number as reported by the SILSO center (https://www.sidc.be/SILSO/datafiles), see Figure 17. The estimated values of the autocovariance function and the partial autocorrelation function are shown in Figure 18 (the estimation will be discussed soon in Chapter 11).

While the plot of the empirical autocovariance function shows a repeating structure, the plot of the estimated partial autocorrelation function shows that the values of $\widehat{\alpha}(k)$ for k up to 9 are relevant. On the other hand, the values $\widehat{\alpha}(k), k \geq 10$, are close to zero and can be considered negligible (note that the first plotted value is $\widehat{\alpha}(1)$). It follows that a reasonable model for the sunspot time series could be an AR(9) model. However, this is only a visual procedure and formal model selection methods are outside the scope of this course.

Definition 10.2 (alternative definition of the partial autocorrelation function)**.** *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, centered, weakly stationary sequence. Let $P_{H_1^k} X_{k+1} = \varphi_1 X_k + \ldots + \varphi_k X_1$ be the best linear prediction of X_{k+1} based on $X_1, \ldots, X_k$, $H_1^k = \mathcal{H}\{X_1, \ldots, X_k\}$. The partial autocorrelation function is defined as $\alpha(k) = \varphi_k, k \in \mathbb{N}$.*

Remark: The concept is the same in Definition 10.2 as in Definition 10.1: after explaining X_{k+1} using $X_2, \ldots, X_k$ (in a linear way), how much does X_1 help to further explain X_{k+1}?

Theorem 10.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, centered, weakly stationary sequence with the autocovariance function R and autocorrelation function r. Assume that $R(0) > 0$ and $R(t) \to 0$, as $t \to \infty$. Then both definitions of the partial autocorrelation function are equivalent, and it holds that $\alpha(1) = r(1)$ and*

$$\alpha(k) = \frac{\begin{vmatrix} 1 & r(1) & \cdots & r(k-2) & r(1) \\ r(1) & 1 & \cdots & r(k-3) & r(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(k-1) & r(k-2) & \cdots & r(1) & r(k) \end{vmatrix}}{\begin{vmatrix} 1 & r(1) & \cdots & r(k-2) & r(k-1) \\ r(1) & 1 & \cdots & r(k-3) & r(k-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(k-1) & r(k-2) & \cdots & r(1) & 1 \end{vmatrix}},\ k > 1. \tag{10.1}$$

Proof. We start by showing the equivalence of the two definitions for $k = 1$. According to the first definition, $\alpha(1) = r(1)$. Considering the second definition, we write $\widehat{X}_2 = P_{H_1^1} X_2 = \varphi_1 X_1$ and $(X_2 - \widehat{X}_2) \perp X_1$, implying $\mathbb{E}(X_2 - \varphi_1 X_1) X_1 = 0$, $R(1) - \varphi_1 R(0) = 0$ and finally $\varphi_1 = r(1)$.

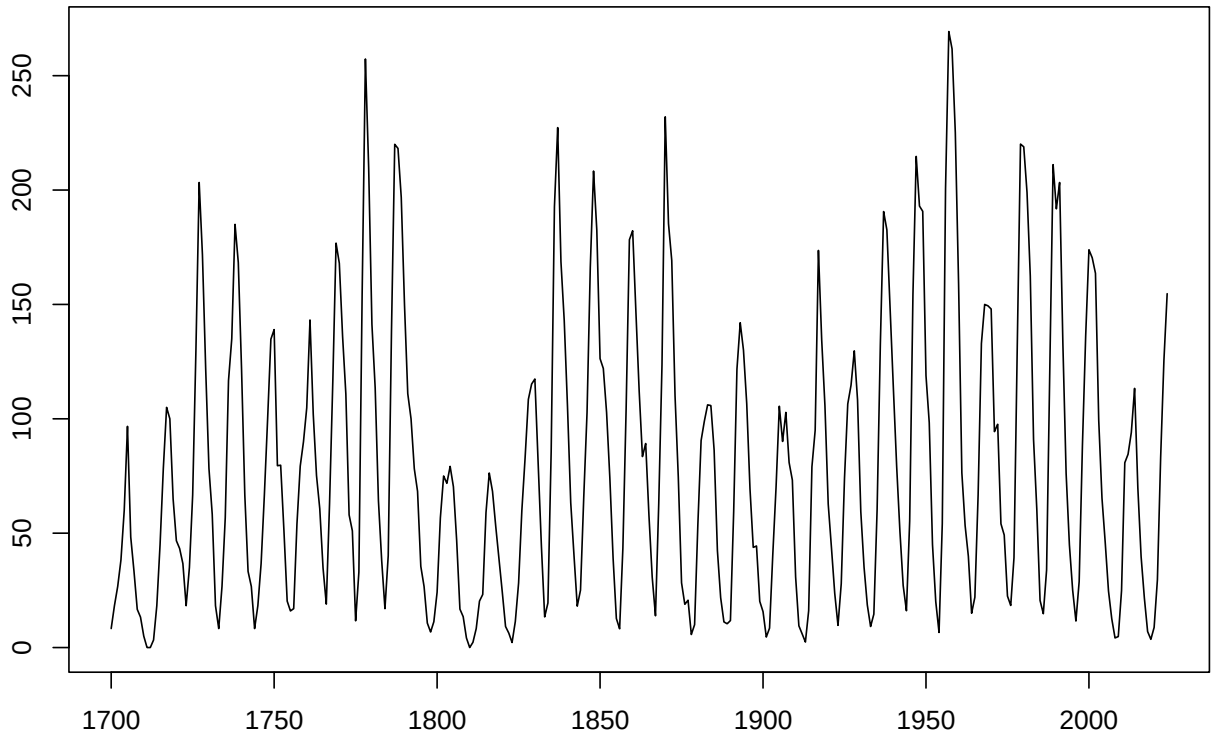

Figure 17: Yearly mean total sunspot number, 1700–2024. Note that this is a discrete-time sequence, and the lines joining the corresponding points in the plot are used only for clarity.

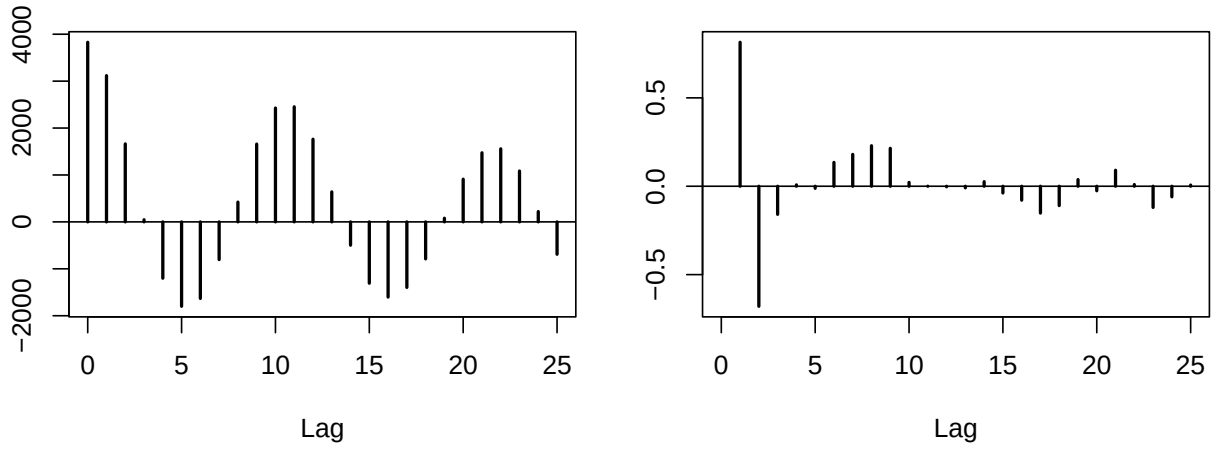

Figure 18: Estimated autocovariance function (left) and estimated partial autocorrelation function (right) for the sunspot time series from Figure 17.

We proceed by showing the equivalence for $k > 1$. We set up some useful notation:

$$H_1^k = \mathcal{H}\{X_1, \ldots, X_k\}, \qquad \widehat{X}_{k+1} = P_{H_1^k} X_{k+1},$$
$$H_2^k = \mathcal{H}\{X_2, \ldots, X_k\}, \qquad \widetilde{X}_{k+1} = P_{H_2^k} X_{k+1}, \qquad \widetilde{X}_1 = P_{H_2^k} X_1, \qquad \widetilde{H} = \mathcal{H}\{X_1 - \widetilde{X}_1\}.$$

In order to work with the second definition, consider the projection $\widehat{X}_{k+1} = \varphi_1 X_k + \ldots + \varphi_k X_1$. We divide the proof of the equivalence into several parts:

A) We show that $(X_{k+1} - \varphi_k(X_1 - \widetilde{X}_1)) \perp \widetilde{H} = \mathcal{H}\{X_1 - \widetilde{X}_1\}$. This means that

$$\begin{aligned} &\mathbb{E}(X_{k+1} - \varphi_k(X_1 - \widetilde{X}_1))(X_1 - \widetilde{X}_1) = 0, \\ &\mathbb{E}X_{k+1}(X_1 - \widetilde{X}_1) - \varphi_k \mathbb{E}(X_1 - \widetilde{X}_1)^2 = 0, \\ &\varphi_k = \frac{\mathbb{E}X_{k+1}(X_1 - \widetilde{X}_1)}{\mathbb{E}(X_1 - \widetilde{X}_1)^2}. \end{aligned} \tag{10.2}$$

B) We show that $(X_1 - \widetilde{X}_1) \perp \widetilde{X}_{k+1}$. It follows that $\mathbb{E}\widetilde{X}_{k+1}(X_1 - \widetilde{X}_1) = 0$ and we may subtract this from the numerator of (10.2) to get

$$\varphi_k = \frac{\mathbb{E}(X_{k+1} - \widetilde{X}_{k+1})(X_1 - \widetilde{X}_1)}{\mathbb{E}(X_1 - \widetilde{X}_1)^2}.$$

C) We argue that $\mathbb{E}(X_1 - \widetilde{X}_1)^2 = \mathbb{E}(X_{k+1} - \widetilde{X}_{k+1})^2$. It follows that

$$\varphi_k = \frac{\mathbb{E}(X_{k+1} - \widetilde{X}_{k+1})(X_1 - \widetilde{X}_1)}{\sqrt{\mathbb{E}(X_1 - \widetilde{X}_1)^2}\sqrt{\mathbb{E}(X_{k+1} - \widetilde{X}_{k+1})^2}} = \operatorname{corr}\left(X_1 - \widetilde{X}_1, X_{k+1} - \widetilde{X}_{k+1}\right) = \alpha(k).$$

Ad A) since $X_1 = \widetilde{X}_1 + (X_1 - \widetilde{X}_1)$, where $\widetilde{X}_1 \in H_2^k$ and $(X_1 - \widetilde{X}_1) \perp H_2^k$, it holds that

$$\widehat{X}_{k+1} = \varphi_1 X_k + \ldots + \varphi_k X_1 = \left[\varphi_1 X_k + \ldots + \varphi_{k-1} X_2 + \varphi_k \widetilde{X}_1\right] + \left[\varphi_k(X_1 - \widetilde{X}_1)\right],$$

where the random variables in the square brackets are orthogonal since $\varphi_1 X_k + \ldots + \varphi_{k-1} X_2 + \varphi_k \widetilde{X}_1 \in H_2^k$ and $\varphi_k(X_1 - \widetilde{X}_1) \perp H_2^k$. It follows that $\varphi_k(X_1 - \widetilde{X}_1) = P_{\widetilde{H}} \widehat{X}_{k+1}$.

We note that $\widetilde{H} \subseteq H_1^k$, since $X_1 \in H_1^k$, $\widetilde{X}_1 \in H_2^k \subseteq H_1^k$ and hence $X_1 - \widetilde{X}_1 \in H_1^k$. From Theorem 9.3 (properties of the projection mapping) we get

$$P_{\widetilde{H}} X_{k+1} = P_{\widetilde{H}}\left(P_{H_1^k} X_{k+1}\right) = P_{\widetilde{H}} \widehat{X}_{k+1} = \varphi_k(X_1 - \widetilde{X}_1).$$

This means that

$$X_{k+1} - P_{\widetilde{H}} X_{k+1} = X_{k+1} - \varphi_k(X_1 - \widetilde{X}_1) \perp \widetilde{H}.$$

Ad B) Again, we consider $X_1 = \widetilde{X}_1 + (X_1 - \widetilde{X}_1)$, where $\widetilde{X}_1 \in H_2^k$ and $(X_1 - \widetilde{X}_1) \perp H_2^k$. Since $\widetilde{X}_{k+1} \in H_2^k$ by definition, we see that $(X_1 - \widetilde{X}_1) \perp \widetilde{X}_{k+1}$.

Ad C) Consider the system of equations we need to solve in order to find the coefficients of the prediction $\widetilde{X}_1$ and compare it to the equations needed for the prediction $\widetilde{X}_{k+1}$. They are the same equations where the corresponding coefficients occur at symmetric places. Furthermore, the values of $\mathbb{E}(X_1 - \widetilde{X}_1)^2$ and $\mathbb{E}(X_{k+1} - \widetilde{X}_{k+1})^2$ depend on these coefficients and the values of the autocovariance function R in the same way. Also, note that for a real-valued, weakly stationary random sequence, $\operatorname{var}(X_2, \ldots, X_k) = \operatorname{var}(X_k, \ldots, X_2)$.

Finally, we prove the formula (10.1). We know that $\widehat{X}_{k+1} = \varphi_1 X_k + \ldots + \varphi_k X_1 \in H_1^k$, $X_{k+1} - \widehat{X}_{k+1} \perp H_1^k$, and therefore

$$\mathbb{E}(X_{k+1} - (\varphi_1 X_k + \varphi_2 X_{k-1} + \ldots + \varphi_k X_1))X_{k+1-j} = 0,\ j = 1, 2, \ldots, k,$$

which gives

$$\begin{aligned}
&R(1) - \varphi_1 R(0) - \ldots - \varphi_k R(k-1) = 0,\\
&R(2) - \varphi_1 R(1) - \ldots - \varphi_k R(k-2) = 0,\\
&\ldots\\
&R(k) - \varphi_1 R(k-1) - \ldots - \varphi_k R(0) = 0.
\end{aligned}$$

We divide both sides by $R(0) > 0$ and write the system of equations in the matrix form:

$$\begin{pmatrix} 1 & r(1) & \cdots & r(k-1) \\ r(1) & 1 & \cdots & r(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(k-1) & r(k-2) & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_k \end{pmatrix} = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(k) \end{pmatrix}.$$

The matrix on the left-hand side is $\Gamma_k / R(0)$ and it is regular because we assumed $R(t) \to 0, t \to \infty$, see Theorem 9.4. Now, the Cramér's rule from linear algebra gives the formula for φ_k. □

Example: Consider again a causal AR(1) model $X_t = \varphi X_{t-1} + Y_t, t \in \mathbb{Z}$, where $|\varphi| < 1$ and $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). The partial autocorrelataion function can be computed, for $k > 1$, as

$$\alpha(k) = \frac{\begin{vmatrix} 1 & \varphi & \cdots & \varphi^{k-2} & \varphi \\ \varphi & 1 & \cdots & \varphi^{k-3} & \varphi^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{k-1} & \varphi^{k-2} & \cdots & \varphi & \varphi^k \end{vmatrix}}{\begin{vmatrix} 1 & \varphi & \cdots & \varphi^{k-2} & \varphi^{k-1} \\ \varphi & 1 & \cdots & \varphi^{k-3} & \varphi^{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{k-1} & \varphi^{k-2} & \cdots & \varphi & 1 \end{vmatrix}}.$$

We see that the last column of the determinant in the numerator is a multiple of the first column. It follows that the determinant is 0 and $\alpha(k) = 0$ for $k > 1$.

11 Estimation of moment properties

11.1 Mean

Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, weakly stationary random sequence with expectation $\mu = \mathbb{E}X_t$, $t \in \mathbb{Z}$, and the autocovariance function $R(s,t) = R(s-t), s,t \in \mathbb{Z}$. Assuming we observe the random variables $X_1, \ldots, X_n$, the natural estimator of μ is the sample mean:

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t.$$

This is an unbiased estimator of μ: $\mathbb{E}\overline{X}_n = \mu$.

Assuming the sequence $\{X_t, t \in \mathbb{Z}\}$ is mean square ergodic, it holds that $\overline{X}_n \to \mu, n \to \infty$, in the mean square and in probability, implying weak consistency of the estimator.

The variance of $\overline{X}_n$ in a weakly stationary random sequence is given by

$$\operatorname{var}\overline{X}_n = \frac{1}{n}\sum_{k=-n+1}^{n-1} R(k)\left(1 - \frac{|k|}{n}\right),$$

see the proof of Theorem 8.2. Additionally, if $\sum_{k=-\infty}^{\infty}|R(k)| < \infty$, it holds that $n\operatorname{var}\overline{X}_n \to \sum_{k=-\infty}^{\infty} R(k) = 2\pi f(0), n \to \infty$, where f is the spectral density of the sequence $\{X_t, t \in \mathbb{Z}\}$.

On the other hand, $\overline{X}_n$ is not the best linear unbiased estimator of μ. Recall that for a linear model $\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbb{E}\varepsilon_i = 0, i = 1, \ldots, n$, $\operatorname{var}\boldsymbol{\varepsilon} = \boldsymbol{\Gamma}$, the best linear unbiased estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = \left(\mathbf{F}^T\boldsymbol{\Gamma}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}$, with $\mathbb{E}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}, \operatorname{var}\widehat{\boldsymbol{\beta}} = \left(\mathbf{F}^T\boldsymbol{\Gamma}^{-1}\mathbf{F}\right)^{-1}$. In our case, $X_t = \mu + \widetilde{X}_t, t \in \mathbb{Z}$, $\mathbb{E}\widetilde{X}_t = 0$, $\mathbf{X}_n = (X_1, \ldots, X_n)^T$, $\widetilde{\mathbf{X}}_n = (\widetilde{X}_1, \ldots, \widetilde{X}_n)^T$, and

$$\operatorname{var}\widetilde{\mathbf{X}}_n = \operatorname{var}\mathbf{X}_n = \Gamma_n = \begin{pmatrix} R(0) & R(1) & \cdots & R(n-1) \\ R(1) & R(0) & \cdots & R(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(n-1) & R(n-2) & \cdots & R(0) \end{pmatrix}.$$

The correspondence with the linear model is the following: $\mathbf{Y} = \mathbf{X}_n$, $\mathbf{F} = (1, \ldots, 1)^T = \mathbf{1}_n$, $\boldsymbol{\beta} = \mu$, $\boldsymbol{\varepsilon} = \widetilde{\mathbf{X}}_n$. Assuming $R(0) > 0$ and $R(t) \to 0, t \to \infty$, Theorem 9.4 gives the regularity of the matrix Γ_n. The best linear unbiased estimator of μ is then

$$\widehat{\mu}_n = \left(\mathbf{1}_n^T\Gamma_n^{-1}\mathbf{1}_n\right)^{-1}\mathbf{1}_n^T\Gamma_n^{-1}\mathbf{X}_n. \tag{11.1}$$

The variance of $\widehat{\mu}_n$ is $\operatorname{var}\widehat{\mu}_n = \left(\mathbf{1}_n^T\Gamma_n^{-1}\mathbf{1}_n\right)^{-1}$.

11.2 Autocovariance and autocorrelation function

The estimator $\widehat{\mu}_n$ from Equation (11.1) assumes that the values of the autocovariance function R are known. The same holds for prediction based on finite history etc.

Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, weakly stationary random sequence, and let $X_1, \ldots, X_n$ be the available observations. The sample autocovariance function is given by

$$\widehat{R}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right), \; k = 0, 1, \ldots, n-1, \tag{11.2}$$

and we define $\widehat{R}(k) = \widehat{R}(-k)$ for $k < 0$ (this is where the assumption of real-valued sequence plays a role). Note that the sample autocovariance function is not an unbiased estimator of the autocovariance function, as $\mathbb{E}\widehat{R}(k) \neq R(k)$. On the other hand, it is asymptotically unbiased under certain assumptions, see Theorem 7.3.4 in Brockwell and Davis (2006).

The matrix

$$\widehat{\Gamma}_n = \begin{pmatrix} \widehat{R}(0) & \widehat{R}(1) & \cdots & \widehat{R}(n-1) \\ \widehat{R}(1) & \widehat{R}(0) & \cdots & \widehat{R}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{R}(n-1) & \widehat{R}(n-2) & \cdots & \widehat{R}(0) \end{pmatrix}$$

is positive semidefinite for each $n \in \mathbb{N}$. To see this, we write $\widehat{\Gamma}_n = \frac{1}{n} U \cdot U^T$, where U is the $n \times 2n$ matrix

$$U = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & Y_1 & Y_2 & \dots & Y_{n-1} & Y_n \\ 0 & 0 & 0 & \dots & Y_1 & Y_2 & Y_3 & \dots & Y_n & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & Y_1 & Y_2 & \dots & Y_{n-1} & Y_n & 0 & \dots & 0 & 0 \end{pmatrix},$$

and $Y_i = X_i - \overline{X}_n, i = 1, \dots, n$. Then for any vector $\mathbf{a}$ of length n we have

$$\mathbf{a}^T \widehat{\Gamma}_n \mathbf{a} = \frac{1}{n} \left(\mathbf{a}^T U\right) \left(\mathbf{a}^T U\right)^T \geq 0.$$

The factor $\frac{1}{n}$ in (11.2) is sometimes replaced by $\frac{1}{n-k}$ but then the matrix $\widehat{\Gamma}_n$ may not be positive semidefinite.

Also note that, for a given $n \in \mathbb{N}$, $\widehat{\Gamma}_n$ is regular if $\widehat{R}(0) > 0$. Indeed, for given values of $X_1, \dots, X_n$, the function

$$\widehat{R}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-|k|} \left(X_t - \overline{X}_n\right) \left(X_{t+|k|} - \overline{X}_n\right), & |k| < n, \\ 0, & |k| \geq n, \end{cases}$$

can be viewed as the autocovariance function of the MA$(n-1)$ sequence with coefficients $b_j = X_{1+j} - \overline{X}_n, j = 1, \dots, n-1$, and $\sigma^2 = \frac{1}{n}$, see Theorem 7.1. Then clearly $\widehat{R}(k) \to 0, k \to \infty$, and assuming $\widehat{R}(k) > 0$, regularity of $\widehat{\Gamma}_n$ follows from Theorem 9.4. This shows that the sample autocovariance function is a relevant estimator, even though it is biased.

From the n observations $X_1, \dots, X_n$ we are able to estimate $R(k)$ for $k = 0, \dots, n-1$. However, the estimated values may not be reliable. Usually, it is recommended to choose $n \geq 50$ and $k \leq \frac{n}{4}$.

Recall that the autocorrelation function r is defined as $r(k) = R(k)/R(0), k \in \mathbb{Z}$. The sample autocorrelation function is given by

$$\widehat{r}(k) = \frac{\widehat{R}(k)}{\widehat{R}(0)} = \frac{\sum_{t=1}^{n-k} \left(X_t - \overline{X}_n\right) \left(X_{t+k} - \overline{X}_n\right)}{\sum_{t=1}^{n} \left(X_t - \overline{X}_n\right)^2}, \; k = 0, 1, \dots, n-1,$$

provided that $\widehat{R}(0) = \frac{1}{n} \sum_{t=1}^{n} \left(X_t - \overline{X}_n\right)^2 > 0$. The asymptotic behavior of the sample autocorrelation function is discussed in the following theorem.

Theorem 11.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a random sequence fulfilling*

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \alpha_j Y_{t-j}, \; t \in \mathbb{Z},$$

where $\{Y_t, t \in \mathbb{Z}\}$ *are independent, identically distributed random variables with zero mean and finite positive variance* σ^2*. Also, let* $\mathbb{E}|Y_t|^4 < \infty$ *and* $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$*. Let* $r(k), k \in \mathbb{Z}$*, be the autocorrelation function of* $\{X_t, t \in \mathbb{Z}\}$ *and* $\widehat{r}_n(k)$ *be the sample autocorrelations based on* $X_1, \ldots, X_n$*. Then for each* $h \in \mathbb{N}$*, as* $n \to \infty$*, the random vector*

$$\sqrt{n}\,(\widehat{\mathbf{r}}_n(h) - \mathbf{r}(h))$$

converges in distribution to the random vector with the multivariate normal distribution $\mathcal{N}_h(\mathbf{0}, \mathbf{W})$*, where*

$$\widehat{\mathbf{r}}_n(h) = (\widehat{r}_n(1), \ldots, \widehat{r}_n(h))^T, \quad \mathbf{r}(h) = (r(1), \ldots, r(h))^T,$$

and $\mathbf{W}$ *is the* $h \times h$ *matrix with elements*

$$w_{ij} = \sum_{k=1}^{\infty} [r(k+i) + r(k-i) - 2r(i)r(k)]\,[r(k+j) + r(k-j) - 2r(j)r(k)]\,,\ i, j = 1, \ldots, h.$$

Proof. See Brockwell and Davis (2006, Theorem 7.2.1). □

Remark: The theorem covers MA(n) and MA(∞) models as well as causal AR(p) and ARMA(p, q) models. Generally speaking, it covers sequences obtained by linear filtration of the strict white noise sequence (having i.i.d. values, not just uncorrelated) with finite fourth moment.

Remark: The formula for w_{ij} is called *the Bartlett formula.* The theorem implies specifically for each $i \in \mathbb{N}$ that

$$\sqrt{n}\,(\widehat{r}_n(i) - r(i)) \xrightarrow{d} \mathcal{N}(0, w_{ii}),\ n \to \infty.$$

Example: Consider an 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. random variables with zero mean, finite positive variance σ^2 and $\mathbb{E}|Y_t|^4 < \infty$. Then $r(k) = \varphi^{|k|}, k \in \mathbb{Z}$. Specifically, $r(1) = \varphi$ and according to Theorem 11.1, we have

$$\sqrt{n}\,(\widehat{r}_n(1) - \varphi) \xrightarrow{d} \mathcal{N}(0, w_{11}),\ n \to \infty,$$

where

$$\begin{aligned} w_{11} &= \sum_{k=1}^{\infty} [r(k+1) + r(k-1) - 2r(1)r(k)]^2 = \sum_{k=1}^{\infty} (\varphi^{k+1} + \varphi^{k-1} - 2\varphi \cdot \varphi^k) \\ &= \sum_{k=1}^{\infty} (\varphi^{k-1} - \varphi^{k+1})^2 = \sum_{k=1}^{\infty} \left[(1-\varphi^2)\varphi^{k-1}\right]^2 = (1-\varphi^2)^2 \sum_{k=1}^{\infty} \varphi^{2(k-1)} \\ &= (1-\varphi^2)^2 \cdot \frac{1}{1-\varphi^2} = 1 - \varphi^2. \end{aligned}$$

If we denote $\widehat{\varphi}_n = \widehat{r}_n(1)$, we can write

$$\sqrt{n}\,(\widehat{\varphi}_n - \varphi) \xrightarrow{d} \mathcal{N}(0, 1-\varphi^2),\ n \to \infty,$$

or

$$\sqrt{n}\frac{\widehat{\varphi}_n - \varphi}{\sqrt{1-\varphi^2}} \xrightarrow{d} \mathcal{N}(0, 1),\ n \to \infty.$$

It follows that $\widehat{\varphi}_n \xrightarrow{\mathbb{P}} \varphi, n \to \infty$ (weak consistency of the estimator) and, by Slutsky's theorem,

$$\sqrt{n}\frac{\widehat{\varphi}_n - \varphi}{\sqrt{1-\widehat{\varphi}_n^2}} \xrightarrow{d} \mathcal{N}(0,1),\ n \to \infty.$$

This allows us to test hypotheses about φ, construct asymptotic confidence intervals, etc. For example, the asymptotic 95% confidence interval is

$$\left(\widehat{\varphi}_n - 1.96\sqrt{\frac{1-\widehat{\varphi}_n^2}{n}}, \widehat{\varphi}_n + 1.96\sqrt{\frac{1-\widehat{\varphi}_n^2}{n}}\right).$$

11.3 Partial autocorrelation function

The sample partial autocorrelation function $\widehat{\alpha}(k), k \in \mathbb{N}$, is obtained using the formula (10.1) by plugging in the values of the sample autocorrelation function $\widehat{r}(k)$:

$$\widehat{\alpha}(1) = \widehat{r}(1), \quad \widehat{\alpha}(k) = \frac{\begin{vmatrix} 1 & \widehat{r}(1) & \cdots & \widehat{r}(k-2) & \widehat{r}(1) \\ \widehat{r}(1) & 1 & \cdots & \widehat{r}(k-3) & \widehat{r}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{r}(k-1) & \widehat{r}(k-2) & \cdots & \widehat{r}(1) & \widehat{r}(k) \end{vmatrix}}{\begin{vmatrix} 1 & \widehat{r}(1) & \cdots & \widehat{r}(k-2) & \widehat{r}(k-1) \\ \widehat{r}(1) & 1 & \cdots & \widehat{r}(k-3) & \widehat{r}(k-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{r}(k-1) & \widehat{r}(k-2) & \cdots & \widehat{r}(1) & 1 \end{vmatrix}}, \ k > 1.$$

The determinant in the denominator is non-zero if $\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}_n\right)^2 > 0$.

The formula (10.1) for $\alpha(k)$ implies that $\alpha(k)$ is a continuous function of $r(1), \ldots, r(k)$, i.e. $\alpha(k) = g(r(1), \ldots, r(k))$ for some continuous function g. Similarly, $\widehat{\alpha}(k) = g(\widehat{r}(1), \ldots, \widehat{r}(k))$. Under the assumptions of Theorem 11.1 it can be shown that $\sqrt{n}\left(\widehat{\alpha}(k) - \alpha(k)\right)$ has asymptotically the $\mathcal{N}(0, \tau^2)$ distribution, where τ^2 depends on the matrix $\mathbf{W}$ from Theorem 11.1 and on the partial derivatives of the function g, provided that g is sufficiently smooth.

12 Estimation of parametric models

12.1 AR(p) sequences

Let us consider a real-valued, weakly stationary, causal AR(p) sequence of order p,

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + Y_t,\ t \in \mathbb{Z},$$

where p is assumed to be known, $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2) and $\varphi_1, \ldots, \varphi_p$ are unknown parameters to be estimated based on observations $X_1, \ldots, X_n$.

Method of moments

Let R be the autocovariance function of the sequence $\{X_t, t \in \mathbb{Z}\}$. Using the Yule-Walker equations we get

$$\begin{aligned} R(0) &= \varphi_1 R(1) + \ldots + \varphi_p R(p) + \sigma^2, \\ R(k) &= \varphi_1 R(k-1) + \ldots + \varphi_p R(k-p),\ k \geq 1. \end{aligned} \tag{12.1}$$

The system of equations for $k = 1, \ldots, p$ can be written in the matrix form $\Gamma\boldsymbol{\varphi} = \boldsymbol{\gamma}$, where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^T$, $\boldsymbol{\gamma} = (R(1), \ldots, R(p))^T$ and

$$\Gamma = \begin{pmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(0) & \cdots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \cdots & R(0) \end{pmatrix}.$$

We replace the values of $R(k)$ in Γ and $\boldsymbol{\gamma}$ by their sample counterparts,

$$\widehat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} \left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right),\ k = 0, 1, \ldots, p,$$

obtaining $\widehat{\Gamma}$ and $\widehat{\boldsymbol{\gamma}}$. Plugging these estimators into the equation $\Gamma\boldsymbol{\varphi} = \boldsymbol{\gamma}$, we obtain the moment estimators of $\varphi_1, \ldots, \varphi_p$ as the solution

$$\widehat{\boldsymbol{\varphi}} = (\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p)^T = \widehat{\Gamma}^{-1}\widehat{\boldsymbol{\gamma}},$$

provided that $\widehat{\Gamma}^{-1}$ exists. We know from Section 11.2 that a sufficient condition for $\widehat{\Gamma}$ 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 estimator of σ^2 is obtained from the equation (12.1) as

$$\widehat{\sigma}^2 = \widehat{R}(0) - \widehat{\varphi}_1\widehat{R}(1) - \ldots - \widehat{\varphi}_p\widehat{R}(p) = \widehat{R}(0) - \widehat{\boldsymbol{\varphi}}^T\widehat{\boldsymbol{\gamma}}.$$

These estimators are often called the Yule-Walker estimators.

Remark: The moment estimators are not very robust (statistically speaking), but are useful for finding approximate estimates as a starting point for more involved methods. Still, the moment estimators are asymptotically normal under certain assumptions.

Theorem 12.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be an AR(p) sequence given by the model*

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + Y_t,\ t \in \mathbb{Z},$$

where $\{Y_t, t \in \mathbb{Z}\}$ *are independent, identically distributed random variables with zero mean and finite positive variance* σ^2*. Suppose that all the roots of the characteristic polynomial* $\lambda^p - \varphi_1\lambda^{p-1} - \ldots - \varphi_p$ *are inside the unit circle, and let* $\widehat{\boldsymbol{\varphi}}_n = (\widehat{\varphi}_{1,n}, \ldots, \widehat{\varphi}_{p,n})^T$ *and* $\widehat{\sigma}_n^2$ *be the moment estimators of* $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^T$ *and* σ^2*, respectively, computed from* $X_1, \ldots, X_n$*. Then,*

$$\sqrt{n}\,(\widehat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}) \xrightarrow{d} \mathcal{N}_p\left(0, \sigma^2\boldsymbol{\Gamma}^{-1}\right),\ n \to \infty,$$

where Γ *is the matrix with elements* $\Gamma_{ij} = R(i-j), 1 \le i, j \le p$*, and* R *is the autocovariance function of* $\{X_t, t \in \mathbb{Z}\}$*. Furthermore, it holds that*

$$\widehat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2,\ n \to \infty.$$

Proof. See Brockwell and Davis (2006, Theorem 8.1.1). □

Remark: Using the method of moments, we have fitted, after subtracting the sample mean, an AR(9) model to the sunspot number time series from Figure 17. The fitted model is

$$\begin{aligned} X_t =& 1.148X_{t-1} - 0.387X_{t-2} - 0.145X_{t-3} + 0.099X_{t-4} - 0.079X_{t-5} \\ &+ 0.037X_{t-6} - 0.016X_{t-7} - 0.028X_{t-8} + 0.214X_{t-9} + Y_t,\ t \in \mathbb{Z}, \end{aligned}$$

with the estimated variance of the white-noise sequence $\widehat{\sigma^2} \doteq 599.8$.

Least squares method

Consider once more the AR(p) sequence given above, and observations $X_1, \ldots, X_n$, where $n > p$. The least squares estimators of $\varphi_1, \ldots, \varphi_p$ are obtained by minimizing the sum of squares

$$\min_{\varphi_1,\ldots,\varphi_p} \sum_{t=p+1}^{n} (X_t - \varphi_1 X_{t-1} - \ldots - \varphi_p X_{t-p})^2.$$

Differentiation with respect to $\varphi_1, \ldots, \varphi_p$ leads to the system of normal equations

$$\sum_{t=p+1}^{n} (X_t - \varphi_1 X_{t-1} - \ldots - \varphi_p X_{t-p}) X_{t-j} = 0,\ j = 1, \ldots, p,$$

or equivalently,

$$\begin{aligned} \varphi_1 \sum_{t=p+1}^{n} X_{t-1}^2 + \ldots + \varphi_p \sum_{t=p+1}^{n} X_{t-1}X_{t-p} &= \sum_{t=p+1}^{n} X_t X_{t-1}, \\ &\vdots \\ \varphi_1 \sum_{t=p+1}^{n} X_{t-1}X_{t-p} + \ldots + \varphi_p \sum_{t=p+1}^{n} X_{t-p}^2 &= \sum_{t=p+1}^{n} X_t X_{t-p}. \end{aligned}$$

If we denote $\mathbf{X}_{t-1} = (X_{t-1}, \ldots, X_{t-p})^T$, we can write this in the matrix form

$$\Sigma_n \boldsymbol{\varphi} = \mathbf{s}_n,$$

where $\mathbf{s}_n = \sum_{t=p+1}^{n} \mathbf{X}_{t-1}X_t = \left(\sum_{t=p+1}^{n} X_{t-1}X_t, \ldots, \sum_{t=p+1}^{n} X_{t-p}X_t\right)^T$ is the vector of the right-hand sides and

$$\Sigma_n = \sum_{t=p+1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}^T = \begin{pmatrix} \sum_{t=p+1}^{n} X_{t-1}^2 & \cdots & \sum_{t=p+1}^{n} X_{t-1}X_{t-p} \\ \vdots & \ddots & \vdots \\ \sum_{t=p+1}^{n} X_{t-1}X_{t-p} & \cdots & \sum_{t=p+1}^{n} X_{t-p}^2 \end{pmatrix}.$$

The solution is obtained as $\widetilde{\boldsymbol{\varphi}}_n = (\widetilde{\varphi}_{1,n}, \ldots, \widetilde{\varphi}_{p,n})^T = \Sigma_n^{-1}\mathbf{s}_n$. The least squares estimator of σ^2 is

$$\widetilde{\sigma}_n^2 = \frac{1}{n-p}\sum_{t=p+1}^{n}\left(X_t - \widetilde{\boldsymbol{\varphi}}^T\mathbf{X}_{t-1}\right)^2.$$

We denote the least squares estimators by $\widetilde{\varphi}_{i,n}$ and $\widetilde{\sigma}_n^2$ to distinguish them from the moment estimators $\widehat{\varphi}_{i,n}$ and $\widehat{\sigma}_n^2$ discussed above. It can be shown that $\widetilde{\boldsymbol{\varphi}}_n$ and $\widetilde{\sigma}_n^2$ have the same asymptotic properties as $\widehat{\boldsymbol{\varphi}}_n$ and $\widehat{\sigma}_n^2$. In particular,

$$\sqrt{n}\left(\widetilde{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}\right) \xrightarrow{d} \mathcal{N}_p\left(0, \sigma^2\boldsymbol{\Gamma}^{-1}\right),\ n \to \infty,$$

where Γ is the matrix from Theorem 12.1, and

$$\widetilde{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2,\ n \to \infty.$$

Maximum likelihood estimation

The maximum likelihood estimation assumes we know the distribution of the random variables used for estimation. Consider first a sequence $\{X_t, t \in \mathbb{Z}\}$ which satisfies the AR(1) model

$$X_t = \varphi X_{t-1} + Y_t,\ t \in \mathbb{Z},$$

where $\{Y_t, t \in \mathbb{Z}\}$ are independent, identically distributed random variables with the $\mathcal{N}(0, \sigma^2)$ distribution. We assume causality of the sequence $\{X_t, t \in \mathbb{Z}\}$, implying $|\varphi| < 1$. Assuming we observe the random variables $X_1, \ldots, X_n$, it follows that X_1 and $(Y_2, \ldots, Y_n)^T$ are independent, with the joint probability density function

$$\begin{aligned} f(x_1, y_2, \ldots, y_n) &= f_1(x_1)f_2(y_2, \ldots, y_n) \\ &= f_1(x_1)\frac{1}{(2\pi\sigma^2)^{(n-1)/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}y_t^2\right\},\ (x_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n. \end{aligned}$$

Using Theorem 3.5, it can be shown that under causality, X_1 has the $\mathcal{N}(0, \tau^2)$ distribution, where $\tau^2 = \frac{\sigma^2}{1-\varphi^2}$.

The transformation theorem implies that the joint probability density function of the random vector $(X_1, \ldots, X_n)^T$ is given by

$$f(x_1, \ldots, x_n) = \frac{\sqrt{1-\varphi^2}}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\left((1-\varphi^2)x_1^2 + \sum_{t=2}^{n}(x_t - \varphi x_{t-1})^2\right)\right\},\ (x_1, \ldots, x_n)^T \in \mathbb{R}^n.$$

The likelihood function $L(\varphi, \sigma^2)$ has the form given by the formula above, with $|\varphi| < 1, \sigma^2 > 0$, where x_i is replaced by $X_i, i = 1, \ldots, n$. The maximum likelihood estimators $\overline{\varphi}$ and $\overline{\sigma}^2$ are the values that maximize $L(\varphi, \sigma^2)$ over the given parametric space. Even for this simple model, it is necessary to solve this non-linear optimization problem numerically.

A simpler solution is provided by the *conditional* maximum likelihood approach. It is easy to see that the conditional density of $(X_2, \ldots, X_n)^T$ given $X_1 = x_1$ in this AR(1) model is

$$f(x_2, \ldots, x_n|x_1) = \frac{1}{(2\pi\sigma^2)^{(n-1)/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(x_t - \varphi x_{t-1})^2\right\},\ (x_1, \ldots, x_n)^T \in \mathbb{R}^n.$$

The conditional maximum likelihood estimators are obtained by maximizing this function with respect to φ and σ^2 over the parametric space $\varphi \in \mathbb{R}, \sigma^2 > 0$. Note that in this approach we did not

assume causality, plus we work with a finite sequence only, meaning all values of $\varphi \in \mathbb{R}$ are indeed relevant and allowed.

If we consider a general AR(p) model $X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + Y_t, t \in \mathbb{Z}$, where $\{Y_t, t \in \mathbb{Z}\}$ are independent, identically distributed random variables with the $\mathcal{N}(0, \sigma^2)$ distribution, the conditional density of $(X_{p+1}, \ldots, X_n)^T$ given $X_1 = x_1, \ldots, X_p = x_p$ is

$$f(x_{p+1}, \ldots, x_n | x_1, \ldots, x_p) = \frac{1}{(2\pi\sigma^2)^{(n-p)/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} (x_t - \boldsymbol{\varphi}^T \boldsymbol{x}_{t-1})^2 \right\}, \ (x_1, \ldots, x_n)^T \in \mathbb{R}^n,$$

where $\boldsymbol{x}_{t-1} = (x_{t-1}, \ldots, x_{t-p})^T$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^T$. By maximization of this function with respect to $\varphi_1, \ldots, \varphi_p$ and σ^2, we obtain the conditional maximum likelihood estimators. It can be shown easily that under the normality assumption, these estimators are numerically equivalent to the least squares estimators (in both cases we look for $\varphi_1, \ldots, \varphi_p$ such that $\sum_{t=p+1}^{n} (X_t - \boldsymbol{\varphi}^T \boldsymbol{X}_{t-1})^2$ is minimal).

12.2 MA(q) sequences

In the previous section, the estimation in the autoregressive models was rather straightforward since we worked with linear regression models where the estimating equations are linear with respect to the parameters. For the moving average sequences (and more generally, the ARMA sequences), this is not the case and the estimation becomes more complicated. Hence, we discuss below only the basic moment method.

Consider the MA(q) sequence given by

$$X_t = Y_t + \theta_1 Y_{t-1} + \ldots + \theta_q Y_{t-q}, \ t \in \mathbb{Z},$$

where $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). Suppose that $\theta_1, \ldots, \theta_q$ and σ^2 are unknown, real-valued parameters to be estimated from the observations $X_1, \ldots, X_n$.

Method of moments

The autocovariance function of the sequence $\{X_t, t \in \mathbb{Z}\}$ is given by the formula

$$R(t) = \begin{cases} \sigma^2 \sum_{k=0}^{q-|t|} \theta_k \theta_{k+|t|}, & |t| \le q, \\ 0, & |t| > q, \end{cases}$$

where we put $\theta_0 = 1$. The moment estimators of $\theta_1, \ldots, \theta_q$ and σ^2 can be obtained from these equations by replacing $R(t)$ with $\widehat{R}(t)$ and solving the system

$$\begin{aligned} \widehat{R}(0) &= \sigma^2 \left(1 + \theta_1^2 + \ldots + \theta_q^2\right) \\ \widehat{R}(1) &= \sigma^2 \left(\theta_1 + \theta_1\theta_2 + \ldots + \theta_{q-1}\theta_q\right), \\ &\vdots \\ \widehat{R}(q) &= \sigma^2 \theta_q. \end{aligned}$$

Note that for such a system of equations, a real-valued solution may not exist or may not be unique.

Example: Consider a MA(1) sequence $X_t = Y_t + \theta Y_{t-1}, t \in \mathbb{Z}$, where $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0, σ^2). Assume we observe the values $X_1, \ldots, X_n$. We know that $R(0) = (1 + \theta^2)\sigma^2$ and $R(1) = \theta\sigma^2$. It follows that $r(1) = R(1)/R(0) = \theta/(1 + \theta^2)$. In order to estimate θ, we first compute $\widehat{r}(1)$ from the data. Then, we solve the equation

$$\widehat{r}(1) = \frac{\theta}{1 + \theta^2}$$

for θ, obtaining $\widehat{\theta}$. Finally, we estimate σ^2 from the variance formula $R(0) = (1+\theta^2)\sigma^2$ as

$$\widehat{\sigma^2} = \frac{\widehat{R}(0)}{1+\widehat{\theta^2}}.$$

We investigate when the estimator $\widehat{\theta}$ exists. We denote $r = \widehat{r}(1)$ and solve the quadratic equation

$$\begin{aligned} r &= \frac{\theta}{1+\theta^2}, \\ 0 &= r\theta^2 - \theta + r, \\ \theta_{1,2} &= \frac{1 \pm \sqrt{1-4r^2}}{2r}. \end{aligned}$$

We see that a real solution exists for $1 - 4r^2 \geq 0$, i.e. for $|r| \leq \frac{1}{2}$. Note also that from the equation $r(1) = \theta/(1+\theta^2)$ it follows that $|r(1)| \leq \frac{1}{2}$ for each $\theta \in \mathbb{R}$.

In summary, for $|\widehat{r}(1)| < \frac{1}{2}$ we have two different real solutions of the quadratic equation, and choosing $\widehat{\theta} = \frac{1-\sqrt{1-4\widehat{r}(1)^2}}{2\widehat{r}(1)}$ leads to an invertible sequence with $|\widehat{\theta}| < 1$.

For $|\widehat{r}(1)| = \frac{1}{2}$, we set $\widehat{\theta} = 1$ for $\widehat{r}(1) = \frac{1}{2}$ and $\widehat{\theta} = -1$ for $\widehat{r}(1) = -\frac{1}{2}$.

For $|\widehat{r}(1)| > \frac{1}{2}$, there is no real solution to the quadratic equation. In a pragmatic way, we choose the estimator for $|\widehat{r}(1)| = \frac{1}{2}$ as our estimator. However, this situation does not occur often since the true value $r(1)$ satisfies $|r(1)| \leq \frac{1}{2}$.

12.3 ARMA(p,q) sequences

Consider the causal, real-valued ARMA(p,q) sequence given by

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \ldots + \theta_q Y_{t-q},\ t \in \mathbb{Z}, \tag{12.2}$$

where $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). Suppose that $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q$ and σ^2 are unknown, real-valued parameters to be estimated from the observations $X_1, \ldots, X_n$.

Method of moments

We use the Yule-Walker equations for the autocovariance function $R_X(k)$ of the sequence $\{X_t, t \in \mathbb{Z}\}$:

$$R_X(k) = \varphi_1 R_X(k-1) + \ldots + \varphi_p R_X(k-p),\ k = q+1, \ldots, q+p.$$

We replace $R_X(k)$ with $\widehat{R}_X(k)$ and solve the system of linear equations to find $\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p$. As a second step, we define $Z_t = X_t - \varphi_1 X_{t-1} - \ldots - \varphi_p X_{t-p}, t \in \mathbb{Z}$, obtaining the MA$(q)$ sequence

$$Z_t = Y_t + \theta_1 Y_{t-1} + \ldots + \theta_q Y_{t-q},\ t \in \mathbb{Z}.$$

The parameters $\theta_1, \ldots, \theta_q$ and σ^2 can be estimated using the approach from Section 12.2, provided that we have the values $\widehat{R}_Z(k)$ available. To find them, we consider the linear filter

$$Z_t = \sum_{j=0}^{p} \beta_j X_{t-j},\ t \in \mathbb{Z},$$

where $\beta_0 = 1, \beta_j = -\varphi_j, j = 1, \ldots, p$, and recall that in this case

$$R_Z(k) = \sum_{j=0}^{p} \sum_{l=0}^{p} \beta_j \beta_l R_X(k+j-l),\ k \in \mathbb{Z}.$$

We plug in the sample autocovariances $\widehat{R}_X(k)$ computed from $X_1, \ldots, X_n$ and the values $\widehat{\beta}_j = -\widehat{\varphi}_j$ estimated above to get $\widehat{R}_Z(k)$.

The moment estimators are under certain assumptions consistent and asymptotically normal, but they have large variance compared to maximum likelihood estimators and are not very robust. Nevertheless, they can provide preliminary estimates as a starting point for more involved estimation methods.

13 Estimation of spectral density

13.1 Periodogram

Definition 13.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a random sequence and let $X_1, \ldots, X_n$ be the available observations.* The periodogram *of $X_1, \ldots, X_n$ is defined as*

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t \mathrm{e}^{-it\lambda} \right|^2, \ \lambda \in [-\pi, \pi].$$

Remark: Periodogram is usually computed at points $\lambda_j = \frac{2\pi j}{n}$ for $j \in \mathbb{Z}$ such that $\lambda_j \in [-\pi, \pi]$, called *the Fourier frequencies*. Some authors even define the periodogram only in these frequencies. It is interesting to note that the periodogram values, computed at the Fourier frequencies, are not influenced by the possible centering of the sequence. This is because for any constant c and any Fourier frequency λ_j it holds that $\sum_{t=1}^{n} c\, \mathrm{e}^{-it\lambda_j} = 0$.

Remark: Periodogram has been proposed as an estimator of the spectral density and as a tool for the identification of periodic components in time series. For illustration, consider the sunspot number time series from Figure 17. The corresponding periodogram is shown in Figure 19. The highest peak in the periodogram is located at the frequency $\lambda \doteq 0.58$, which corresponds to the period of $2\pi/\lambda \doteq 10.8$ time units (years). This result is consistent with the well-known 11-year period of the solar cycle. Note that since the sequence is real-valued, its spectral density is symmetric around the origin, and the periodogram is plotted for non-negative frequencies only, as usual in practice. The same applies to other figures in this chapter.

Remark: To actually compute the values of the periodogram, it is more convenient to use the form

$$I_n(\lambda) = \frac{1}{4\pi} \left[A(\lambda)^2 + B(\lambda)^2 \right], \ \lambda \in [-\pi, \pi],$$

where

$$A(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \cos(t\lambda), \quad B(\lambda) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \sin(t\lambda), \ \lambda \in [-\pi, \pi].$$

Remark: Alternatively, for a real-valued sequence the periodogram can be expressed as

$$\begin{aligned} I_n(\lambda) &= \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} X_t X_s \mathrm{e}^{-i(t-s)\lambda} = \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} \sum_{s=\max(1,1-k)}^{\min(n,n-k)} X_s X_{s+k} \mathrm{e}^{-ik\lambda} \\ &= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \mathrm{e}^{-ik\lambda} C_k, \end{aligned} \tag{13.1}$$

where

$$C_k = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}, & k \geq 0, \\ C_{-k}, & k < 0. \end{cases}$$

Remark: Note that the formula (13.1) resembles the inverse formula for computing the spectral density: $f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-ik\lambda} R(k), \lambda \in [-\pi, \pi]$. Since C_k can be viewed as an estimator of $R(k)$, $I_n(\lambda)$ can be viewed as an estimator of $f(\lambda)$.

Theorem 13.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real-valued, weakly stationary random sequence with the mean value μ, spectral density f and the autocovariance function R such that $\sum_{k=-\infty}^{\infty} |R(k)| < \infty$.*

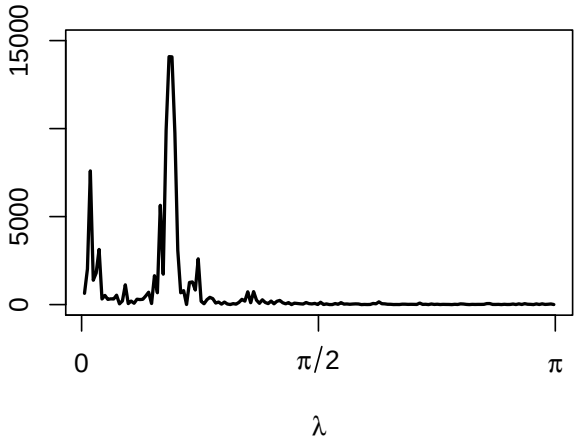

Figure 19: Periodogram of the sunspot number time series.

1. *If* $\mu = 0$, *then* $\mathbb{E}I_n(\lambda) \to f(\lambda), n \to \infty, \lambda \in [-\pi, \pi]$.

2. *If* $\mu \neq 0$, *then for* $n \to \infty$ *we have*

$$\mathbb{E}I_n(\lambda) \to f(\lambda),\ \lambda \in [-\pi, \pi] \setminus \{0\},$$
$$\mathbb{E}I_n(0) - \frac{n\mu^2}{2\pi} \to f(0).$$

Proof. 1. For a centered sequence we have

$$\mathbb{E}I_n(\lambda) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{e}^{-i(t-s)\lambda} \mathbb{E}X_t X_s = \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} \mathrm{e}^{-ik\lambda} R(k)(n - |k|),\ \lambda \in [-\pi, \pi],$$

see the computation of $\varphi_n(\lambda)$ in the proof of Theorem 5.4.

According to Theorem 5.7, the spectral density of $\{X_t, t \in \mathbb{Z}\}$ exists and is given by

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

Using the same arguments as in the proof of Theorem 5.7 we have for $\lambda \in [-\pi, \pi]$:

$$|f(\lambda) - \mathbb{E}I_n(\lambda)| \leq \frac{1}{2\pi} \sum_{|k| \geq n} |R(k)| + \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} |R(k)| \cdot |k| \to 0,\ n \to \infty,$$

where the first term in the upper bound is a remainder of a convergent series, and the second term converges to 0 due to the Kronecker lemma.

2. For a non-centered sequence with mean μ we have

$$\mathbb{E}I_n(\lambda) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{e}^{-i(t-s)\lambda} \mathbb{E}X_t X_s = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{e}^{-i(t-s)\lambda} (R(t-s) + \mu^2)$$
$$= \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} \mathrm{e}^{-ik\lambda} R(k)(n - |k|) + \frac{\mu^2}{2\pi n} \left| \sum_{t=1}^{n} \mathrm{e}^{-it\lambda} \right|^2,\ \lambda \in [-\pi, \pi].$$

The first term converges to $f(\lambda)$ for $\lambda \in [-\pi, \pi]$ according to the first part of the proof. Concerning the second term, for $\lambda \neq 0$ it equals $\frac{\mu^2}{2\pi n} \frac{1-\cos(n\lambda)}{1-\cos\lambda}$ and converges to 0 for $n \to \infty$; for $\lambda = 0$ it equals $\frac{n\mu^2}{2\pi}$. □

Remark: The previous theorem establishes asymptotic unbiasedness of $I_n(\lambda)$ as the estimator of $f(\lambda)$. However, consistency does not hold, e.g. it can be shown that for any Gaussian, stationary sequence with continuous spectral density f it holds that

$$\lim_{n\to\infty} \operatorname{var} I_n(\lambda) = \begin{cases} f(\lambda)^2, & \lambda \in (-\pi,\pi) \setminus \{0\}, \\ 2f(\lambda)^2, & \lambda \in \{-\pi, 0, \pi\}, \end{cases}$$

see Anděl (1976, p. 103). Even in this simple case, the variance of $I_n(\lambda)$ does not converge to 0.

Remark: The periodogram can detect hidden periodic components in a time series. For illustration, consider a sequence $\{X_t, t \in \mathbb{Z}\}$ such that $X_t = \alpha \mathrm{e}^{it\lambda_0} + Y_t, t \in \mathbb{Z}$, where $\alpha \neq 0$ is a constant, $\lambda_0 \in [-\pi, \pi]$, and $\{Y_t, t \in \mathbb{Z}\}$ is a white noise sequence WN(0,σ^2). Then,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \mathrm{e}^{-it\lambda} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \mathrm{e}^{-it\lambda} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha \mathrm{e}^{-it(\lambda-\lambda_0)}, \ \lambda \in [-\pi, \pi].$$

This means that for $\lambda = \lambda_0$, the nonrandom part of the periodogram, corresponding to the second term of the previous formula, tends to $+\infty$ as $n \to \infty$. For $\lambda \neq \lambda_0$, this term is close to 0, see the second part of the proof of Theorem 13.1 for $\sum_{t=1}^{n} \mathrm{e}^{-it\lambda}$. It means that if there is a single periodic component at frequency λ_0, the periodogram takes its largest value at this frequency. Usually, the frequency λ_0 is not known, and it is reasonable to consider the maximum value of the periodogram at the Fourier frequencies when looking for periodic components.

Theorem 13.2. *Let $\{X_t, t \in \mathbb{Z}\}$ be a Gaussian random sequence of independent, identically distributed random variables with zero mean and variance $\sigma^2 \in (0, \infty)$. 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*

$$W = \frac{\max_{1 \le r \le m} I_n(\lambda_r)}{I_n(\lambda_1) + \ldots + I_n(\lambda_m)}$$

has the probability density function

$$g(x) = m(m-1) \sum_{j=1}^{\lfloor 1/x \rfloor} (-1)^{j-1} \binom{m-1}{j-1} (1 - jx)^{m-2}, \ 0 < x < 1,$$

and furthermore

$$\mathbb{P}(W > x) = 1 - \sum_{k=0}^{\lfloor 1/x \rfloor} (-1)^k \binom{m}{k} (1 - kx)^{m-1}, \ 0 < x < 1. \tag{13.2}$$

Proof. See Anděl (1976, p. 79–82). □

13.2 Fisher test of periodicity

Based on the Theorem 13.2, we want to test the null hypothesis of no periodic component in the time series, $H_0 : X_1, \ldots, X_n$ are i.i.d. with the $\mathcal{N}(0, \sigma^2)$ distribution, against the alternative that H_0 is violated. The alternative hypothesis H_1 assumes the model

$$X_t = \sum_{j=1}^{p} (\alpha_j \cos(\lambda_j t) + \beta_j \sin(\lambda_j t)) + \varepsilon_t, \ t \in \{1, \ldots, n\}, \tag{13.3}$$

where $p, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p$ are unknown parameters, $\lambda_1, \ldots, \lambda_p$ are unknown frequencies and $\varepsilon_1, \ldots, \varepsilon_n$ are independent, identically distributed random variables with the $\mathcal{N}(0, \sigma^2)$ distribution.

Under the null hypothesis, all the values $I_n(2\pi/n), \ldots, I_n(2\pi m/n)$ should be similar, and the value of the test statistic W from Theorem 13.2 should be close to $1/m$. Large values of W indicate a violation of H_0 and the presence of periodic components in the time series (or correlations in the time series, but we do not consider that in H_1). We reject H_0 on the significance level α if $W > c_\alpha$, where the critical value c_α can be computed from (13.2).

Remark: For $m \leq 50$, we can use for finding c_α just the first two terms in (13.2), i.e. the approximation $\mathbb{P}(W > x) \approx m(1-x)^{m-1}$. For $m > 50$, we can use the asymptotic distribution

$$\lim_{m\to\infty} \mathbb{P}\left(W > \frac{x + \ln m}{m}\right) = 1 - \exp\{-\exp\{-x\}\},\ x > 0.$$

Remark: If the Fisher test rejects the null hypothesis, we accept the alternative that the mean of $\{X_t, t \in \mathbb{Z}\}$ contains a periodic component with frequency $\lambda_{(1)}$ corresponding to the maximum value of the periodogram $V_{(1)}$, where we denote $V_{(1)} \geq V_{(2)} \geq \ldots \geq V_{(m)}$ the ordered values of $I_n(\lambda_r), r = 1, \ldots, m$.

Now we can determine further periodic components with a modification of the Fisher test. Note that the test was derived under the null hypothesis which we have just rejected, so the test is not exact, but it works well in practice (Anděl, 1976, p. 86). The original Fisher test is based on the test statistic

$$W = W_1 = \frac{V_{(1)}}{V_{(1)} + \ldots + V_{(m)}}$$

and the critical values are obtained from (13.2). The modified test is based on the test statistic

$$W_2 = \frac{V_{(2)}}{V_{(2)} + \ldots + V_{(m)}}$$

and the critical values are obtained from (13.2) with $m-1$ in place of m. We proceed like this until all remaining frequencies are not significant. If the frequencies $\lambda_{(1)}, \ldots, \lambda_{(p)}$ were found significant, we accept the model (13.3) with these frequencies. The parameters $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p$ can be estimated using e.g. the least squares method.

13.3 Estimation of spectral density

We have shown in Theorem 13.1 that the periodogram is an asymptotically unbiased estimator of the spectral density. However, it is not consistent even for the Gaussian white noise sequence. On the other hand, kernel estimation can be applied in this setting. We remark that in this case, a centered sequence should be used for estimation due to possible interpretability issues – artificially high values tend to occur for low frequencies when a weakly stationary sequence with non-zero mean is used directly for estimation.

Let K be a kernel function on the interval $[-\pi, \pi]$, satisfying

$$K(\lambda) \geq 0, \quad K(\lambda) = K(-\lambda), \quad \int_{-\pi}^{\pi} K(\lambda)\,\mathrm{d}\lambda = 1, \quad \int_{-\pi}^{\pi} K^2(\lambda)\,\mathrm{d}\lambda < \infty.$$

Under certain assumptions, $\int_{-\pi}^{\pi} I_n(\lambda)K(\lambda)\,\mathrm{d}\lambda$ can be shown to be asymptotically unbiased and consistent estimator of $\int_{-\pi}^{\pi} f(\lambda)K(\lambda)\,\mathrm{d}\lambda$. If the function K is concentrated around 0, the statistic

$$\widehat{f_n}(\lambda_0) = \int_{-\pi}^{\pi} K(\lambda - \lambda_0) I_n(\lambda)\,\mathrm{d}\lambda$$

can be considered a consistent estimator of $f(\lambda_0), \lambda_0 \in [-\pi, \pi]$.

If we expand the function K into the Fourier series $K(\lambda) = \sum_{k=-\infty}^{\infty} w_k \mathrm{e}^{ik\lambda}, \lambda \in [-\pi, \pi]$, where $w_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-ik\lambda} K(\lambda)\,\mathrm{d}\lambda$ and use the formula (13.1), we get for $\lambda_0 \in [-\pi.\pi]$:

$$\begin{aligned}
\widehat{f}_n(\lambda_0) &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \mathrm{e}^{-ik\lambda} C_k K(\lambda - \lambda_0)\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \int_{-\pi}^{\pi} \mathrm{e}^{-ik\lambda} K(\lambda - \lambda_0)\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \int_{-\pi}^{\pi} \mathrm{e}^{-ik\lambda} \sum_{j=-\infty}^{\infty} w_j \mathrm{e}^{ij(\lambda-\lambda_0)}\,\mathrm{d}\lambda \\
&= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} C_k \sum_{j=-\infty}^{\infty} w_j \mathrm{e}^{-ij\lambda_0} \int_{-\pi}^{\pi} \mathrm{e}^{ij\lambda - ik\lambda}\,\mathrm{d}\lambda \\
&= \sum_{k=-n+1}^{n-1} C_k w_k \mathrm{e}^{-ik\lambda_0} = C_0 w_0 + 2 \sum_{k=1}^{n-1} C_k w_k \cos(k\lambda_0).
\end{aligned}$$

We note that only the Fourier coefficient $w_0, \ldots, w_{n-1}$ play a role, hence the same estimator is given for all the kernel functions sharing the same set of Fourier coefficients $w_0, \ldots, w_{n-1}$.

Remark: The coefficients C_k have large variability for large values of k (close to n). Often, truncated estimators are considered:

$$\widehat{f}_n(\lambda_0) = \sum_{k=-M}^{M} C_k w_k \mathrm{e}^{-ik\lambda_0} = C_0 w_0 + 2 \sum_{k=1}^{M} C_k w_k \cos(k\lambda_0),$$

where the truncation point is usually chosen as $\frac{n}{6} < M < \frac{n}{5}$.

Remark: There are many ways how to choose the kernel function K (or more frequently the coefficients $w_0, \ldots, w_k$) so that the estimator of the spectral density exhibits desirable properties such as non-negative values, low bias, or a good order of convergence in consistency. Popular choices include the *Bartlett estimator*

$$w_k = \begin{cases} \frac{1}{2\pi}\left(1 - \frac{|k|}{M}\right), & |k| \le M, \\ 0, & |k| > M, \end{cases}$$

or the *Blackman-Tukey estimator*

$$w_k = \begin{cases} \frac{1}{2\pi}\left[1 - 2a + 2a\cos\left(\frac{\pi k}{M}\right)\right], & |k| \le M, \\ 0, & |k| > M, \end{cases}$$

where $0 < a \le 0.25$. Both the Bartlett estimator and the Blackman-Tukey estimator with $a = 0.25$ provide non-negative estimates. The Blackman-Tukey estimator with $a = 0.23$ (another traditional choice) has a smaller bias but can result in negative estimates.

One of the most frequently used estimators is also the *Parzen estimator*

$$w_k = \begin{cases} \frac{1}{2\pi}\left[1 - 6\left(\frac{|k|}{M}\right)^2 + 6\left(\frac{|k|}{M}\right)^3\right], & |k| \le \frac{M}{2}, \\ \frac{1}{2\pi}\left[2\left(1 - \frac{|k|}{M}\right)^3\right], & \frac{M}{2} < |k| \le M, \\ 0, & |k| > M. \end{cases}$$

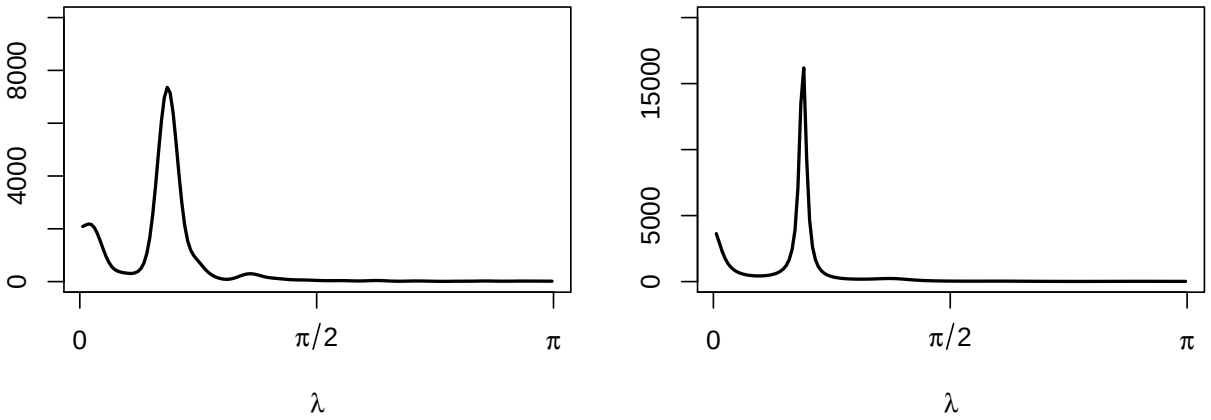

Figure 20: Nonparametric estimate of the spectral density of the sunspot number time series, using the Parzen window with $M = 60$ (left), and the parametric estimate based on the fitted AR(9) model (right).

This estimator also provides non-negative estimates and leads to smoother estimates of the spectral density. For illustration, such nonparametric estimate is plotted in the left panel of Figure 20 for the sunspot number time series.

Remark: The previous estimator is nonparametric. If a parametric model for the observed sequence is available, meaning that a parametric formula for the spectral density is at hand, we can plug in the estimated values of the model parameters to obtain a parametric estimate of the spectral density. As an illustration, the parametric estimate for the sunspot number time series is plotted in the right panel of Figure 20.

References

Anděl, J. (1976) *Statistická analýza časových řad.* SNTL, Prague.

Brockwell, P. J. and Davis, R. A. (2006) *Time series: theory and methods.* Springer-Verlag, New York.

Kallenberg, O. (2002) *Foundations of modern probability.* Springer, New York.

Loève, M. (1963) *Probability theory.* Van Nostrand, Princeton.

Priestley, M. B. (1981) *Spectral analysis and time series*, vol. I. Academic Press, London.

Rao, R. C. (1978) *Lineární metody statistické indukce a jejich aplikace.* Academia, Prague.

Rudin, W. (2003) *Analýza v reálném a komplexním oboru.* Academia, Prague.

Štěpán, J. (1987) *Teorie pravděpodobnosti: matematické základy.* Academia, Prague.