# Lecture Notes
# Time Series Analysis

*Summer semester 2022*

Friedrich-Schiller-Universität Jena

Michael H. Neumann

This version: May 18, 2022

## Important information

- Lecture period: April 11 – July 15

- Examination period: July 18 – August 12
  There will be oral examinations, in the second half of the official examination period. Dates for these examinations will be fixed in good time.

- Contact
  - Office: room 3516
  - Phone: 03641/946260
  - Email: michael.neumann@uni-jena.de

# Literature


[1] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd Edition. Springer. New York.

[2] Kreiß, J.-P. and Neuhaus, G. (2006). *Einführung in die Zeitreihenanalyse*. Springer. Berlin. (in German)

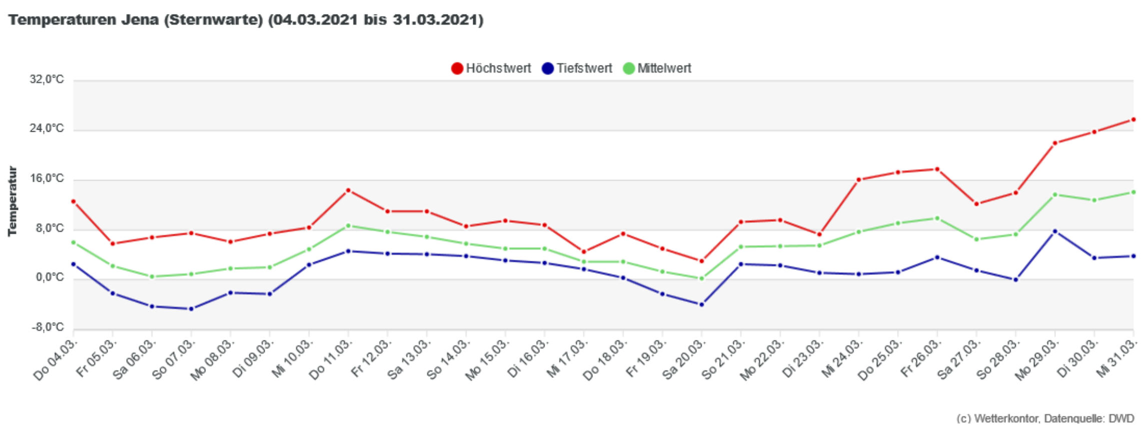
# Contents

# 1 Models for time series

This course is intended to familiarize you with some popular models for time series and their properties, as well as with statistical methods for estimating unknown parameters. The following table shows measurements of the daily temperatures (in degree centigrade) in the city of Jena in March 2021. These data, which are recorded at discrete times, form a so-called time series. Loosely speaking, a time series is a set $(x_t)_{t\in T}$ of observations, each one being recorded at a specified time $t$. In our case, we have three series, $(x_t^{(i)})_{t\in\{4,5,\dots,31\}}$, where $x_t^{(i)}$ is the maximum, minimum or average temperature measured at day $t$, for $i = 1, 2, 3$, respectively.

| day | maximum | minimum | average | day | maximum | minimum | average |
|---|---|---|---|---|---|---|---|
| 3/4 | 12.6 | 2.5 | 6.0 | 3/18 | 7.4 | 0.3 | 2.9 |
| 3/5 | 5.8 | -2.2 | 2.2 | 3/19 | 5.0 | -2.3 | 1.3 |
| 3/6 | 6.8 | -4.3 | 0.5 | 3/20 | 3.0 | -4.0 | 0.2 |
| 3/7 | 7.5 | -4.7 | 0.9 | 3/21 | 9.3 | 2.5 | 5.3 |
| 3/8 | 6.1 | -2.1 | 1.8 | 3/22 | 9.6 | 2.3 | 5.4 |
| 3/9 | 7.4 | -2.3 | 2.0 | 3/23 | 7.3 | 1.1 | 5.5 |
| 3/10 | 8.4 | 2.4 | 4.9 | 3/24 | 16.1 | 0.9 | 7.7 |
| 3/11 | 14.4 | 4.6 | 8.7 | 3/25 | 17.3 | 1.2 | 9.1 |
| 3/12 | 11.0 | 4.2 | 7.7 | 3/26 | 17.8 | 3.6 | 9.9 |
| 3/13 | 11.0 | 4.1 | 6.9 | 3/27 | 12.2 | 1.5 | 6.5 |
| 3/14 | 8.6 | 3.8 | 5.8 | 3/28 | 14.0 | 0.0 | 7.3 |
| 3/15 | 9.5 | 3.1 | 5.0 | 3/29 | 22.0 | 7.8 | 13.7 |
| 3/16 | 8.8 | 2.7 | 5.0 | 3/30 | 23.8 | 3.5 | 12.8 |
| 3/17 | 4.5 | 1.7 | 2.9 | 3/31 | 25.8 | 3.8 | 14.1 |

Some structure in these three time series can be detected in the picture on the bottom of this page. The red curve shows the maximum temperatures while the blue and the green curves show the minimum and average temperatures, respectively. When we think of these measurements as **realizations of random variables** we can guess that these random variables show a similar behavior over the entire period of measurements, a property which will be called "stationarity". Moreover, temperatures measured at day $t$ are not far from those measured at day $t-1$. This indicates that an appropriate model for our data should allow for dependence between the random variables. This course will familiarize you with a few simple models for time series and with tools to deal with time series data.



## 1.1 Basic concepts, the Daniell-Kolmogorov existence theorem

In this subsection we introduce a few basic concepts which will be needed throughout this course. We begin with a formal definition.

**Definition.** A **stochastic process** is a family of random variables $\mathbf{X} = (X_t)_{t \in T}$ defined on a common probability space $(\Omega, \mathcal{F}, P)$.
For each $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ on $T$ is called **realization**, **trajectory** or **sample path** of the process $\mathbf{X}$.
The term **time series** is used for the process $\mathbf{X}$ but also for a realization of $\mathbf{X}$, where the index set $T$ is usually some set of equidistant points in $\mathbb{R}$, usually but not necessarily thought of time points.

In this course we restrict our attention to the following cases:

- $T = \mathbb{N} = \{1, 2, \ldots\}$, $T = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ or $T = \mathbb{Z}$
- $X_t$ takes values in $\mathbb{R}$ or (sometimes) in $\mathbb{C}$ or $\mathbb{R}^d$

**Problem:** Suppose we have a (real) time series $(X_t)_{t \in T}$ which is defined on a probability space $(\Omega, \mathcal{F}, P)$. Since each of the $X_t$ is a random variable on $(\Omega, \mathcal{F}, P)$ it is clear that $X^{-1}(B) := \big\{\omega \in \Omega \colon X(\omega) \in B\big\}$ holds for all Borel sets $B \in \mathcal{B}$, and so $P\big(\{\omega \colon X_t(\omega) \in B\}\big)$ is well-defined. (Note that $X^{-1}$ does **not** denote the inverse mapping; $X^{-1}(B)$ is the inverse image (preimage) of the set $B$.) What is less clear, however, is to what extent we can draw conclusions about the random behavior of this process over **finite** or even **infinite time periods**.
Let, for definiteness, $T = \mathbb{N}$. Does the probability measure $P$ carry information about the "joint distribution" of a finite or even an infinite number of random variables $X_1, X_2, \ldots$? In other words, for what kind of sets $C \subseteq \mathbb{R}^\infty := \big\{(x_1, x_2, \ldots) \colon x_t \in \mathbb{R}\big\}$ is the probability of the event $\{\omega \in \Omega \colon (X_1(\omega), X_2(\omega), \ldots) \in C\}$ specified by the given probability measure $P$? Since $P$ is defined on the $\sigma$-Algebra $\mathcal{F}$, we have to identify sets $C \subseteq \mathbb{R}^\infty$ such that

$$\{\omega \in \Omega \colon (X_1(\omega), X_2(\omega), \ldots) \in C\} \in \mathcal{F}. \tag{1.1.1}$$

To this end, we make use of the following result from measure theory:

**Lemma 1.1.1.** *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, and let $f \colon \Omega \to \Omega'$ be an arbitrary mapping. If $f^{-1}(E') \in \mathcal{F}$ holds for all $E' \in \mathcal{E}'$, where $\mathcal{E}'$ is a collection of subsets of $\Omega'$ such that $\sigma\big(\mathcal{E}'\big) = \mathcal{F}'$, then the mapping $f$ is $(\mathcal{F} - \mathcal{F}')$-measurable.*

*Proof.* We use the **good set principle** and define the system of good sets,

$$\mathcal{G} := \big\{E' \subseteq \Omega' \colon f^{-1}(E') \in \mathcal{F}\big\}.$$

The set $\mathcal{G}$ is a $\sigma$-algebra on $\Omega'$. Indeed, we have:

a) $f^{-1}(\Omega') = \Omega \in \mathcal{F}$, hence $\Omega' \in \mathcal{G}$.

b) If $E' \in \mathcal{G}$, then $f^{-1}(E') \in \mathcal{F}$, and so $f^{-1}(E'^c) = \big(f^{-1}(E')\big)^c \in \mathcal{F}$, which means that $E'^c \in \mathcal{G}$.

c) If $E'_1, E'_2, \ldots \in \mathcal{G}$, then $f^{-1}(E'_1), f^{-1}(E'_2), \ldots \in \mathcal{F}$, and hence $f^{-1}\Big(\bigcup_{i=1}^\infty E'_i\Big) = \bigcup_{i=1}^\infty f^{-1}\big(E'_i\big) \in \mathcal{F}$. This implies that $\bigcup_{i=1}^\infty E'_i \in \mathcal{G}$.

Since by assumption $\mathcal{E}' \subseteq \mathcal{G}$ we therefore obtain that

$$\mathcal{F}' \,=\, \sigma\big(\mathcal{E}'\big) \,\subseteq\, \sigma\big(\mathcal{G}\big) \,=\, \mathcal{G},$$

i.e. $f^{-1}(E') \in \mathcal{F}$ for all $E' \in \mathcal{F}'$. Hence, the mapping $f\colon\ \Omega \to \Omega'$ is $(\mathcal{F} - \mathcal{F}')$-measurable. $\square$

By assumption, $X_1, X_2, \ldots$ are random variables which means that $X_t\colon\ \Omega \to \mathbb{R}$ is $(\mathcal{F}-\mathcal{B})$-measurable, i.e. $X_t^{-1}(B) := \{\omega\colon\ X_t(\omega) \in B\} \in \mathcal{F}$ holds for all $B \in \mathcal{B}$. It follows from Lemma 1.1.1 that $(X_1, \ldots, X_n)\colon\ \Omega \to \mathbb{R}^n$ is $(\mathcal{F} - \mathcal{B}^n)$-measurable. Indeed, we have, for arbitrary $B_1, \ldots, B_n \in \mathcal{B}$,

$$\begin{aligned}(X_1, \ldots, X_n)^{-1}\big(B_1 \times \cdots \times B_n\big) \;&=\; \big\{\omega\colon\ \big(X_1(\omega), \ldots, X_n(\omega)\big) \in B_1 \times \cdots \times B_n\big\} \\ &=\; \underbrace{\big\{\omega\colon\ X_1(\omega) \in B_1\big\}}_{\in \mathcal{F}} \cap \ldots \cap \underbrace{\big\{\omega\colon\ X_n(\omega) \in B_n\big\}}_{\in \mathcal{F}} \;\in\; \mathcal{F}.\end{aligned}$$

Since $\sigma\big(\{B_1 \times \cdots \times B_n\colon\ B_1, \ldots, B_n \in \mathcal{B}\}\big) = \mathcal{B}^n$ we obtain by Lemma 1.1.1 that

$$(X_1, \ldots, X_n)^{-1}\big(B\big) \,\in\, \mathcal{F} \qquad \forall B \in \mathcal{B}^n. \tag{1.1.2}$$

(1.1.2) means that the joint distribution of a **finite** number of random variables is specified by $P$.

The transition from the finite-dimensional to the infinite-dimensional case is achieved by using so-called **finite-dimensional sets** (cylinder sets). Let

$$\mathcal{C}_n \,:=\, \Big\{\big\{(x_1, x_2, \ldots)\colon\ (x_1, \ldots, x_n) \in B, x_{n+1}, x_{n+2}, \ldots \in \mathbb{R}\big\}\colon\ B \in \mathcal{B}^n\Big\} \,=\, \big\{B \times \mathbb{R}^\infty\colon\ B \in \mathcal{B}^n\big\}$$

be the collection of all $n$-dimensional cylinder sets. Then the union of these sets,

$$\mathcal{C} \,=\, \bigcup_{n=1}^\infty \mathcal{C}_n,$$

is the family of all **cylinder sets**. $\mathcal{C}$ is an algebra, and therefore also a ring on $\mathbb{R}^\infty$ but **not** a $\sigma$-algebra. ($\mathcal{C}$ contains $\mathbb{R}^\infty$ and is stable under the formations of complementation and finite unions; but is not stable under the formation of countable unions.)
Let $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots)$ and let $C \in \mathcal{C}$ be an arbitrary cylinder set. Then, there exist some $n \in \mathbb{N}$ and $C_n \in \mathcal{B}^n$ such that $C = C_n \times \mathbb{R}^\infty$. Hence

$$\mathbf{X}^{-1}(C) \,=\, \big\{\omega\colon\ \mathbf{X}(\omega) \in C\big\} \,=\, \big\{\omega\colon\ \big(X_1(\omega), \ldots, X_n(\omega)\big) \in C_n\big\} \,\in\, \mathcal{F}.$$

This implies, again by Lemma 1.1.1, that

$$\{\omega \in \Omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\} \,=\, \mathbf{X}^{-1}(C) \,\in\, \mathcal{F} \qquad \forall C \in \sigma(\mathcal{C}). \tag{1.1.3}$$

While the $\sigma$-algebra $\sigma(\mathcal{C})$ generated by the cylinder sets is too small for some purposes in the case of processes in continuous time (i.e. $T = [0, \infty)$ or $T = \mathbb{R}$), this set is usually rich enough when we deal with processes in discrete time. For example, the set $\big\{x \in \mathbb{R}^\infty \colon\ n^{-1}\sum_{t=1}^n x_t \underset{n\to\infty}{\longrightarrow} \mu\big\}$ ($\mu \in \mathbb{R}$) is contained in $\sigma(\mathcal{C})$ which means that the probability $P\big(\{\omega\colon\ \bar{X}_n(\omega) \underset{n\to\infty}{\longrightarrow} \mu\}\big)$ is well-defined, where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$ is the mean of a sample of size $n$. Although the probabilities of events $\{\omega \in \Omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\}$ are well-

specified for all sets $C \in \sigma(\mathcal{C})$, it might be difficult or even impossible to compute the probabilities $P\big(\{\omega \in \Omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\}\big)$ explicitly unless the set $C \in \sigma(\mathcal{C})$ has a very simple structure. This is because $\mathbf{X} = (X_1, X_2, \ldots)$ is an infinite-dimensional object and sets $C$ in $\sigma(\mathcal{C})$ may have a complex structure, leaving the simple case of cylinder sets aside. Fortunately, it turns out that important aspects of the behavior of a process $(X_t)_{t\in T}$ can be read off from the so-called **finite-dimensional distributions**. Here is a formal definition:

**Definition.** Let $(X_t)_{t\in T}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$. For $k \in \mathbb{N}$ and distinct times $t_1, \ldots, t_k \in T$, $P^{X_{t_1},\ldots,X_{t_k}}$ is a **finite-dimensional distribution** of the process $\mathbf{X}$.
$(P^{X_{t_1},\ldots,X_{t_k}}(B) := P\big(\{\omega \in \Omega\colon\ (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in B\}\big)$ for $B \in \mathcal{B}^n$.)

Knowledge of the finite-dimensional distributions is sufficient for many purposes. Actually, if for a process $(X_t)_{t\in\mathbb{N}}$ the probabilities $P^{X_1,\ldots,X_n}(B) = P\big(\{\omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in B \times \mathbb{R}^\infty\}\big)$ are given, it follows from the uniqueness theorem of measure theory that $P\big(\{\omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\}\big)$ is specified for all $C \in \sigma(\mathcal{C})$. Therefore, it should not come as a surprise when the definition of stationarity given in the next Subsection 1.2S2.1 is based on the finite-dimensional distributions.

## Exercises

**Ex. 1.1.1** Show that $\mathcal{C}$ is an algebra but not a $\sigma$-algebra on $\mathbb{R}^\infty$.

**Ex. 1.1.2** Show that, for $\mu \in \mathbb{R}$,

$$\Big\{x \in \mathbb{R}^\infty\colon\ \frac{1}{n}\sum_{t=1}^n x_t \underset{n\to\infty}{\longrightarrow} \mu\Big\} \in \sigma\big(\mathcal{C}\big).$$

**Ex. 1.1.3** Let $(X_t)_{t\in[0,\infty)}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ such that

- $X_0 = 0$ with probability 1,
- for $0 < t_1 < t_2 < \ldots < t_k$, $k \in \mathbb{N}$, the increments $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are stochastically independent.
- for $s < t$, $X_t - X_s \sim \mathcal{N}\big(0, t-s\big)$.

Find the finite-dimensional distributions $P^{X_{t_1},\ldots,X_{t_k}}$.

With our definition of a time series $(X_t)_{t\in T}$ we have tacitly assumed that we have an infinite number of random variables on a probability space $(\Omega, \mathcal{F}, P)$ at our disposal. But does such a probability space that supports a countable or even uncountable number of random variables with given properties exist a all? Few results in probability theory are more fundamental or more well-known than the Daniell-Kolmogorov existence theorem. It was first discovered by the British mathematician Percy John Daniell in a slightly different setting, and later rediscovered by the famous Russian mathematician Andrey Nikolaevich Kolmogorov. This theorem is also referred to Kolmogorov existence theorem, Kolmogorov extension theorem or Kolmogorov consistency theorem. It basically states, for any "reasonable" family $\{\mu_{t_1,\dots,t_k}\colon t_1,\dots,t_k \in T\ (t_i \neq t_j \text{ for } i \neq j), k \in \mathbb{N}\}$ of probability distributions ($\mu_{t_1,\dots,t_k}$ is a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$), that there exists a suitable probability space $\big(\Omega, \mathcal{F}, P\big)$ and a stochastic process $\mathbf{X} = (X_t)_{t\in T}$ on $\big(\Omega, \mathcal{F}, P\big)$ such that

$$P^{X_{t_1},\dots,X_{t_k}} = \mu_{t_1,\dots,t_k} \qquad \forall k \in \mathbb{N},\ \forall t_1,\dots,t_k \in T\ \ (t_i \neq t_j \text{ for } i \neq j).$$

Before we state this theorem, we take a closer look at the finite-dimensional distributions of a given stochastic process $(X_t)_{t\in T}$. It is obvious that, for each $k \in \mathbb{N}$ and arbitrary distinct $t_1,\dots,t_k \in T$, the following properties are fulfilled:

1) For all permutations $\pi$ of $\{1,\dots,k\}$ and all $B_1,\dots,B_k \in \mathcal{B}$,

$$P^{X_{t_{\pi(1)}},\dots,X_{t_{\pi(k)}}}\big(B_{\pi(1)} \times \cdots \times B_{\pi(k)}\big) = P^{X_{t_1},\dots,X_{t_k}}\big(B_1 \times \cdots \times B_k\big).$$

2) For $k \geq 2$ and all $B_1,\dots,B_{k-1} \in \mathcal{B}$,

$$P^{X_{t_1},\dots,X_{t_k}}\big(B_1 \times \cdots \times B_{k-1} \times \mathbb{R}\big) = P^{X_{t_1},\dots,X_{t_{k-1}}}\big(B_1 \times \cdots \times B_{k-1}\big).$$

While the validity of these two properties is clear, it is far less obvious that some sort of **converse statement** holds true.

**Theorem 1.1.2. (Daniell-Kolmogorov existence theorem)**
*Let $T$ be a non-empty set. For each $k \in \mathbb{N}$ and distinct $t_1,\dots,t_k \in T$, let $\mu_{t_1,\dots,t_k}$ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Suppose that these probability measures satisfy the following* **consistency conditions***:*

*1) For all permutations $\pi$ of $\{1,\dots,k\}$ and all $B_1,\dots,B_k \in \mathcal{B}$,*

$$\mu_{t_{\pi(1)},\dots,t_{\pi(k)}}\big(B_{\pi(1)} \times \cdots \times B_{\pi(k)}\big) = \mu_{t_1,\dots,t_k}\big(B_1 \times \cdots \times B_k\big). \tag{1.1.4}$$

*2) For $k \geq 2$ and all $B_1,\dots,B_{k-1} \in \mathcal{B}$,*

$$\mu_{t_1,\dots,t_k}\big(B_1 \times \cdots \times B_{k-1} \times \mathbb{R}\big) = \mu_{t_1,\dots,t_{k-1}}\big(B_1 \times \cdots \times B_{k-1}\big). \tag{1.1.5}$$

*Then there exist a probability space $\big(\Omega, \mathcal{F}, P\big)$ and a stochastic process $\mathbf{X} = (X_t)_{t\in T}$ on $\big(\Omega, \mathcal{F}, P\big)$ such that, for each $k \in \mathbb{N}$ and distinct $t_1,\dots,t_k \in T$,*

$$P^{X_{t_1},\dots,X_{t_k}} = \nu_{t_1,\dots,t_k}, \tag{1.1.6}$$

*i.e. the process $\mathbf{X}$ has $\mu_{t_1,\dots,t_k}$ as its finite-dimensional distribution relative to $t_1,\dots,t_k$.*

Before we turn to a proof of this theorem we recall two well-known results from measure theory and prove two auxiliary lemmas. The first of these theorems from measure theory, named after the Greek mathematician Constantin Carathéodory, is one of the main tools for the construction of measures. It states that a non-negative and $\sigma$-additive set function $\mu$ on a ring $\mathcal{R}$ can be extended to a measure on the $\sigma$-algebra generated by $\mathcal{R}$.

**Theorem 1.1.3. (Carathéodory's extension theorem)**
*Suppose that $\Omega$ is a non-empty set and that $\mathcal{R}$ is a collection of subsets of $\Omega$ such that*

- $\emptyset \in \mathcal{R}$,
- *if* $A, B \in \mathcal{R}$, *then* $A \setminus B \in \mathcal{R}$,
- *if* $A, B \in \mathcal{R}$, *then* $A \cup B \in \mathcal{R}$.

*($\mathcal{R}$ is a so-called* **ring** *(of sets) on $\Omega$.)*
*Suppose further that $\mu_0\colon \mathcal{R} \to [0, \infty]$ is a set function such that*

- $\mu_0(\emptyset) = 0$,
- *if* $A_1, A_2, \ldots \in \mathcal{R}$ *are pairwise disjoint sets such that* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$, *then*

$$\mu_0\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu_0\big(A_n\big).$$

*($\mu_0$ is a* **pre-measure** *on $\mathcal{R}$.)*
*Then there exists a measure $\mu$ on $\sigma(\mathcal{R})$ such that*

$$\mu\big(A\big) = \mu_0\big(A\big) \qquad \forall A \in \mathcal{R}.$$

In other words, any pre-measure $\mu_0$ on a ring $\mathcal{R}$ can be extended to a measure $\mu$ on the $\sigma$-algebra $\sigma(\mathcal{R})$ generated by $\mathcal{R}$. Note in passing that any ring $\mathcal{R}$ of sets is intersection-stable, if $A, B \in \mathcal{R}$, then $A \cap B \in \mathcal{R}$. Indeed, this follows from $A \cap B = A \setminus B^c = A \setminus (A \cap B^c) = A \setminus (A \setminus B)$.

Carathéodory's extension theorem will be complemented by the Uniqueness theorem which ensures that such an extension is unique if there exists a sequence $(E_n)_{n\in\mathbb{N}}$ of sets that belong to $\mathcal{R}$ such that $\bigcup_{n=1}^{\infty} E_n = \Omega$ and $\mu_0\big(E_n\big) < \infty$.

**Theorem 1.1.4. (Uniqueness theorem)**
*Suppose that $\Omega$ is a non-empty set and that $\mathcal{E}$ is an intersection-stable collection of subsets of $\Omega$. Let $\mu$ and $\nu$ be measures on $\sigma(\mathcal{E})$ such that*

*(i)* $\mu(E) = \nu(E) \qquad \forall E \in \mathcal{E}$,

*(ii)* *there exist sets* $E_1 \subseteq E_2 \subseteq \ldots$ *that belong to* $\mathcal{E}$, $\bigcup_{n=1}^{\infty} E_n = \Omega$, *and* $\mu(E_n) = \nu(E_n) < \infty \quad \forall n \in \mathbb{N}$.

*Then*

$$\mu(A) = \nu(A) \qquad \forall A \in \sigma\big(\mathcal{E}\big).$$

In particular, a probability measure $P$ on a $\sigma$-algebra $\mathcal{A}$ is completely specified by its values on an intersection-stable collection of sets $\mathcal{E}$ which generates $\mathcal{A}$.

In order to make the proof of our main result in this section, the Daniell-Kolmogorov theorem, transparent, we put some of the technical considerations in two lemmas. Since we make use of Theorems 1.1.3 and 1.1.4 it essentially remains two show that a certain set function which is obviously a content (i.e. finitely additive) is also countably additive. Recall that one typically uses in the simpler case of the construction of Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$ two major arguments: The ring of sets on which the construction is started is given by the collection of finite unions of half-open rectangles. In this case it is obvious that such sets can be approximated from below by closed rectangles such that the content of the difference set is arbitrarily small. And it is easy to see that the intersection of a non-increasing sequence of non-empty closed rectangles is also non-empty. In the present case the situation is less obvious. Instead of the simple Lebesgue measure we have to deal with arbitrary probability measures and since we work in the space $\mathbb{R}^\infty$ we cannot directly use the result for intersections of compact sets mentioned above. The following two lemmas provide corresponding results which are tailor-made for our proof of the Daniell-Kolmogorov theorem.

**Lemma 1.1.5.** *Let $P$ be a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$, and let $A \in \mathcal{B}^d$ and $\epsilon > 0$ be arbitrary. Then there exists a compact (i.e. closed and bounded) set $C$ such that $C \subseteq A$ and $P(A \setminus C) \leq \epsilon$.*

*Proof.* In a first step we show that there exists a *closed* set $F$ such that $F \subseteq A$ and $P(F \setminus A) \leq \varepsilon/2$. To this end, we define a suitable collection of **good sets**,

$$\mathcal{G} \;:=\; \big\{B \in \mathcal{B}^d\colon \text{ for all } \delta > 0 \text{ there exist a closed set } F_\delta \text{ and an open set } U_\delta \\ \text{such that } F_\delta \subseteq B \subseteq U_\delta \text{ and } P(U_\delta \setminus F_\delta) \leq \delta\big\}.$$

It is easy to see that $\mathcal{G}$ is a $\sigma$-algebra on $\mathbb{R}^d$. Indeed, we have that

a) $\mathbb{R}^d \in \mathcal{G}$ since $\mathbb{R}^d$ itself is both closed and open.

b) Let $B \in \mathcal{G}$ be arbitrary, i.e. for each $\delta > 0$ there exist a closed set $F$ and an open set $U$ such that $F \subseteq B \subseteq U$ and $P(U \setminus F) \leq \delta$. Note that the set $U^c$ is, as the complement of an open set, a closed set, and $F^c$ is, as the complement of a closed set, an open set. It holds that $U^c \subseteq B^c \subseteq F^c$ and $P(F^c \setminus U^c) = P(U \setminus F) \leq \delta$. Hence, $B^c$ also belongs to $\mathcal{G}$.

c) Suppose that $B_1, B_2, \ldots$ are sets that belong to $\mathcal{G}$. Then there are open sets $U_1, U_2, \ldots$ such that $B_n \subseteq U_n$ and $P(U_n \setminus B_n) \leq 2^{-(n+1)}\delta$ for all $n \in \mathbb{N}$. The set $U := \bigcup_{n=1}^\infty U_n$ is an open set, $\bigcup_{n=1}^\infty B_n \subseteq U$, and

$$P\Big(U \setminus \bigcup_{n=1}^\infty B_n\Big) \leq \sum_{n=1}^\infty P\Big(U_n \setminus \bigcup_{k=1}^\infty B_k\Big) \leq \sum_{n=1}^\infty P\big(U_n \setminus B_n\big) \leq \delta/2. \tag{1.1.7}$$

Furthermore, there exist closed sets $F_1, F_2, \ldots$ such that $F_n \subseteq B_n$ and $P(B_n \setminus F_n) \leq 2^{-(n+2)}\delta$ for all $n \in \mathbb{N}$. Unfortunately, the set $F := \bigcup_{n=1}^\infty F_n$ is not necessarily a closed set. However, since $P$ is continuous from below and $\bigcup_{n=1}^N F_n \nearrow F$ as $N \to \infty$ there exists some $N_\delta$ such that $P\big(F \setminus \bigcup_{n=1}^{N_\delta} F_n\big) \leq \delta/4$. The set $\widetilde{F} := \bigcup_{n=1}^{N_\delta} F_n$ is, as a **finite** union of closed sets, a closed set, and it holds that $\widetilde{F} \subseteq \bigcup_{n=1}^{\infty} B_n$ as well as

$$P\Big(\bigcup_{n=1}^{\infty} B_n \setminus \widetilde{F}\Big) = P\Big(\bigcup_{n=1}^{\infty} B_n \setminus F\Big) + P\Big(F \setminus \widetilde{F}\Big) \le \delta/4 + \delta/4 = \delta/2. \tag{1.1.8}$$

To summarize, we have that $\widetilde{F} \subseteq \bigcup_{n=1}^{\infty} B_n \subseteq U$ and it follows from (1.1.7) and (1.1.8) that $P(U \setminus \widetilde{F}) \le \delta$. Hence, $B_1, B_2, \ldots \in \mathcal{G}$ implies that $\bigcup_{n=1}^{\infty} B_n \in \mathcal{G}$.

It follows from a) to c) that $\mathcal{G}$ is a $\sigma$-algebra on $\mathbb{R}^d$.

Next we show that $\mathcal{G}$ contains all half-open rectangles. Indeed, for an arbitrary half-open rectangle $(a, b] := (a_1, b_1] \times \cdots \times (a_d, b_d]$ such that $-\infty < a_i \le b_i < \infty$ for all $i = 1, \ldots, d$, we define $F_n := [a_1 + 1/n, b_1] \times \cdots \times [a_d + 1/n, b_d]$ and $U_n := (a_1, b_1 + 1/n) \times \cdots \times (a_d, b_d + 1/n)$. Since $F_n \nearrow (a, b]$ it follows from continuity from below that $P\big(F_n\big) \nearrow P\big((a, b]\big)$. Similarly, since $U_n \searrow (a, b]$ we obtain from continuity from above that $P\big(U_n\big) \searrow P\big((a, b]\big)$. Hence, there exists for each $\delta > 0$ some sufficiently large $N = N(\delta)$ such that, besides $F_N \subseteq (a, b] \subseteq U_N$, $P\big(U_N \setminus F_N\big) = P\big(U_N \setminus (a, b]\big) + P\big((a, b] \setminus F_N\big) \le \delta$.

Now we can complete the proof of the first step in the usual way. Let $\mathcal{I}^d := \big\{(a, b] \colon -\infty < a_i \le b_i < \infty \quad \forall i = 1, \ldots, d\big\}$ denote the collection of all half-open rectangles. Since $\mathcal{I}^d \subseteq \mathcal{G}$ we obtain that $\sigma\big(\mathcal{I}^d\big) \subseteq \sigma\big(\mathcal{G}\big)$. It is well-known that $\sigma\big(\mathcal{I}^d\big) = \mathcal{B}^d$. Furthermore, since $\mathcal{G}$ is a $\sigma$-algebra we have that $\sigma\big(\mathcal{G}\big) = \mathcal{G}$. This implies in particular that there exists a closed set $F$ such that $F \subseteq A$ and $P(F \setminus A) \le \varepsilon/2$.

The sets $F_n := F \cap ([-n, n] \times \cdots \times [-n, n])$ are closed and bounded, hence compact sets. Since $F_n \nearrow F$ we obtain that $P\big(F_n\big) \nearrow P\big(F\big)$. Therefore, there exists some $N < \infty$ such that $P\big(F \setminus F_N\big) \le \varepsilon/2$ and we have, besides $F_N \subseteq A$, $P\big(A \setminus F_N\big) = P\big(A \setminus F\big) + P\big(F \setminus F_N\big) \le \varepsilon$. $\square$

Before we turn to the next auxiliary result we note that, for an arbitrary sequence $(D_n)_{n \in \mathbb{N}}$ of non-empty compact sets of $\mathbb{R}^d$ such that $D_{n+1} \subseteq D_n$ holds for all $n \in \mathbb{N}$, the intersection of these sets is also nonempty. To see this, assume the contrary, i.e. $\bigcap_{n=1}^{\infty} D_n = \emptyset$. Then $D_1 \subseteq \Big(\bigcap_{n=2}^{\infty} D_n\Big)^c = \bigcup_{n=2}^{\infty} D_n^c$. The sets $D_2^c, D_3^c, \ldots$ are, as complements of compact sets, open sets and they cover the compact set $D_1$. We can find a finite subcover, e.g. $D_2^c, \ldots, D_N^c$, i.e. $D_1 \subseteq \big(\bigcap_{n=2}^{N} D_n\big)^c$ which implies that $D_N = \bigcup_{n=1}^{N} D_n = \emptyset$. This, however, contradicts our hypothesis that all sets $D_1, D_2, \ldots$ are non-empty. The following lemma provides a corresponding result in the infinite-dimensional case.

**Lemma 1.1.6.** *Suppose that $(D_n)_{n \in \mathbb{N}}$ is a sequence of non-empty compact sets, $D_n \subseteq \mathbb{R}^n$, such that*

$$D_{n+1} \subseteq D_n \times \mathbb{R} \qquad \forall n \in \mathbb{N}.$$

*Then there exists a sequence $(x_n)_{n \in \mathbb{N}}$ such that*

$$(x_1, x_2, \ldots) \in \bigcap_{n=1}^{\infty} \big(D_n \times \mathbb{R}^{\infty}\big).$$

*Proof.* In order to find an appropriate sequence, we first direct our focus on single components. To start with, we pick for each $n \in \mathbb{N}$ an arbitrary $\big(x_1^{(n)}, \ldots, x_n^{(n)}\big) \in D_n$.

Since $D_1$ is compact there exists a subsequence $\big(n_k^{(1)}\big)_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that $\big(x_1^{(n_k^{(1)})}\big)_{k\in\mathbb{N}}$ converges to some $x_1^{[1]} \in D_1$. The sequence $\big(n_k^{(1)}\big)_{k\in\mathbb{N}}$ contains a further subsequence $\big(n_k^{(2)}\big)_{k\in\mathbb{N}}$ such that $\big((x_1^{(n_k^{(2)})}, x_2^{(n_k^{(2)})})\big)_{k\in\mathbb{N}}$ converges to some limit, say $(x_1^{[2]}, x_2^{[2]}) \in D_2$. This procedure will be continued as follows. Suppose that the $(m-1)$th subsequence $\big(n_k^{(m-1)}\big)_{k\in\mathbb{N}}$ has already been chosen. Then there exists a further subsequence $\big(n_k^{(m)}\big)_{k\in\mathbb{N}}$ such that $\big((x_1^{(n_k^{(m)})}, \ldots, x_m^{(n_k^{(m)})})\big)_{k\in\mathbb{N}}$ converges to some limit $(x_1^{[m]}, \ldots, x_m^{[m]}) \in D_m$. Since this construction is based on subsequences we obtain that $x_m^{[m]} = x_m^{[m+1]} = x_m^{[m+2]} = \ldots$. The sought sequence $(x_n)_{n\in\mathbb{N}}$ is given by $x_n := x_n^{[n]}$. Indeed, we have that $\big(x_1, \ldots, x_n\big) \in D_n$ for all $n \in \mathbb{N}$, which implies that

$$\big(x_1, x_2, \ldots\big) \in \bigcap_{n=1}^{\infty} \big(D_n \times \mathbb{R}^\infty\big).$$

$\square$

After these preparatory considerations we are in a position to prove the Daniell-Kolmogorov theorem.

*Proof of Theorem 1.1.2.* We have to find a suitable probability space $\big(\Omega, \mathcal{F}, P\big)$ and to define a stochastic process $\mathbf{X} = (X_t)_{t\in T}$ on this space such that (1.1.6) is fulfilled. To simplify notation and in order not to obscure the main ideas by too many details of minor importance, we consider the simple case where $T = \mathbb{N}$. We choose

$$\begin{aligned} \Omega &= \mathbb{R}^\infty = \big\{(\omega_1, \omega_2, \ldots)\colon\ x_t \in \mathbb{R}\big\}, \\ \mathcal{F} &= \sigma(\mathcal{C}), \end{aligned}$$

and define, for each $t \in T$,

$$X_t(\omega) = \omega_t.$$

It follows from the construction that all mappings $X_t$ are $(\mathcal{F} - \mathcal{B})$-measurable. Indeed, we have for $B \in \mathcal{B}$

$$X_t^{-1}\big(B\big) = \big\{\omega\colon\ X_t(\omega) \in B\big\} = \big\{\omega\colon\ (\omega_1, \ldots, \omega_t) \in \mathbb{R}^{t-1} \times B\big\} \subseteq \mathcal{C}_t.$$

To some extent, the choice of the probability measure $P$ is now canonical. In order not to violate condition (1.1.6), we have to choose $P$ on the collection of cylinder sets $\mathcal{C}$ such that, for each $k \in \mathbb{N}$ and arbitrary $B \in \mathcal{B}^k$,

$$P\big(B \times \mathbb{R}^\infty\big) = P\Big(\big\{\omega \in \mathbb{R}^\infty\colon\ \underbrace{(\omega_1, \ldots, \omega_k)}_{=(X_1(\omega), \ldots, X_k(\omega))} \in B\big\}\Big) = \mu_{1,\ldots,k}(B). \tag{1.1.9}$$

As a starting point, we define the set function $P_0\colon \mathcal{C} \to [0,1]$ as

$$P_0\big(B \times \mathbb{R}^\infty\big) := \mu_{1,\ldots,k}(B). \tag{1.1.10}$$

We obtain from the second consistency condition (1.1.5) that such a definition does not lead to a contradiction. Note that it is easy to see that the family of cylinder sets $\mathcal{C}$ is an algebra in $\Omega$ but **not** a $\sigma$-algebra. Therefore, it remains to extend the definition of $P_0$ to a probability measure $P$ on a suitable $\sigma$-algebra. At first sight, such an extension may seem to be out of reach since e.g. the $\sigma$-algebra $\sigma(\mathcal{C})$ also contains sets with a very involved structure. Fortunately, an **explicit** definition of all probabilities is not necessary.

It can be shown that $P_0$ is a so-called **content** (a finitely additive set function) on the algebra $\mathcal{C}$. We will show that $P_0$ is even a **pre-measure** (a $\sigma$-additive set function) on $\mathcal{C}$. Suppose that $A_1, A_2, \dots$ are pairwise disjoint sets that belong to $\mathcal{C}$, and that $A := \bigcup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{C}$. We will show that

$$P_0\Big(\bigcup_{n=1}^{\infty} A_n\Big) \;=\; \sum_{n=1}^{\infty} P_0\big(A_n\big). \tag{1.1.11}$$

Since $P_0$ is a content and therefore finitely additive we have

$$\begin{aligned} P_0\Big(\bigcup_{n=1}^{\infty} A_n\Big) &= P_0\Big(A \setminus \bigcup_{k=1}^{n} A_k\Big) + P_0\Big(\bigcup_{k=1}^{n} A_k\Big) \\ &= P_0\big(B_n\big) + \sum_{k=1}^{n} P_0\big(A_k\big), \end{aligned}$$

where $B_n := A \setminus \bigcup_{k=1}^{n} A_k$. Since $\sum_{k=1}^{n} P_0\big(A_k\big) \underset{n\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} P_0\big(A_k\big)$ it remains to show that

$$P_0\big(B_n\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{1.1.12}$$

Since $B_{n+1} \subseteq B_n$ for all $n$ the sequence $\big(P_0(B_n)\big)_{n\in\mathbb{N}}$ is non-increasing and therefore converges. Let us assume that (1.1.12) is not true, i.e. there exists some $\epsilon > 0$ such that

$$P_0\big(B_n\big) \;\geq\; \epsilon \qquad \forall n \in \mathbb{N}. \tag{1.1.13}$$

We shall prove that in that case

$$\bigcap_{n=1}^{\infty} B_n \neq \emptyset,$$

which is obviously wrong since $\bigcup_{k=1}^{n} A_k \nearrow A$, and so $B_n \searrow \emptyset$. Suppose for simplicity of notation (otherwise we can add sets in the sequence of sets $B_n$) that $B_n = C_n \times \mathbb{R}^{\infty}$, for some $C_n \in \mathcal{B}^n$. It follows from Lemma 1.1.5 that there exist compact sets $C_n^* \subseteq C_n$ such that

$$P_0\Big(\big(C_n \times \mathbb{R}^{\infty}\big) \setminus \big(C_n^* \times \mathbb{R}^{\infty}\big)\Big) \;=\; \mu_{1,\dots,n}\big(C_n \setminus C_n^*\big) \;\leq\; \epsilon\, 2^{-n}.$$

Let

$$D_n \;:=\; \big(C_1^* \times \mathbb{R}^{n-1}\big) \cap \dots \cap \big(C_{n-1}^* \times \mathbb{R}^1\big) \cap C_n^*.$$

It follows that $D_n \subseteq C_n$ and $D_{n+1} \subseteq D_n \times \mathbb{R}$. Moreover,

$$\begin{aligned} \mu_{1,\dots,n}\big(D_n\big) &= \mu_{1,\dots,n}\big(C_n\big) - \mu_{1,\dots,n}\big(C_n \setminus D_n\big) \\ &\geq \mu_{1,\dots,n}\big(C_n\big) - \mu_{1,\dots,n}\Big(C_n \cap \big((C_1^* \times \mathbb{R}^{n-1}) \cap \dots \cap (C_{n-1}^* \times \mathbb{R}^1) \cap C_n^*\big)^c\Big) \\ &\geq \mu_{1,\dots,n}\big(C_n\big) - \mu_{1,\dots,n}\big(C_n \setminus (C_1^* \times \mathbb{R}^{n-1})\big) - \dots - \mu_{1,\dots,n}\big(C_n \setminus (C_{n-1}^* \times \mathbb{R}^1)\big) \\ &\qquad - \mu_{1,\dots,n}\big(C_n \setminus C_n^*\big) \\ &\geq \mu_{1,\dots,n}\big(C_n\big) - \mu_1\big(C_1 \setminus C_1^*\big) - \dots - \mu_{1,\dots,n-1}\big(C_{n-1} \setminus C_{n-1}^*\big) - \mu_{1,\dots,n}\big(C_n \setminus C_n^*\big) \\ &\geq \epsilon - \epsilon\big(2^{-1} + \dots + 2^{-n}\big) > 0. \end{aligned}$$

Hence, $D_n \neq \emptyset$ and it follows from Lemma 1.1.6 that $\bigcap_{n=1}^{\infty} D_n \times \mathbb{R}^{\infty} \neq \emptyset$. Since $\bigcap_{n=1}^{\infty} D_n \times \mathbb{R}^{\infty} \subseteq \bigcap_{n=1}^{\infty} B_n$ we get a contradiction and we conclude that (1.1.12) holds true. Hence, $P_0$ is a pre-measure on $\mathcal{C}$. At this point, we can simply use Caratheodory's extension theorem (Theorem 1.1.3) and the Uniqueness theorem (Theorem 1.1.4) to conclude that there exists a **unique extension** of $P_0$ to a probability measure $P$ on the $\sigma$-algebra $\sigma(\mathcal{C})$.

It follows from (1.1.10) for each $k \in \mathbb{N}$ and distinct $t_1, \ldots, t_k \in T$, that

$$P^{X_{t_1}, \ldots, X_{t_k}} = \mu_{t_1, \ldots, t_k},$$

as required. □

Many classical theorems in probability require that there exists a sequence $(X_t)_{t \in \mathbb{N}}$ of independent and identically distributed random variables satisfying appropriate regularity conditions. One such example is the strong law of large numbers which states that with probability 1 the sequence of sample means $\bar{X}_n$ converges to $EX_t$, as $n$ tends to infinity. As usual in mathematics, "existence" means that one can construct a corresponding model free of contradiction. To obtain such a model for a sequence $(X_t)_{t \in \mathbb{N}}$ of independent random variables following a common distribution $Q$, choose a family of distributions as

$$\mu_{t_1, \ldots, t_k} = \underbrace{Q \otimes \cdots \otimes Q}_{k \text{ times}}, \qquad t_1, \ldots, t_k \in \mathbb{N},\ t_i \neq t_j \text{ for } i \neq j.$$

This family satisfies the consistency conditions (1.1.5) and (1.1.6). According to the proof of Theorem 1.1.2, we can choose a probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = \mathbb{R}^{\infty}$, $\mathcal{F} = \sigma(\mathcal{C})$, and $P$ such that $P(\{\omega\colon (\omega_{t_1, \ldots, t_k}) \in C\}) = \mu_{t_1, \ldots, t_k}(C)$ for $t_i \neq t_j$ in case of $i \neq j$ and $C \in \mathcal{B}^k$. Then the random variables $X_t$ given as $X_t(\omega) = \omega_t$ are independent, $X_t \sim Q$.

In cases where $T = \mathbb{N}$, it is convenient to use the order structure of $\mathbb{N}$ and take the $\mu_{t_1, \ldots, t_k}$ to be specified initially only for $k$-tuples $(t_1, \ldots, t_k) = (1, 2, \ldots, k)$. Using the consistency conditions (1.1.5) and (1.1.6) this completely specifies the finite-dimensional distributions $\mu_{t_1, \ldots, t_k}$ for **all** $k$-tuples of distinct points of $\mathbb{N}$.

## 1.2 Stationarity

Typical problems in the statistical analysis of time series are the estimation of model parameters or the prediction of future values. Suppose that we observe realizations $x_1, \ldots, x_n$ (our data) of the random variables $X_1, \ldots, X_n$, where $(X_t)_{t\in\mathbb{N}}$ is our time series. The prediction of future values $X_{n+1}, X_{n+2}, \ldots$ can only be successful if the $X_t$ are not completely unrelated to each other. Consistency of (sequences of) estimators for certain parameters will be possible if we obtain an increasing amount of new information about the underlying situation as the sample size $n$ tends to infinity. This is usually the case if the dependence between the observed random variables is not too strong and if the parameters do not change over time. The latter requirement leads to the important concept of stationarity of a process. Stationarity roughly means that the properties of the process do not change as time proceeds. An exact definition of such a notion will be given on the basis of the finite-dimensional distributions of a process.

**Definition.** Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a (real-valued) stochastic process on $(\Omega, \mathcal{F}, P)$.

(i) $\mathbf{X}$ is said to be **strictly stationary** if

$$P^{X_{t_1},\ldots,X_{t_k}} = P^{X_{t_1+t},\ldots,X_{t_k+t}} \qquad \forall t_1, \ldots, t_k, t \in \mathbb{Z},\ \forall k \in \mathbb{N}.$$

(ii) $\mathbf{X}$ is said to be **stationary** (**weakly stationary**) if

a) $EX_t^2 < \infty \quad \forall t \in \mathbb{Z}$,

b) $EX_t = \mu \quad \forall t \in \mathbb{Z}$ and some $\mu \in \mathbb{R}$,

c) $\mathrm{cov}(X_r, X_s) = \mathrm{cov}(X_{r+t}, X_{s+t}) \quad \forall r, s, t \in \mathbb{Z}$.

In this case, $\gamma_X(k) := \mathrm{cov}(X_{t+k}, X_t)$ is the **autocovariance** at lag $k$, $\gamma_X\colon\ \mathbb{Z} \to \mathbb{R}$ is the **autocovariance function**.

If $T = \mathbb{N}_0$ or $T = \mathbb{N}$, then the above definition has to be adapted accordingly.

As already mentioned, a process $(X_t)_{t\in\mathbb{N}}$ with the property of strict stationarity as defined above has also a shift-invariant behavior when infinite stretches are considered. Indeed, since the finite-dimensional distribution determine probabilities such as $P\big(\{\omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\}\big)$ for $C \in \sigma(\mathcal{C})$, we conclude that

$$P\big(\{\omega\colon\ (X_1(\omega), X_2(\omega), \ldots) \in C\}\big) = P\big(\{\omega\colon\ (X_{k+1}(\omega), X_{k+2}(\omega), \ldots) \in C\}\big) \quad \forall k \in \mathbb{N},\ \forall C \in \sigma(\mathcal{C}).$$

Therefore the above definition of strict stationarity will be sufficient for (almost?) all purposes. The following proposition clarifies the relation between strict and weak stationarity.

**Proposition 1.2.1.** *Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a real-valued process on $(\Omega, \mathcal{F}, P)$.*

*(i) If $\mathbf{X}$ is strictly stationary and $EX_0^2 < \infty$, then $\mathbf{X}$ is also weakly stationary.*

*(ii) The converse statement is not true in general.*

*Proof.* (i) First of all, since $EX_0^2 < \infty$ the expectation of $X_0$ exists and is finite. Therefore we obtain from $P^{X_t} = P^{X_{t+s}}$ that $EX_t = EX_{t+s}\ \forall t, s \in \mathbb{Z}$. Furthermore, $\mathrm{cov}(X_r, X_s)$ also exists and it follows from $P^{X_r,X_s} = P^{X_{r+t},X_{s+t}}$ that $\mathrm{cov}(X_r, X_s) = \mathrm{cov}(X_{r+t}, X_{s+t})$ for all $r, s, t \in \mathbb{Z}$.
(ii) Here is a counter-example: Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a (two-sided) sequence of independent random variables, where

$$X_t \sim \begin{cases} \text{Poisson}(\lambda), & \text{if } t \text{ is odd}, \\ N(\lambda, \lambda), & \text{if } t \text{ is even} \end{cases}.$$

Then $EX_t = \lambda\ \forall t \in \mathbb{Z}$ and

$$\mathrm{cov}(X_t, X_s) = \begin{cases} \lambda, & \text{if } t = s, \\ 0, & \text{if } t \neq s \end{cases}.$$

Hence, $\mathbf{X}$ is weakly but not strictly stationary. □

An important special case is that of a Gaussian process.

**Definition.** $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a **Gaussian process** on $(\Omega, \mathcal{F}, P)$ if all finite-dimensional distributions are Gaussian.

**Lemma 1.2.2.** *If $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a Gaussian process on a probability space $(\Omega, \mathcal{F}, P)$, then the following two statements are equivalent.*

*a) $\mathbf{X}$ is strictly stationary,*

*b) $\mathbf{X}$ is weakly stationary.*

*Proof.* [a) $\Longrightarrow$ b)] Gaussianity implies that $EX_0^2 < \infty$. Therefore, b) follows from (i) of Proposition 1.2.1.
[b) $\Longrightarrow$ a)] Suppose that $\mathbf{X}$ is weakly stationary. Let $k \in \mathbb{N}$ and $t_1, \ldots, t_k, t \in \mathbb{Z}$ be arbitrary. We have to show that

$$P^{X_{t_1},\ldots,X_{t_k}} = P^{X_{t_1+t},\ldots,X_{t_k+t}}. \tag{1.2.1}$$

We have that

$$P^{X_{t_1},\ldots,X_{t_k}} \sim N(\mu, \Sigma),$$

where

$$\mu = \left(EX_{t_1}, \ldots, EX_{t_k}\right)^T \quad \text{and} \quad \Sigma = \begin{pmatrix} \mathrm{cov}(X_{t_1}, X_{t_1}) & \ldots & \mathrm{cov}(X_{t_1}, X_{t_k}) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X_{t_k}, X_{t_1}) & \ldots & \mathrm{cov}(X_{t_k}, X_{t_k}) \end{pmatrix}.$$

Analogously,

$$P^{X_{t_1+t},\ldots,X_{t_k+t}} \sim N(\widetilde{\mu}, \widetilde{\Sigma}).$$

We obtain from weak stationarity that

$$\widetilde{\mu} = \left(EX_{t_1+t}, \ldots, EX_{t_k+t}\right)^T = \mu$$

and $\operatorname{cov}(X_{t_i+t}, X_{t_j+t}) = \operatorname{cov}(X_{t_i}, X_{t_j})$, which implies that

$$\widetilde{\Sigma} = \Sigma.$$

Therefore, (1.2.1) is satisfied which means that $\mathbf{X}$ is strictly stationary. □

## Examples

1) Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables. This process is strictly stationary. If in addition $E[\varepsilon_t^2] < \infty$, then it is also weakly stationary.

2) Let $Y$ and $Z$ be uncorrelated random variables such that $EY = EZ = 0$ and $EY^2 = EZ^2 = 1$. We define, for any $\theta \in [-\pi, \pi]$,

$$X_t := Y \cos(\theta t) + Z \sin(\theta t).$$

Then

$$EX_t = 0 \qquad \forall t$$

and

$$\begin{aligned} \operatorname{cov}(X_{t+r}, X_t) &= E\left[\{Y \cos(\theta(t+r)) + Z \sin(\theta(t+r))\}\{Y \cos(\theta t) + Z \sin(\theta t)\}\right] \\ &= \cos(\theta(t+r)) \cos(\theta t) + \sin(\theta(t+r)) \sin(\theta t) \\ &= \cos(\theta r). \end{aligned}$$

The latter equation follows from one of the trigonometric identities. We have, on the one hand,

$$e^{i(u-v)} = \cos(u-v) + i \sin(u-v),$$

and, on the other hand,

$$\begin{aligned} e^{i(u-v)} &= e^{iu} e^{-iv} = \big(\cos(u) + i \sin(u)\big)\big(\cos(v) - i \sin(v)\big) \\ &= \cos(u) \cos(v) + \sin(u) \sin(v) + i \big(\sin(u) \cos(v) - \cos(u) \sin(v)\big). \end{aligned}$$

Therefore, the autocovariances are also shift-invariant which means that the process $(X_t)_{t\in\mathbb{Z}}$ is weakly stationary.

3) Let $X_0, \varepsilon_1, \varepsilon_2, \ldots$ be independent, $X_0 \sim N(0, \sigma_X^2)$, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ $\forall t \in \mathbb{N}$.

We obtain a so-called **autoregressive process** $(X_t)_{t\in\mathbb{N}_0}$ by defining recursively

$$X_t := \alpha X_{t-1} + \varepsilon_t \quad \forall t \in \mathbb{N},$$

where $\alpha$ is a real constant, $|\alpha| < 1$.

**Question:** Is it possible to choose $\sigma_X^2$ such that the process $(X_t)_{t\in\mathbb{N}_0}$ is stationary?

Since

$$X_1 \,=\, \alpha\, X_0 \,+\, \varepsilon_1 \,\sim\, N(0, \alpha^2\, \sigma_X^2 + \sigma_\varepsilon^2)$$

a necessary condition for any kind of stationarity is that $\mathrm{var}(X_0) = \sigma_X^2 = \mathrm{var}(X_1) = \alpha^2\sigma_X^2 + \sigma_\varepsilon^2$, i.e. $\sigma_X^2 = \sigma_\varepsilon^2/(1-\alpha^2)$.

Suppose that $\sigma_X^2 = \sigma_\varepsilon^2/(1-\alpha^2)$. We have that

$$X_t \,=\, \varepsilon_t \,+\, \alpha\, X_{t-1} \,=\, \ldots \,=\, \varepsilon_t \,+\, \alpha\, \varepsilon_{t-1} \,+\, \ldots \,+\, \alpha^{t-1}\, \varepsilon_1 \,+\, \alpha^t\, X_0.$$

Since $(X_0, X_1, \ldots, X_t)^T = M_t\,(X_0, \varepsilon_1, \ldots, \varepsilon_t)^T$, for some $(t+1)\times(t+1)$-matrix $M_t$ we see that $(X_0, X_1, \ldots, X_t)^T$ has a multivariate normal distribution, and so $(X_t)_{t\in\mathbb{N}_0}$ is a zero mean Gaussian process. We obtain, as above, that

$$\mathrm{var}(X_0) \,=\, \mathrm{var}(X_1) \,=\, \ldots \,=\, \mathrm{var}(X_t) \quad \forall t \in \mathbb{N}.$$

Furthermore, for $k \in \mathbb{N}$, we obtain from independence of $X_t, \varepsilon_{t+1}, \ldots \varepsilon_{t+k}$ that

$$\mathrm{cov}(X_{t+k}, X_t) \,=\, \mathrm{cov}\left(\varepsilon_{t+k} + \alpha\varepsilon_{t+k-1} + \cdots + \alpha^{k-1}\varepsilon_{t+1} + \alpha^k X_t,\; X_t\right) \,=\, \alpha^k\, \sigma_X^2.$$

Hence, $(X_t)_{t\in\mathbb{N}_0}$ is a weakly stationary process. Moreover, as a Gaussian process, it is also strictly stationary.

The following lemma contains a few elementary properties of the autocovariance function of a stationary process.

**Lemma 1.2.3.** *Let $\gamma$ be the autocovariance function of a real-valued stationary process $(X_t)_{t\in\mathbb{Z}}$. Then*

*(i)* $\gamma(0) \,\geq\, 0$,

*(ii)* $|\gamma(r)| \,\leq\, \gamma(0) \quad \forall r \in \mathbb{Z}$,

*(iii)* $\gamma(r) \,=\, \gamma(-r) \quad \forall r \in \mathbb{Z}$.

*Proof.* (i) is a statement of the obvious fact that

$$\gamma(0) \,=\, \mathrm{var}(X_t) \,\geq\, 0.$$

(ii) is an immediate consequence of the Cauchy-Schwarz (Cauchy-Bunyakovsky-Schwarz) inequality,

$$|\gamma(r)| \,=\, |\,\mathrm{cov}(X_{t+r}, X_t)| \,\leq\, \sqrt{\mathrm{var}(X_{t+r})}\sqrt{\mathrm{var}(X_t)} \,=\, \gamma(0).$$

Finally, (iii) follows from

$$\gamma(r) \,=\, \mathrm{cov}(X_{t+r}, X_t) \,=\, \mathrm{cov}(X_t, X_{t+r}) \,=\, \gamma(-r).$$

□

**Remark 1.2.4.** *If* $(X_t)_{t\in\mathbb{Z}}$ *is a complex-valued process and* $E[|X_t|^2] < \infty\ \forall t$*, then*

$$\operatorname{cov}(X_{t+r}, X_t) := E\big[(X_{t+r} - EX_{t+r})\,\overline{(X_t - EX_t)}\big],$$

*where* $\overline{z}$ *denotes the complex conjugate of a complex number* $z$*. Since, for a complex-valued random variable* $Y$*,* $|EY| \le E|Y|$[1]*, an analogue of the Cauchy-Schwarz inequality holds true:*

$$\big|E[Y_{t+r}\overline{Y}_t]\big| \le E\left[|Y_{t+r}\overline{Y}_t|\right] \le \sqrt{E[|Y_{t+r}|^2]}\,\sqrt{E[|Y_t|^2]}.$$

Next we intend to find a characterization of autocovariance functions.

**Definition.** A real-valued function on the integers, $\kappa\colon \mathbb{Z} \to \mathbb{R}$, is said to be **non-negative definite** (positive semidefinite) if

$$\sum_{i,j=1}^{n} a_i\,\kappa(t_i - t_j)\,a_j \ge 0 \qquad \forall n \in \mathbb{N}, \forall a_1, \ldots, a_n \in \mathbb{R}, \forall t_1, \ldots, t_n \in \mathbb{Z}.$$

**Theorem 1.2.5.** *Let* $\gamma\colon\ \mathbb{Z} \to \mathbb{R}$ *be a real-valued function. Then the following statements are equivalent.*

*(i)* $\gamma$ *is the autocovariance function of a real-valued stationary process* $(X_t)_{t\in\mathbb{Z}}$*,*

*(ii)* $\gamma$ *is an* **even** *and* **non-negative definite** *function.*

*Proof.* [(i) $\Longrightarrow$ (ii)] Assume that $\gamma$ is the autocovariance function of a real-valued stationary process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$. Then

$$\gamma(k) = \operatorname{cov}(X_k, X_0) = \operatorname{cov}(X_0, X_k) = \gamma(-k),$$

i.e. $\gamma$ is an even function. Moreover, we have, for arbitrary $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$, $t_1, \ldots, t_n \in \mathbb{Z}$,

$$\begin{aligned}\sum_{i,j=1}^{n} a_i\,\gamma(t_i - t_j)\,a_j &= \sum_{i,j=1}^{n} a_i\ \operatorname{cov}\left(X_{t_i}, X_{t_j}\right)\,a_j \\ &= \sum_{i,j=1}^{n} \operatorname{cov}\left(a_i X_{t_i}, a_j X_{t_j}\right) = \operatorname{var}\Big(\sum_{i=1}^{n} a_i X_{t_i}\Big) \ge 0,\end{aligned}$$

i.e., $\gamma$ is non-negative definite.

[1] Suppose that $E|Y| < \infty$. Then

$$\begin{aligned}|EY|^2 &= EY \cdot \overline{EY} = E\big[Y \cdot \overline{EY}\big] \\ &= \operatorname{Re}\big(E\big[Y \cdot \overline{EY}\big]\big) = E\big[\underbrace{\operatorname{Re}\big(Y \cdot \overline{EY}\big)}_{\le |Y|\cdot|\overline{EY}|}\big] \\ &\le E\big[|Y| \cdot |EY|\big] = E|Y| \cdot |EY|,\end{aligned}$$

which implies that $\big|EY\big| \le E|Y|$.

[(ii) $\implies$ (i)] Let $\gamma$ be an arbitrary even and non-negative definite function. For each $m, n \in \mathbb{Z}$, $m \leq n$, we consider the matrix

$$\Gamma_{m,n} := \begin{pmatrix} \gamma(m-m) & \dots & \gamma(m-n) \\ \vdots & \ddots & \vdots \\ \gamma(n-m) & \dots & \gamma(n-n) \end{pmatrix}.$$

Since $\gamma$ is an even function it follows that $\Gamma_{m,n}$ is a symmetric matrix. Moreover, $\Gamma_{m,n}$ is a non-negative definite matrix. Actually, let $c = (c_m, \dots, c_n)^T \in \mathbb{R}^{n-m+1}$ be arbitrary. Then $c^T \Gamma_{m,n} c = \sum_{i,j=m}^n c_i \gamma(i-j) c_j \geq 0$. Hence, $\Gamma_{m,n}$ has the properties of a covariance matrix. We define

$$P_{m,n} := \mathcal{N}_{n-m+1}(0_{n-m+1}, \Gamma_{m,n}),$$

where $0_k = (0, \dots, 0)^T$ denotes a vector of length $k$ consisting of zeroes. It follows that the family of distributions $(P_{m,n})_{m \leq n}$ satisfies the consistency condition of Kolmogorov's existence theorem. Therefore, there exists a stochastic process $\mathbf{X} = (X_t)_{t \in \mathbb{Z}}$ on a suitable probability space $(\Omega, \mathcal{F}, P)$ such that

$$P^{X_m, \dots, X_n} = P_{m,n}.$$

It follows that the process $(X_t)_{t \in \mathbb{Z}}$ is both weakly and strictly stationary and that

$$\operatorname{cov}(X_{t+k}, X_t) = \gamma(k) \qquad \forall k \in \mathbb{Z},$$

as required. □

## Exercises

**Ex. 1.2.1** Suppose that $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables, $E\varepsilon_0 = 0$, $E\varepsilon_0^2 =: \sigma_\varepsilon^2 < \infty$, and

$$X_t := \varepsilon_t + \beta\, \varepsilon_{t-1}.$$

Is the process $(X_t)_{t \in \mathbb{Z}}$ (weakly) stationary?

**Ex. 1.2.2** Let $(\beta_k)_{k \in \mathbb{Z}}$ be a sequence of real numbers with $\sum_{k=-\infty}^\infty \beta_k^2 < \infty$. The function $\gamma\colon \mathbb{Z} \to \mathbb{R}$ is defined by $\gamma(k) = \sum_{j=-\infty}^\infty \beta_{j+k} \beta_j$.

Is $\gamma$ an autocovariance function?

## 1.3 Hilbert spaces

First of all, the contents of this section seems to be completely out of place in a course on time series. But why do we pay attention to such an abstract subject as a Hilbert space? In what follows we will be faced with the following questions.

(i) As a simple class of models for time series, we consider so-called linear processes. Given an underlying process $(\varepsilon_t)_{t\in\mathbb{Z}}$, we consider a process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$, where

$$X_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}. \qquad (1.3.1)$$

This raises the following questions.

– Does the infinite sum on the right-hand side of (1.3.1) converge? And if so, in which sense?

– What is the covariance structure of $(X_t)_{t\in\mathbb{Z}}$?
Of course, $\operatorname{cov}(\sum_{j=0}^m \beta_j \varepsilon_{s-j}, \sum_{k=0}^m \beta_k \varepsilon_{t-k})$ can be easily computed since we can take out the finite sums. But what about $\operatorname{cov}(\sum_{j=0}^\infty \beta_j \varepsilon_{s-j}, \sum_{k=0}^\infty \beta_k \varepsilon_{t-k})$ where infinite sums are involved?

Answers to these questions can be easily deduced in the general context of Hilbert spaces.

(ii) Suppose that we observe realizations $x_1, \ldots, x_n$ of random variables $X_1, \ldots, X_n$, where $(X_t)_{t\in\mathbb{N}}$ is a stationary process. How can we best predict future values, e.g. $X_{n+1}$?
We will see that a best linear predictor is given by the orthogonal projection of $X_{n+1}$ onto the linear space spanned by $X_1, \ldots, X_n$. This can be reformulated as a projection in an appropriate Hilbert space and its characterization is most conveniently derived in such an abstract context.

**Definition.** Let $\mathcal{H}$ be a complex (real) vector space which is closed under the operations of vector addition (if $x, y \in \mathcal{H}$ then $x + y \in \mathcal{H}$) and complex (real) scalar multiplication (if $x \in \mathcal{H}$ and $\alpha \in \mathbb{C}$ or $\alpha \in \mathbb{R}$ then $\alpha x \in \mathcal{H}$).

For a complex vector space $\mathcal{H}$, a mapping $\langle \cdot, \cdot \rangle \colon \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ is called **inner product** (scalar product) if

(i) $\langle x, x \rangle \geq 0 \quad \forall x \in \mathcal{H}$
$\langle x, x \rangle = 0 \quad \Leftrightarrow \quad x = \mathbf{0}$ ($\mathbf{0}$ denotes the zero element of $\mathcal{H}$.),

(ii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (The bar denotes complex conjugation.),

(iii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \quad \forall x, y \in \mathcal{H}, \forall \alpha \in \mathbb{C}$,

(iv) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \quad \forall x, y, z \in \mathcal{H}$.

The complex number $\langle x, y \rangle$ is called the inner product of $x$ and $y$. A vector space $\mathcal{H}$ equipped with an inner product is called **inner-product space** (scalar space).

Note that it follows from (ii) and (iii) that $\langle x, \alpha y \rangle = \overline{\langle \alpha y, x \rangle} = \overline{\alpha \langle y, x \rangle} = \overline{\alpha}\, \overline{\langle y, x \rangle} = \overline{\alpha} \langle x, y \rangle$.

**Remark 1.3.1.** *A* **real** *vector space $\mathcal{H}$ is an inner product space if for all $x, y \in \mathcal{H}$ there exists a* **real** *number $\langle x, y\rangle$ such that suitably adapted versions of (i) to (iv) are satisfied. (Of course, (ii) obviously reduces to $\langle x, y\rangle = \langle y, x\rangle$ and (iii) has to be satisfied for all real $\alpha$.)*

## Examples

1) Real Euclidean space $\mathbb{R}^d$

$$\langle x, y\rangle \;=\; \sum_{i=1}^{d} x_i\, y_i$$

2) Complex Euclidean space $\mathbb{C}^d$

$$\langle x, y\rangle \;=\; \sum_{i=1}^{d} x_i\, \bar{y}_i$$

**Definition.** Let $\mathcal{H}$ be an inner-product space, $x \in \mathcal{H}$. The **norm** of $x$ is defined to be

$$\|x\| \;:=\; \sqrt{\langle x, x\rangle}.$$

Actually, in order to justify the term norm we still have to prove that $\|\cdot\|$ satisfies all axioms of a norm. In particular, validity of the triangle inequality has to be verified. Before we turn to this point, we state a few auxiliary results.

**Lemma 1.3.2.** *(Cauchy-Bunyakovsky-Schwarz inequality)*
*Let $\mathcal{H}$ be an inner-product space. Then*

*(i)* $\quad |\langle x, y\rangle| \;\le\; \|x\|\,\|y\| \qquad \forall x, y \in \mathcal{H},$

*(ii)* $\quad |\langle x, y\rangle| \;=\; \|x\|\,\|y\|$ $\qquad$ *if and only if* $\qquad \|y\|^2\, x \;=\; \langle x, y\rangle\, y.$

*Proof.* (i) Let $x, y \in \mathcal{H}$ be arbitrary. Using the axioms of an inner product we obtain that

$$\begin{aligned} 0 \;&\le\; \langle \|y\|^2\, x \;-\; \langle x, y\rangle\, y,\; \|y\|^2\, x \;-\; \langle x, y\rangle\, y\rangle \\ &=\; \|x\|^2\, \|y\|^4 \;+\; \langle x, y\rangle\, \overline{\langle x, y\rangle}\, \|y\|^2 \;-\; \|y\|^2\, \langle x, y\rangle\, \overline{\langle x, y\rangle} \;-\; \langle x, y\rangle\, \|y\|^2\, \langle y, x\rangle \\ &=\; \|y\|^2\, \left\{\|x\|^2\, \|y\|^2 \;-\; |\langle x, y\rangle|^2\right\}. \end{aligned}$$

Now we distinguish between two cases.
Case 1: If $\|y\| \neq 0$, then the term in curly braces is non-negative which yields that assertion (i) holds true.
Case 2: If $\|y\| = 0$, then $y = \mathbf{0}$ which implies that $\langle x, y\rangle = 0$. In this case, the term in curly braces is equal to 0.

(ii) $(\Longrightarrow)$
If $|\langle x,y\rangle| = \|x\|\,\|y\|$, then $\{\|x\|^2\,\|y\|^2 - |\langle x,y\rangle|^2\} = 0$, which implies that

$$\langle \|y\|^2\,x \;-\; \langle x,y\rangle\,y,\; \|y\|^2\,x \;-\; \langle x,y\rangle\,y\rangle \;=\; 0$$

and, therefore, $\|y\|^2\,x = \langle x,y\rangle\,y$.
$(\Longleftarrow)$
If $\|y\|^2\,x = \langle x,y\rangle\,y$, then

$$\|y\|^2\;\left\{\|x\|^2\,\|y\|^2 \;-\; |\langle x,y\rangle|^2\right\} \;=\; 0.$$

Case 1: If $\|y\| = 0$, then $|\langle x,y\rangle| = 0 = \|x\|\,\|y\|$.
Case 2: If $\|y\| \neq 0$, then $|\langle x,y\rangle| = \|x\|\,\|y\|$. □

Now we are in a position to verify that $\|\cdot\|$ shares all axioms of a norm.

**Lemma 1.3.3.** *Let $\mathcal{H}$ be an inner-product space and let $\|x\| = \sqrt{\langle x,x\rangle}$. Then*

*(i)* $\|x\| \geq 0$ *and* $\|x\| = 0$ *if and only if* $x = \mathbf{0}$,

*(ii)* $\|\alpha\,x\| = |\alpha|\,\|x\| \quad \forall x \in \mathcal{H}, \forall \alpha \in \mathbb{C}$,

*(iii)* $\|x+y\| \leq \|x\| + \|y\| \quad \forall x,y \in \mathcal{H}$.

*Proof.* (i) is obvious. (ii) follows from $\|\alpha\,x\|^2 = \langle \alpha\,x, \alpha\,x\rangle = \alpha\,\bar{\alpha}\,\langle x,x\rangle = |\alpha|^2\,\|x\|^2$. Finally, we have

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \;=\; \|x\|^2 \;+\; \langle x,y\rangle \;+\; \langle y,x\rangle \;+\; \|y\|^2 \\
&= \|x\|^2 \;+\; \|y\|^2 \;+\; \underbrace{2\,\mathrm{Re}(\langle x,y\rangle)}_{\leq 2\,|\langle x,y\rangle| \,\leq\, 2\,\|x\|\,\|y\|} \\
&\leq (\|x\| \;+\; \|y\|)^2\,.
\end{aligned}$$

□

The following lemma provides an important property, the so-called **continuity of the inner product**. This allows us, among others, to compute autocovariances of a linear process $(X_t)_{t\in\mathbb{Z}}$, where $X_t = \sum_{k=-\infty}^{\infty} \beta_k\,\varepsilon_{t-k}$.

**Lemma 1.3.4.** *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences of elements of an inner-product space $\mathcal{H}$ with $\|x_n - x\| \underset{n\to\infty}{\longrightarrow} 0$ and $\|y_n - y\| \underset{n\to\infty}{\longrightarrow} 0$, for some $x,y \in \mathcal{H}$. Then*

*(i)* $\|x_n\| \underset{n\to\infty}{\longrightarrow} \|x\|$,

*(ii)* $\langle x_n, y_n\rangle \underset{n\to\infty}{\longrightarrow} \langle x,y\rangle$. *("**continuity of the inner product**")*

*Proof.* (i) We obtain from the triangle inequality

$$\|x_n\| = \|x + (x_n - x)\| \le \|x\| + \|x_n - x\|$$

as well as

$$\|x\| = \|x_n + (x - x_n)\| \le \|x_n\| + \|x_n - x\|.$$

Therefore,

$$\big|\|x_n\| - \|x\|\big| \le \|x_n - x\| \underset{n\to\infty}{\longrightarrow} 0.$$

(ii) It follows from linearity of the inner product and by the Cauchy-Schwarz inequality that

$$\begin{aligned} |\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle| \\ &\le \underbrace{\|x_n\|}_{\text{bounded}} \|y_n - y\| + \|x_n - x\| \|y\| \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

□

In the next section of these lecture notes, we consider linear processes $(X_t)_{t\in\mathbb{Z}}$, where $X_t = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$. The following definitions and results will be used to deduce convergence of the infinite series.

**Definition.** A sequence $(x_n)_{n\in\mathbb{N}}$ of elements of the inner-product space $\mathcal{H}$ is said to be a **Cauchy sequence** if

$$\|x_n - x_m\| \longrightarrow 0 \quad \text{as } m, n \longrightarrow \infty,$$

that is, for every $\epsilon > 0$, there exists some $N(\epsilon) \in \mathbb{N}$ such that

$$\|x_n - x_m\| \le \epsilon \quad \forall m, n \ge N(\epsilon).$$

**Definition.** A **Hilbert space** $\mathcal{H}$ is an inner-product space which is **complete**, that is, an inner-product space in which every Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ converges in norm to some element $x \in \mathcal{H}$ ($\|x_n - x\| \to_{n\to\infty} 0$).

**Example** $\mathcal{H} = \mathbb{R}^d$
Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}^d$, i.e. $\|x_n - x_m\| \longrightarrow 0$ as $m, n \longrightarrow \infty$. Since $\|x_n - x_m\|^2 = \sum_{i=1}^d (x_{ni} - x_{mi})^2$ we have that

$$|x_{ni} - x_{mi}| \longrightarrow 0 \quad \text{as } m, n \longrightarrow \infty,$$

i.e. $(x_{ni})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. By completeness of $\mathbb{R}$, there exists some $x_{0i} \in \mathbb{R}$ such that

$$x_{ni} \underset{n\to\infty}{\longrightarrow} x_{0i}.$$

This yields, for $x = (x_{01}, \ldots, x_{0d})^T$,

$$\|x_n - x\| \underset{n\to\infty}{\longrightarrow} 0.$$

## The space $L^2(\Omega, \mathcal{F}, P)$

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We define

$$\begin{aligned} L^2(\Omega, \mathcal{F}, P) \;&:=\; \{X\colon\; X \text{ is a real-valued random variable on } (\Omega, \mathcal{F}, P), \\ &\qquad \int_\Omega X^2(\omega)\, dP(\omega) < \infty\}. \end{aligned}$$

$L^2(\Omega, \mathcal{F}, P)$ is a real vector space. In particular, if $X, Y \in L^2(\Omega, \mathcal{F}, P)$ and $\alpha \in \mathbb{R}$, then $X + Y \in L^2(\Omega, \mathcal{F}, P)$ and $\alpha X \in L^2(\Omega, \mathcal{F}, P)$. Moreover, the axioms of a vector space (commutativity, associativity, ...) are fulfilled. We are going to define an inner product as

$$\langle X, Y \rangle \;:=\; E[X\,Y] \;=\; \int_\Omega X(\omega)\, Y(\omega)\, dP(\omega).$$

(Since $E[|XY|] \leq (EX^2 + EY^2)/2 < \infty$ the inner product of $X$ and $Y$ is well-defined and finite.) Moreover, for $X, Y, Z \in L^2(\Omega, \mathcal{F}, P)$ and $\alpha \in \mathbb{R}$,

$$\begin{aligned} \langle X, X \rangle \;&\geq\; 0, \\ \langle X, Y \rangle \;&=\; \langle Y, X \rangle, \\ \langle \alpha X, Y \rangle \;&=\; \alpha\, \langle X, Y \rangle, \\ \langle X + Y, Z \rangle \;&=\; \langle X, Z \rangle \,+\, \langle Y, Z \rangle. \end{aligned}$$

However, if $\langle X, X \rangle = 0$, then it does not necessarily follow that $X(\omega) = 0$ for all $\omega \in \Omega$. Only $P(X \neq 0) = 0$ follows in general. In view of this, we have to consider **equivalence classes** and we say that random variables $X$ and $Y$ are **equivalent** if $P(X \neq Y) = 0$. This relation partitions $L^2(\Omega, \mathcal{F}, P)$ into equivalence classes, and the space $L^2(\Omega, \mathcal{F}, P)$ has to be defined as the collection of these classes with an inner product defined as above. With this agreement, we actually have that

$$\langle X, X \rangle \;=\; 0 \quad \text{ if and only if } \quad X = \mathbf{0},$$

where $\mathbf{0}$ is the class of those random variables such that $P(X \neq 0) = 0$.

To simplify notation, we will continue to use the symbols $X, Y, \ldots$ for elements of $L^2(\Omega, \mathcal{F}, P)$. But we should keep in mind that we actually have to deal with **classes** of random variables. Next we will show that $L^2(\Omega, \mathcal{F}, P)$ is complete which means that this space is actually a Hilbert space.

**Theorem 1.3.5.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then $L^2(\Omega, \mathcal{F}, P)$ is complete.*

*Proof.* Let $(X_n)_{n\in\mathbb{N}}$ be an arbitrary Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$. We have to show that there exists some $X \in L^2(\Omega, \mathcal{F}, P)$ such that

$$\|X_n \,-\, X\| \underset{n\to\infty}{\longrightarrow} 0.$$

**(i) (Identification of a prospective limit)**

Since $(X_n)_{n\in\mathbb{N}}$ is a Cauchy sequence we can find a strictly increasing subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that

$$\|X_n \,-\, X_m\| \;\leq\; 2^{-k} \qquad \forall n, m \geq n_k,$$

which implies in particular that

$$\|X_{n_{k+1}} - X_{n_k}\| \le 2^{-k} \qquad \forall k \in \mathbb{N}.$$

With $n_0 = 0$ and $X_{n_0} = \mathbf{0}$, we obtain that

$$\begin{aligned} E\left[\sum_{j=1}^{\infty} |X_{n_j} - X_{n_{j-1}}|\right] &= \sum_{j=1}^{\infty} E|X_{n_j} - X_{n_{j-1}}| \qquad \text{(by monotone convergence)} \\ &\le \sum_{j=1}^{\infty} \|X_{n_j} - X_{n_{j-1}}\| \qquad \text{(by Cauchy-Schwarz)} \\ &\le \|X_{n_1}\| + \sum_{j=2}^{\infty} \underbrace{\|X_{n_j} - X_{n_{j-1}}\|}_{\le 2^{-(j-1)}} < \infty. \end{aligned}$$

Therefore, the random variable $\sum_{j=1}^{\infty} |X_{n_j} - X_{n_{j-1}}|$ is finite with probability 1, and

$$X_{n_k} = \sum_{j=1}^{k} (X_{n_j} - X_{n_{j-1}}) \underset{k\to\infty}{\longrightarrow} \sum_{j=1}^{\infty} (X_{n_j} - X_{n_{j-1}})$$

holds true with probability 1. We define

$$X(\omega) := \begin{cases} \lim_{k\to\infty} X_{n_k}(\omega) & \text{if } \sum_{j=1}^{\infty} |X_{n_j} - X_{n_{j-1}}| < \infty \\ 0 & \text{otherwise} \end{cases}.$$

**(ii) (Convergence of the full sequence)**
Using the fact that $|X_n - X|^2 = \liminf_{k\to\infty} |X_n - X_{n_k}|^2$ holds true with probability one we obtain by Fatou's lemma that

$$\int_{\Omega} |X_n - X|^2 \, dP = \int_{\Omega} \liminf_{k\to\infty} |X_n - X_{n_k}|^2 \, dP \le \liminf_{k\to\infty} \int_{\Omega} |X_n - X_{n_k}|^2 \, dP.$$

The right-hand side of this display can be made arbitrarily small by choosing $n$ large enough. This shows that $\int |X_n - X|^2 \, dP \underset{n\to\infty}{\longrightarrow} 0$.

**(iii)** $(X \in L^2(\Omega, \mathcal{F}, P))$
We obtain, again by Fatou's lemma, that

$$\begin{aligned} \int_{\Omega} X^2 \, dP &= \int_{\Omega} \liminf_{k\to\infty} X_{n_k}^2 \, dP \\ &\le \liminf_{k\to\infty} \int_{\Omega} X_{n_k}^2 \, dP \\ &\le \liminf_{k\to\infty} \left( \sum_{j=1}^{k} \|X_{n_j} - X_{n_{j-1}}\| \right)^2 < \infty. \end{aligned}$$

□

## Exercise

**Ex. 1.3.1** Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, P)$ and $(\beta_k)_{k\in\mathbb{Z}}$ be a sequence of real numbers. Assume that $E\varepsilon_t = 0$, $\sigma_\varepsilon^2 := E\varepsilon_t^2 < \infty$, and $\sum_{k=-\infty}^{\infty} \beta_k^2 < \infty$.

(i) Show that $\big(X_{t,m}\big)_{m\in\mathbb{N}}$ defined by

$$X_{t,m} := \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}$$

is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$.

(ii) Let $X_t$ be the $L^2$-limit of $\big(X_{t,m}\big)_{m\in\mathbb{N}}$.
Compute $\operatorname{cov}(X_{t+k}, X_t)$.

(iii) Let $\pi\colon \mathbb{Z} \to \mathbb{Z}$ be a bijective function. Then, for each $t \in \mathbb{Z}$, $\big(\beta_{\pi(k)}\varepsilon_{t-\pi(k)}\big)_{k\in\mathbb{Z}}$ is a rearrangement of the sequence $\big(\beta_k\varepsilon_{t-k}\big)_{t\in\mathbb{Z}}$.

a) Show that $\big(\widetilde{X}_{t,m}\big)_{m\in\mathbb{N}}$ defined by

$$\widetilde{X}_{t,m} := \sum_{k=-m}^{m} \beta_{\pi(k)} \varepsilon_{t-\pi(k)}$$

is also a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$.

b) Show that

$$\big\|X_{t,m} - \widetilde{X}_{t,m}\big\| \underset{m\to\infty}{\longrightarrow} 0$$

and conclude that

$$P\big(X_t = \widetilde{X}_t\big) = 1,$$

where $\widetilde{X}_t$ denotes the $L^2$-limit of $\big(\widetilde{X}_{t,m}\big)_{m\in\mathbb{N}}$.

## Projections in Hilbert spaces

I what follows we consider orthogonal projections in a Hilbert space and derive an easily applicable characterization of it. This sounds again rather abstract but we will see that such a general result is quite useful when "best" predictors of future values of a process have to be specified. (The word "best" is in quotation marks since we still have to specify what we mean by it.) Suppose that $X_1$, $X_2$ and $Y$ are squared integrable real-valued random variables on $(\Omega, \mathcal{F}, P)$. We observe realizations $x_1$ and $x_2$ of $X_1$ and $X_2$, respectively, and we may wish to approximate the value of $Y$ by using a linear combination $\widehat{Y} = \phi_1 X_1 + \phi_2 X_2$ of $X_1$ and $X_2$, which minimizes the **mean squared error of prediction** (MSEP),

$$S(\phi_1, \phi_2) \;=\; E\big[|Y \,-\, \phi_1 X_1 \,-\, \phi_2 X_2|^2\big] \;=\; \|Y \,-\, \phi_1 X_1 \,-\, \phi_2 X_2\|^2,$$

where $\|\cdot\|$ denotes the norm in $L^2(\Omega, \mathcal{F}, P)$. Suppose, for simplicity, that $X_1$ and $X_2$ are not collinear, i.e., neither one of $X_1$, $X_2$ is a multiple of the other. This means that the Cauchy-Schwarz inequality is strict, i.e. $E|X_1 X_2| < \sqrt{E[X_1^2]}\sqrt{E[X_2^2]}$; see Lemma 1.3.2. We have that

$$\begin{aligned} S(\phi_1, \phi_2) &= E\big[Y^2\big] + \phi_1^2 E\big[X_1^2\big] + \phi_2^2 E\big[X_2^2\big] + 2\phi_1\phi_2 E\big[X_1 X_2\big] - 2\phi_1 E\big[Y X_1\big] - 2\phi_2 E\big[Y X_2\big] \\ &= E\big[Y^2\big] \,+\, (\phi_1, \phi_2)\, M \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \,-\, 2\, E\Big[(Y X_1, Y X_2)\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}\Big], \end{aligned}$$

where

$$M \;=\; \begin{pmatrix} E\big[X_1^2\big] & E\big[X_1 X_2\big] \\ E\big[X_1 X_2\big] & E\big[X_2^2\big] \end{pmatrix}.$$

Since $\det(M) \,=\, E\big[X_1^2\big]\, E\big[X_2^2\big] - \big(E\big[X_1 X_2\big]\big)^2 \,>\, 0$ we conclude that the matrix $M$ is regular. Therefore we obtain that

$$S(\phi_1, \phi_2) \,\geq\, E\big[Y^2\big] \,+\, \Big\{\underbrace{\lambda_{\min}(M)}_{>0}\,\Big\|\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}\Big\| \,-\, 2\,\Big\|\begin{pmatrix} E[Y X_1] \\ E[Y X_2] \end{pmatrix}\Big\|\Big\} \,\cdot\, \Big\|\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}\Big\| \underset{\|\phi\|\to\infty}{\longrightarrow} \infty$$

which means that we can restrict our search for a minimizer $\binom{\phi_{10}}{\phi_{20}}$ of $S$, if it exists at all, to a sufficiently large **compact** subset $C \,:=\, \big\{\phi \,\in\, \mathbb{R}^2\colon\, \|\phi\| \,\leq\, c\big\}$ of $\mathbb{R}^2$. The function $S\colon\, \mathbb{R}^2 \to \mathbb{R}$ is continuous which implies that its infimum on $C$ is attained. As a necessary condition for a minimum, the partial derivatives of $S$ must be zero. It holds that $\frac{\partial}{\partial\phi_1} S(\phi_1, \phi_2) = \frac{\partial}{\partial\phi_2} S(\phi_1, \phi_2) = 0$ if and only if

$$\underbrace{\begin{pmatrix} E\big[X_1^2\big] & E\big[X_1 X_2\big] \\ E\big[X_1 X_2\big] & E\big[X_2^2\big] \end{pmatrix}}_{=M} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \;=\; \begin{pmatrix} E\big[Y X_1\big] \\ E\big[Y X_2\big] \end{pmatrix},$$

i.e., the values of $\phi_{10}$ and $\phi_{20}$ we are seeking are solutions to the so-called normal equation. Since the matrix $M$ is regular,

$$\begin{pmatrix} \phi_{10} \\ \phi_{20} \end{pmatrix} \;=\; M^{-1} \begin{pmatrix} E\big[Y X_1\big] \\ E\big[Y X_2\big] \end{pmatrix}$$

is the sought solution and the **best linear predictor** is given by

$$\widehat{Y} \;=\; \phi_{10} X_1 \,+\, \phi_{20} X_2.$$

It can be conjectured from these computations that the computation of best predictors may get quite cumbersome in more involved situations. For example, it is not clear what happens if the matrix $M$ were singular. Therefore, we use again the abstract context of Hilbert spaces to derive a general characterization which will be easily applicable. We begin with a definition.

**Definition.** A linear subspace $\mathcal{M}$ of a Hilbert space $\mathcal{H}$ is said to be a **closed subspace** if $\mathcal{M}$ contains all of its limit points, i.e., if $x_n \in \mathcal{M}$ $\forall n \in \mathbb{N}$ and $\|x_n - x\| \underset{n\to\infty}{\longrightarrow} 0$ for some $x \in \mathcal{H}$, then $x \in \mathcal{M}$.

**Theorem 1.3.6.** *If $\mathcal{M}$ is a closed subspace of the (real or complex) Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$, then*

*(i)* *there is a unique element $\widehat{x} \in \mathcal{M}$ such that*

$$\|x - \widehat{x}\| = \inf_{y\in\mathcal{M}} \|x - y\|,$$

*($\widehat{x}$ is the projection of $x$ onto $\mathcal{M}$, denoted $P_{\mathcal{M}}x$.)*

*(ii)* $\widehat{x} \in \mathcal{M}$ *and* $\|x - \widehat{x}\| = \inf_{y\in\mathcal{M}} \|x - y\|$
*if and only if*
$\widehat{x} \in \mathcal{M}$ *and* $\langle x - \widehat{x}, y\rangle = 0 \quad \forall y \in \mathcal{M}$.

To summarize, the above theorem ensures that a projection always exists and is unique. Moreover, part (ii) provides a criterion which can be used to determine this projection almost effortlessly, even in complex situations. This will be illustrated by the example given after the proof of this theorem.

*Proof of Theorem 1.3.6.* (i) Let $d := \inf_{y\in\mathcal{M}} \|x - y\|$. Then there exists a sequence $(y_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{M}$ such that $\|x - y_n\| \underset{n\to\infty}{\longrightarrow} d$. We show that $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. To this end, we use the so-called parallelogram law:

$$2\,\|a\|^2 + 2\,\|b\|^2 = \|a+b\|^2 + \|a-b\|^2 \qquad \forall a, b \in \mathcal{H}. \tag{1.3.2}$$

Indeed, we have

$$\begin{aligned} \|a+b\|^2 + \|a-b\|^2 &= \|a\|^2 + \|b\|^2 + \langle a,b\rangle + \langle b,a\rangle + \|a\|^2 + \|b\|^2 - \langle a,b\rangle - \langle b,a\rangle \\ &= 2\,\|a\|^2 + 2\,\|b\|^2. \end{aligned}$$

We obtain from (1.3.2)

$$\begin{aligned} \|y_m - y_n\|^2 &= \|(y_m - x) - (y_n - x)\|^2 \\ &= 2\,\|y_m - x\|^2 + 2\,\|y_n - x\|^2 - \|(y_m + y_n) - 2x\|^2 \\ &= 2\,\|y_m - x\|^2 + 2\,\|y_n - x\|^2 - 4\,\|\underbrace{(y_m + y_n)/2}_{\in\mathcal{M}} - x\|^2 \\ &\le 2\,\|y_m - x\|^2 + 2\,\|y_n - x\|^2 - 4d^2 \underset{m,n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Hence, $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and there exists an $\widehat{x} \in \mathcal{H}$ such that

$$\|y_n - \widehat{x}\| \underset{n\to\infty}{\longrightarrow} 0.$$

Since $\mathcal{M}$ is closed, we conclude that $\widehat{x} \in \mathcal{M}$. By continuity of the inner product (see (ii) of Lemma 1.3.4) we obtain that

$$\begin{aligned} \|x - \widehat{x}\|^2 &= \langle x - \widehat{x}, x - \widehat{x}\rangle \\ &= \lim_{n\to\infty} \langle x - y_n, x - y_n\rangle \\ &= \lim_{n\to\infty} \|x - y_n\|^2 = d^2. \end{aligned}$$

To establish uniqueness, suppose that $\widetilde{x} \in \mathcal{M}$ is an arbitrary projection. Then

$$\|x - \widetilde{x}\| = \|x - \widehat{x}\| = d$$

and, again by the parallelogram law (1.3.2),

$$\|\underbrace{(\widehat{x}+\widetilde{x})/2}_{\in\mathcal{M}} - x\|^2 + \underbrace{\|(\widehat{x}-x)/2 - (\widetilde{x}-x)/2\|^2}_{=\|\widehat{x}-\widetilde{x}\|^2/4} = \frac{1}{2}\|\widehat{x}-x\|^2 + \frac{1}{2}\|\widetilde{x}-x\|^2 = d^2.$$

If $\widetilde{x} \neq \widehat{x}$, then $\|\widehat{x} - \widetilde{x}\| > 0$, and so $\|(\widehat{x} + \widetilde{x})/2 - x\|^2 < d^2$, which contradicts our assumption that both $\widehat{x}$ and $\widetilde{x}$ are projections of $x$ onto $\mathcal{M}$. Hence $\widetilde{x} = \widehat{x}$.
(ii) $(\Longrightarrow)$
Suppose that $\widehat{x} \in \mathcal{M}$ and $\|x - \widehat{x}\| = \inf_{y\in\mathcal{M}} \|x - y\|$. Suppose further that there exists some $y \in \mathcal{M}$ such that

$$\langle x - \widehat{x}, y\rangle \neq 0.$$

We will show that there exists some $\widetilde{x} \in \mathcal{M}$ which is closer to $x$ than $\widehat{x}$. As a possible candidate, we take $\widetilde{x} = \widehat{x} + \alpha y$, where $\alpha \in \mathbb{C}$. (In case of a real Hilbert space, $\alpha \in \mathbb{R}$.) Since $\mathcal{M}$ is a linear space we also have that $\widetilde{x} \in \mathcal{M}$. Then

$$\begin{aligned} \|x - \widetilde{x}\|^2 &= \|x - \widehat{x} - \alpha y\|^2 = \langle (x - \widehat{x}) - \alpha y, (x - \widehat{x}) - \alpha y\rangle \\ &= \|x - \widehat{x}\|^2 + |\alpha|^2 \|y\|^2 - \alpha\langle y, x - \widehat{x}\rangle - \overline{\alpha\langle y, x - \widehat{x}\rangle}. \end{aligned}$$

Now we specify $\alpha$ as $\alpha = \epsilon\langle x - \widehat{x}, y\rangle$, where $\epsilon \in \mathbb{R}$. With this choice,

$$\|x - \widetilde{x}\|^2 = \|x - \widehat{x}\|^2 + \epsilon\,|\langle x - \widehat{x}, y\rangle|^2 \left\{\epsilon\|y\|^2 - 2\right\} < \|x - \widehat{x}\|^2$$

holds for sufficiently small $\epsilon > 0$. This is a contradiction to the assumption that $\widehat{x}$ is the projection.
$(\Longleftarrow)$
Suppose that $\widehat{x} \in \mathcal{M}$ and $\langle x - \widehat{x}, y\rangle = 0\ \forall y \in \mathcal{M}$. Let $\widetilde{x} \in \mathcal{M}$ be arbitrary. Then

$$\begin{aligned} \|x - \widetilde{x}\|^2 &= \langle x - \widehat{x} + \widehat{x} - \widetilde{x}, x - \widehat{x} + \widehat{x} - \widetilde{x}\rangle \\ &= \|x - \widehat{x}\|^2 + \|\widehat{x} - \widetilde{x}\|^2 + \underbrace{\langle x - \widehat{x}, \widehat{x} - \widetilde{x}\rangle}_{=0} + \underbrace{\langle \widehat{x} - \widetilde{x}, x - \widehat{x}\rangle}_{=0}. \end{aligned}$$

This implies that

$$\|x - \widehat{x}\| = \inf_{y\in\mathcal{M}} \|x - y\|.$$

□

## Application: Best linear prediction of a stationary process

Let $\mathbf{X} = (X_t)_{t\in\mathbb{N}}$ be a (weakly) stationary real-valued process on $(\Omega, \mathcal{F}, P)$ and let $\gamma_X$ be the autocovariance function of this process. To simplify matters, we assume that $EX_t = 0$. Suppose that realizations $x_1, \ldots, x_n$ (our data) of $X_1, \ldots, X_n$ are observed. We want to find the best linear predictor of $X_{n+1}$,

$$\widehat{X}_{n+1} = \sum_{j=1}^{n} \phi_{j0} X_{n+1-j},$$

where

$$E\big[\big(X_{n+1} - \widehat{X}_{n+1}\big)^2\big] = S(\phi_{10}, \ldots, \phi_{n0}) := \inf_{\phi_1,\ldots,\phi_n} E\Big[\Big(X_{n+1} - \sum_{j=1}^{n} \phi_j X_{n+1-j}\Big)^2\Big],$$

that is, $\widehat{X}_{n+1}$ minimizes the **mean squared error of prediction**.

To determine $\widehat{X}_{n+1}$, we could use basic calculus as above and set the partial derivatives of the functional $S$ equal to zero. If, in addition, the analogue of the matrix $M$ in the above example is regular, then there actually exists a unique solution. On the other hand, we could also employ the results of Theorem 1.3.6. This theorem tells us that a unique solution exists in any case, no matter if the counterpart of the matrix $M$ is regular or not. Moreover, it will be shown below that this solution is easily obtained using part (ii) of this theorem.

Let $\mathcal{H} = L^2(\Omega, \mathcal{F}, P)$ and $\mathcal{M} = \{\sum_{j=1}^{n} \alpha_j X_{n+1-j} \colon\ \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}$. It is clear that $\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$. Since

$$E\Big[\Big(X_{n+1} - \sum_{j=1}^{n} \phi_j X_{n+1-j}\Big)^2\Big] = \Big\|X_{n+1} - \sum_{j=1}^{n} \phi_j X_{n+1-j}\Big\|^2$$

it follows that the sought best predictor is just the orthogonal projection of $X_{n+1}$ onto the subspace $\mathcal{M}$. Therefore we see without hesitation, that the best linear predictor exists (i.e. the corresponding infimum is actually attained) and is unique. Part (ii) of Theorem 1.3.6 helps us to find coefficients $\phi_{10}, \ldots, \phi_{n0}$ such that $\widehat{X}_{n+1} = \sum_{j=1}^{n} \phi_{j0} X_{n+1-j}$. Since $\langle X_{n+1} - \widehat{X}_{n+1}, X\rangle$ has to be zero for all $X \in \mathcal{M}$ these coefficients have to solve the following system of equations:

$$\langle X_{n+1} - \sum_{j=1}^{n} \phi_{j0} X_{n+1-j}, X_k\rangle = 0 \qquad \forall k = n, n-1, \ldots, 1.$$

This is fulfilled if and only if

$$\underbrace{\begin{pmatrix} \langle X_{n+1}, X_n\rangle \\ \vdots \\ \langle X_{n+1}, X_1\rangle \end{pmatrix}}_{=:\,\gamma} = \underbrace{\begin{pmatrix} \langle X_{n+1-1}, X_n\rangle & \ldots & \langle X_{n+1-n}, X_n\rangle \\ \vdots & \ddots & \vdots \\ \langle X_{n+1-1}, X_1\rangle & \ldots & \langle X_{n+1-n}, X_1\rangle \end{pmatrix}}_{=:\,\Gamma_n} \underbrace{\begin{pmatrix} \phi_{10} \\ \vdots \\ \phi_{n0} \end{pmatrix}}_{=:\,\Phi_n}.$$

As already mentioned, Theorem 1.3.6 guarantees that there exists at least one solution, no matter whether or not the matrix $\Gamma_n$ is regular. If $\Gamma_n$ is singular, then there exist infinitely many solutions. However, Theorem 1.3.6 guarantees that every solution provides the same (uniquely defined) predictor $\widehat{X}_{n+1}$.

## 1.4 Linear processes

In this section, we consider so-called linear processes. Their simple structure allows us to derive their properties without much effort. Even a counterpart to the Lindeberg-Lévy central limit theorem can be easily derived. Moreover, it will be shown in the next section that certain processes with a more involved structure can be represented as such a linear process. We begin with a few definitions.

**Definition.** The process $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ is said to be **white noise** if

$$E\varepsilon_t \;=\; 0 \quad \forall t, \qquad \operatorname{cov}(\varepsilon_{t+h}, \varepsilon_t) \;=\; \begin{cases} \sigma_\varepsilon^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}.$$

Notation: $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$.
If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables such that $E\varepsilon_t = 0$ and $\operatorname{var}(\varepsilon_t) = \sigma_\varepsilon^2$, then we use the notation

$$(\varepsilon_t)_{t\in\mathbb{Z}} \;\sim\; \mathrm{IID}(0, \sigma_\varepsilon^2).$$

**Definition.** Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a real-valued process on $(\Omega, \mathcal{F}, P)$. Then the process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ defined by

$$X_t \;=\; \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$$

is said to be a **linear process**. In this context, the $\varepsilon_t$ are called **innovations**.

Special cases:

- $X_t \;=\; \sum_{k=0}^{q} \beta_k \varepsilon_{t-k}$
  Then $(X_t)_{t\in\mathbb{Z}}$ is an MA($q$) process (**moving average process** of order $q$).
- $X_t \;=\; \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$
  Then $(X_t)_{t\in\mathbb{Z}}$ is an MA($\infty$) process. (**causal linear process**)

**Remark 1.4.1.** *Regarding the sequence of innovations, there are different definitions of linear processes in the literature. For example, Brockwell and Davis ("Time Series: Theory and Methods") suppose that* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma_\varepsilon^2)$ *whereas Kreiß and Neuhaus ("Einführung in die Zeitreihenanalyse") assume that* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim IID(0, \sigma_\varepsilon^2)$*. In what follows, we adapt our assumption on the innovation process to the respective purpose.*

Note that the definition of a linear process involves an infinite series and it is not clear whether or not this series converges. The following proposition provides sufficient conditions for their convergence.

**Proposition 1.4.2.** *Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of real-valued random variables on $(\Omega, \mathcal{F}, P)$ and $(\beta_k)_{k\in\mathbb{Z}}$ be an absolutely convergent series, i.e. $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$.*

*(i) If $\sup_t \{E|\varepsilon_t|\} < \infty$, then the series*

$$X_t = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$$

*converges absolutely with probability 1.*

*(ii) If $\sup_t \{E[\varepsilon_t^2]\} < \infty$, then the series converges in mean square to the same limit $X_t$, i.e.,*

$$X_{t,m} := \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k} \xrightarrow{L^2} X_t.$$

*Proof.* (i) The monotone convergence theorem yields that

$$\begin{aligned} E\Big[\sum_{k=-\infty}^{\infty} |\beta_k \varepsilon_{t-k}|\Big] &= \lim_{m\to\infty} E\Big[\sum_{k=-m}^{m} |\beta_k|\,|\varepsilon_{t-k}|\Big] \\ &= \lim_{m\to\infty} \sum_{k=-m}^{m} |\beta_k|\, E|\varepsilon_{t-k}| \\ &\le \Big(\lim_{m\to\infty} \sum_{k=-m}^{m} |\beta_k|\Big) \sup_t \{E|\varepsilon_t|\} < \infty. \end{aligned}$$

Therefore,

$$P\Big(\sum_{k=-\infty}^{\infty} |\beta_k \varepsilon_{t-k}| < \infty\Big) = 1,$$

that is, the series $\sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$ converges absolutely with probability 1. We denote the limit by $X_t$.
(ii) Let $X_{t,m} := \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}$. Then, for $m < n$,

$$\begin{aligned} \|X_{t,n} - X_{t,m}\|^2 &= \Big\langle \sum_{m<|j|\le n} \beta_j \varepsilon_{t-j}, \sum_{m<|k|\le n} \beta_k \varepsilon_{t-k} \Big\rangle \\ &= \sum_{m<|j|\le n} \sum_{m<|k|\le n} \beta_j \beta_k E[\varepsilon_{t-j}\varepsilon_{t-k}] \\ &\le \underbrace{\Big(\sum_{m<|j|\le n} |\beta_j|\Big)^2}_{\to 0 \text{ as } m,n\to\infty} \sup_t \{E[\varepsilon_t^2]\}. \end{aligned}$$

Therefore,

$$\|X_{t,n} - X_{t,m}\| \longrightarrow 0 \qquad \text{as } m, n \to \infty,$$

i.e., $(X_{t,m})_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$. It follows from Theorem 1.3.5 (completeness of $L^2(\Omega, \mathcal{F}, P)$) that there exists some $\widetilde{X}_t \in L^2(\Omega, \mathcal{F}, P)$ such that

$$\|X_{t,m} - \widetilde{X}_t\| \underset{m\to\infty}{\longrightarrow} 0.$$

It follows from (i) that $X_{t,m} \xrightarrow{a.s.} X_t$ as $m \to \infty$. Therefore, we obtain by Fatou's lemma that

$$\begin{aligned} E\Big[|\widetilde{X}_t - X_t|^2\Big] &= E\Big[\liminf_{m\to\infty} |\widetilde{X}_t - X_{t,m}|^2\Big] \\ &\leq \liminf_{m\to\infty} E\Big[|\widetilde{X}_t - X_{t,m}|^2\Big] = 0, \end{aligned}$$

which implies that

$$P\left(\widetilde{X}_t = X_t\right) = 1.$$

□

## Stationarity of linear processes

In the following we investigate issues of stationarity. But why is this important? We have already learned that stationarity means that the statistical properties of a process generating a time series do not change over time. It does not mean that the series does not change over time, just that the way it changes does not itself change over time. As we have seen, future values of stationary processes can be **predicted**, as the way they change is predictable. Furthermore, besides the goal of predicting future values, stationarity also means that targets of a statistical analysis, as for example the mean of the random variables or the autocovariance structure, do not change over time. This means, the more values of the process we observe, the more information about such a fixed object we collect. This also makes a meaningful asymptotic analysis of statistical procedures possible. Typically, we can prove **consistency** of (sequences of) statistical estimators, that is, such a sequence of estimators converges to the target quantity as the size of the sample tends to infinity. Next we show that a linear process inherits properties of stationarity from the underlying innovation process.

**Proposition 1.4.3.** *Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a real-valued process on $(\Omega, \mathcal{F}, P)$, $X_t = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$, where $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$.*

*(i)* *If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and $E|\varepsilon_0| < \infty$, then $(X_t)_{t\in\mathbb{Z}}$ defined by $X_t = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$ is also strictly stationary.*

*(ii)* *Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is weakly stationary with autocovariance function $\gamma_\varepsilon$. Then $(X_t)_{t\in\mathbb{Z}}$ defined by $X_t = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$ is also weakly stationary and*

$$\begin{aligned} EX_t &= E\varepsilon_0 \Big(\sum_{k=-\infty}^{\infty} \beta_k\Big), \\ \gamma_X(h) &= \operatorname{cov}(X_{t+h}, X_t) = \sum_{j,k=-\infty}^{\infty} \beta_j \beta_k \gamma_\varepsilon(h - j + k). \end{aligned}$$

*Proof.* (i) To prove strict stationarity of the process $(X_t)_{t\in\mathbb{Z}}$, we have to show that

$$P^{X_{t_1},\dots,X_{t_k}} = P^{X_{t_1+t},\dots,X_{t_k+t}} \qquad \forall t_1,\dots,t_k,t\in\mathbb{Z},\ \forall k\in\mathbb{N}.$$

Let $t_1,\dots,t_k,t \in \mathbb{Z}$, $k \in \mathbb{N}$ be arbitrary. To simplify notation, we assume that $t_1 \le \dots \le t_k$. It is quite easy to show strict stationarity for a linear process with finite memory. We consider the truncated variables $X_{t,m} = \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}$. We have that

$$\big(X_{t_1,m},\dots,X_{t_k,m}\big)^T = \Big(\sum_{j=-m}^{m} \beta_j \varepsilon_{t_1-j},\dots,\sum_{j=-m}^{m} \beta_j \varepsilon_{t_k-j}\Big)^T = g(\varepsilon_{t_1-m},\dots,\varepsilon_{t_k+m}),$$

for some function $g\colon\ \mathbb{R}^{t_k-t_1+1+2m} \to \mathbb{R}^k$. We can represent $\big(X_{t_1+t,m},\dots,X_{t_k+t,m}\big)^T$ in an analogous manner,

$$\big(X_{t_1+t,m},\dots,X_{t_k+t,m}\big)^T = g(\varepsilon_{t_1+t-m},\dots,\varepsilon_{t_k+t+m}),$$

Since $(\varepsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary, we have

$$P^{\varepsilon_{t_1-m},\dots,\varepsilon_{t_k+m}} = P^{\varepsilon_{t_1+t-m},\dots,\varepsilon_{t_k+t+m}},$$

which implies that

$$P^{X_{t_1,m},\dots,X_{t_k,m}} = P^{g(\varepsilon_{t_1-m},\dots,\varepsilon_{t_k+m})} = P^{g(\varepsilon_{t_1+t-m},\dots,\varepsilon_{t_t+m})} = P^{X_{t_1+t,m},\dots,X_{t_k+t,m}}. \qquad (1.4.1)$$

This means that the truncated variables $X_{t,m}$ form a strictly stationary process.

To obtain stationarity of the process of interest, we make use of the convergence results stated in Proposition 1.4.2. According to statement (i) of this proposition, we have

$$X_{s,m} \xrightarrow{a.s.} X_s \quad \forall s\in\mathbb{Z}, \qquad \text{as } m\to\infty,$$

which yields

$$\big(X_{t_1,m},\dots,X_{t_k,m}\big)^T \xrightarrow{a.s.} \big(X_{t_1},\dots,X_{t_k}\big)^T$$

as well as

$$\big(X_{t_1+t,m},\dots,X_{t_k+t,m}\big)^T \xrightarrow{a.s.} \big(X_{t_1+t},\dots,X_{t_k+t}\big)^T.$$

It is well-known that almost sure convergence of a sequence of random variables implies its convergence in distribution. Since this is equivalent to weak convergence of the corresponding laws we conclude that

$$P^{X_{t_1,m},\dots,X_{t_k,m}} \Longrightarrow P^{X_{t_1},\dots,X_{t_k}}$$

and

$$P^{X_{t_1+t,m},\dots,X_{t_k+t,m}} \Longrightarrow P^{X_{t_1+t},\dots,X_{t_k+t}}.$$

Now it follows from (1.4.1) and uniqueness of the weak limit that

$$P^{X_{t_1},\dots,X_{t_k}} = P^{X_{t_1+t},\dots,X_{t_k+t}}.$$

(ii) While the expected value of $X_{t,m}$ is easily computed, the justification for the expected value of $X_t$ requires more care. By weak stationarity of $(\varepsilon_t)_{t\in\mathbb{Z}}$,

$$E\varepsilon_t^2 = E\varepsilon_0^2 < \infty \quad \forall t\in\mathbb{Z}.$$

Therefore, we obtain by the second statement of Proposition 1.4.2 that

$$E\,|X_{t,m} \;-\; X_t|^2 \underset{m\to\infty}{\longrightarrow} 0.$$

Since $E|X_{t,m} - X_t| \leq \sqrt{E(X_{t,m} - X_t)^2}$ we see that

$$\begin{aligned} EX_t &= \lim_{m\to\infty} EX_{t,m} = \lim_{m\to\infty} E\Big[\sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}\Big] \\ &= E\varepsilon_0 \Big(\lim_{m\to\infty} \sum_{k=-m}^{m} \beta_k\Big) = E\varepsilon_0 \Big(\sum_{k=-\infty}^{\infty} \beta_k\Big). \end{aligned}$$

Before we derive the second order properties of the process $(X_t)_{t\in\mathbb{Z}}$, note that linearity of the inner product applies to finite sums. Since $\|X_{t,m} - X_t\| \underset{m\to\infty}{\longrightarrow} 0$ we can use the continuity of the inner product (see (ii) of Lemma 1.3.4) and we obtain that

$$\begin{aligned} E\,[X_{t+h}X_t] &= \langle X_{t+h}, X_t\rangle \\ &= \lim_{m\to\infty} \langle X_{t+h,m}, X_{t,m}\rangle \\ &= \lim_{m\to\infty} \langle \sum_{j=-m}^{m} \beta_j \varepsilon_{t+h-j}, \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}\rangle \\ &= \lim_{m\to\infty} \sum_{j,k=-m}^{m} \beta_j\beta_k \underbrace{\langle \varepsilon_{t+h-j}, \varepsilon_{t-k}\rangle}_{\gamma_\varepsilon(h-j+k) + \big(E\varepsilon_0\big)^2} \\ &= \sum_{j,k=-\infty}^{\infty} \beta_j\beta_k\gamma_\varepsilon(h-j+k) \;+\; \big(E\varepsilon_0\big)^2 \Big(\sum_{k=-\infty}^{\infty} \beta_k\Big)^2. \end{aligned}$$

This implies that

$$\gamma_X(h) \;=\; \sum_{j,k=-\infty}^{\infty} \beta_j\beta_k\gamma_\varepsilon(h-j+k).$$

□

## Exercises

**Ex. 1.4.1** Suppose that $Y$ and $Z$ are uncorrelated random variables with $EY = EZ = 0$ and $EY^2 = EZ^2 = 1$. For $t \in \mathbb{N}$, let $X_t = Y\cos(\theta t) + Z\sin(\theta t)$, where $\theta \in \mathbb{R}$.

Show that $\widehat{X}_3 = 2\cos(\theta)X_2 - X_1$ is the best linear predictor of $X_3$ given $X_1$, $X_2$.

*Hint:* $E[X_sX_t] = \cos(\theta(s-t))$ *and* $\cos(2\theta) = (\cos(\theta))^2 - (\sin(\theta))^2$.

**Ex. 1.4.2** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$ and $X_t = \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}$, for some $\alpha \in \mathbb{R}$, $|\alpha| < 1$.

Show that $\widehat{X}_{n+1} := \alpha X_n$ is the best linear predictor of $X_{n+1}$ given $X_1, \ldots, X_n$.
*Hint: Argue that* $X_{n+1} - \alpha X_n = \varepsilon_{n+1}$ *and use (ii) of Lemma 1.3.4.*

## A central limit theorem for linear processes

In what follows we derive a central limit theorem for linear processes. Because of their particular structure, we can build on a well-known CLT for sequences of independent and identically distributed random variables. As a reminder, we quote a version of a CLT which is usually attributed to the Finnish mathematician Jarl Waldemar Lindeberg and the French mathematician Paul Pierre Lévy.

**Theorem 1.4.4.** *(Lindeberg-Lévy central limit theorem)*
*Suppose that* $(X_t)_{t\in\mathbb{N}}$ *is a sequence of i.i.d. random variables such that* $EX_t = 0$ *and* $\mathrm{var}(X_t) =: \sigma^2 \in [0,\infty)$*. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \stackrel{d}{\longrightarrow} Y \sim N(0,\sigma^2).$$

We use this well-known result to prove the following central limit theorem for linear processes.

**Theorem 1.4.5.** *Let* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim IID(0,\sigma_\varepsilon^2)$*. Suppose that* $X_t = \sum_{k=-\infty}^{\infty}\beta_k\varepsilon_{t-k}$*, where* $(\beta_k)_{k\in\mathbb{Z}}$ *is a sequence of real numbers such that* $\sum_{k=-\infty}^{\infty}|\beta_k| < \infty$*. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \stackrel{d}{\longrightarrow} Z \sim N(0,v),$$

*where*

$$v = \sigma_\varepsilon^2\Big(\sum_{k=-\infty}^{\infty}\beta_k\Big)^2.$$

This theorem will be proved in two steps. First we prove such a result for a simpler case where the random variables $X_t$ are replaced by their truncated versions $X_{t,m}$. This allows us to derive the desired result mainly by a simple re-arrangement of the terms in certain double sums.

**Lemma 1.4.6.** *Let* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim IID(0,\sigma_\varepsilon^2)$*,* $X_{t,m} = \sum_{k=-m}^{m}\beta_k\varepsilon_{t-k}$*. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m} \stackrel{d}{\longrightarrow} Z_m \sim N(0,v_m),$$

*where*

$$v_m = \sigma_\varepsilon^2\Big(\sum_{k=-m}^{m}\beta_k\Big)^2.$$

*Proof.* First of all, we change the order of summation,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{k=-m}^{m} \beta_k \varepsilon_{t-k} \\ &= \frac{1}{\sqrt{n}}\sum_{s=1-m}^{n+m} \varepsilon_s \sum_{k\colon\, |k|\le m,\, 1\le s+k\le n} \beta_k.\end{aligned}$$

Now we split up

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m} &= \frac{1}{\sqrt{n}}\sum_{s=1}^{n} \varepsilon_s \Big(\sum_{k=-m}^{m} \beta_k\Big) \\ &\quad - \frac{1}{\sqrt{n}}\sum_{s=1}^{n} \varepsilon_s \Big(\sum_{k\colon\, |k|\le m,\, s+k\notin\{1,\dots,n\}} \beta_k\Big) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{s\in\{1-m,\dots,0\}\cup\{n+1,\dots,n+m\}} \varepsilon_s \sum_{k\colon\, |k|\le m,\, 1\le s+k\le n} \beta_k \\ &=: T_{n,1} + T_{n,2} + T_{n,3},\end{aligned}$$

say. It follows from the Lindeberg-Lévy central limit theorem (Theorem 1.4.4) that

$$T_{n,1} \xrightarrow{d} Z_m. \tag{1.4.2}$$

The terms $T_{n,2}$ and $T_{n,3}$ both consist of a bounded number of summands with bounded expectation. (To see this for $T_{n,2}$, note that $\#\big\{k\colon |k| \le m, s+k \notin \{1,\dots,n\}\big\} = 0$ if $s \in \{m+1,\dots,n-m\}$.) Therefore,

$$E|T_{n,2} + T_{n,3}| = O(1/\sqrt{n}),$$

which implies that

$$T_{n,2} + T_{n,3} \xrightarrow{P} 0. \tag{1.4.3}$$

(1.4.2) and (1.4.3) yield the assertion. □

*Proof of Theorem 1.4.5.* The proof of this result will be split into three steps. We show that

a) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t$ can be well approximated by $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m}$,

b) by Lemma 1.4.6, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m}$ converges to a random variable with a $N(0,v_m)$ distribution,

c) $N(0,v_m) \Longrightarrow N(0,v)$, as $m\to\infty$.

We begin with justifying the first approximation. Let $\widetilde{X}_{t,m} := X_t - X_{t,m}$. Note that

$$\begin{aligned}\operatorname{cov}\big(\widetilde{X}_{s,m}, \widetilde{X}_{0,m}\big) &= \operatorname{cov}\Big(\sum_{k\colon |k|>m} \beta_k\varepsilon_{s-k}, \sum_{j\colon |j|>m} \beta_j\varepsilon_{-j}\Big) \\ &= \sum_{k,j\colon |k|,|j|>m} \beta_k\beta_j \underbrace{\langle \varepsilon_{s-k}, \varepsilon_{-j}\rangle}_{=0 \text{ if } j\ne k-s} \\ &= \sigma_\varepsilon^2 \sum_{k\colon |k|,|k-s|>m} \beta_k\beta_{k-s}.\end{aligned}$$

Therefore we obtain that

$$\begin{aligned} E\Big[\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \;-\; \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,m}\Big)^2\Big] &= \frac{1}{n}\,\mathrm{var}\Big(\sum_{t=1}^{n}\widetilde{X}_{t,m}\Big) \\ &= \frac{1}{n}\sum_{s,t=1}^{n}\mathrm{cov}\,\big(\widetilde{X}_{s,m},\widetilde{X}_{t,m}\big) \\ &= \sum_{s=-(n-1)}^{n-1}\underbrace{\frac{n-|s|}{n}}_{\le 1}\;\underbrace{\mathrm{cov}\,\big(\widetilde{X}_{s,m},\widetilde{X}_{0,m}\big)}_{=\sigma_\varepsilon^2\sum_{k:\;|k|,|k-s|>m}\beta_k\beta_{k-s}} \\ &\le \sigma_\varepsilon^2\Big(\sum_{k:\;|k|>m}|\beta_k|\Big)^2 \;=:\; C_m \qquad \forall n\in\mathbb{N}. \end{aligned}$$

Furthermore, $\sum_{k=-\infty}^{\infty}|\beta_k|<\infty$ implies that

$$C_m \underset{m\to\infty}{\longrightarrow} 0.$$

Note that a) to c) contain results with different modes of convergence: While a) is an $L^2$-approximation, b) states convergence in distribution for sequences of random variables, and c) is related to weak convergence of probability measures. Furthermore, approximations a) and c) require that $m\to\infty$ whereas b) holds for fixed $m$ and is based on $n\to\infty$. To bring together these results, we employ the concept of characteristic functions.

Let $\varphi_S$ denote the characteristic function of a generic random variable $S$. According to Lévy's continuity theorem, convergence in distribution of a sequence of random variables is equivalent to pointwise convergence of their characteristic functions. Therefore, it suffices to show that

$$\varphi_{n^{-1/2}\sum_{t=1}^{n}X_t}(u) \underset{n\to\infty}{\longrightarrow} \varphi_Z(u) \qquad \forall u\in\mathbb{R}. \tag{1.4.4}$$

Let $u\in\mathbb{R}$ and $\epsilon>0$ be arbitrary. We split up:

$$\begin{aligned} \big|\varphi_{n^{-1/2}\sum_{t=1}^{n}X_t}(u) - \varphi_Z(u)\big| &\le \big|\varphi_{n^{-1/2}\sum_{t=1}^{n}X_t}(u) - \varphi_{n^{-1/2}\sum_{t=1}^{n}X_{t,m}}(u)\big| \\ &\quad + \big|\varphi_{n^{-1/2}\sum_{t=1}^{n}X_{t,m}}(u) - \varphi_{Z_m}(u)\big| \\ &\quad + \big|\varphi_{Z_m}(u) - \varphi_Z(u)\big| \\ &=: T_{1,m,n} + T_{2,m,n} + T_{3,m}, \end{aligned}$$

say. Since $|e^{ix}-e^{iy}|\le|x-y|\;\forall x,y\in\mathbb{R}$ we obtain

$$\begin{aligned} T_{1,m,n} &\le E\big|e^{iu(n^{-1/2}\sum_{t=1}^{n}X_t)} - e^{iu(n^{-1/2}\sum_{t=1}^{n}X_{t,m})}\big| \\ &\le u\,E\Big|n^{-1/2}\sum_{t=1}^{n}X_t - n^{-1/2}\sum_{t=1}^{n}X_{t,m}\Big| \\ &\le u\sqrt{C_m} \qquad \forall n\in\mathbb{N}. \end{aligned}$$

From Lemma 1.4.6 we conclude that

$$T_{2,m,n} \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall m\in\mathbb{N}.$$

Finally, we obtain from $v_m \longrightarrow_{m\to\infty} v$ that

$$T_{3,m} = \big|e^{-u^2v_m/2} - e^{-u^2v/2}\big| \underset{m\to\infty}{\longrightarrow} 0.$$

For the $\epsilon$ chosen above, we can find some $m_0 = m_0(\epsilon)$ such that

$$T_{1,m_0,n} + T_{3,m_0} \leq \epsilon/2 \qquad \forall n \in \mathbb{N}.$$

Moreover, there exists some $n_0 = n_0(\varepsilon, m_0)$ such that

$$T_{2,m_0,n} \leq \epsilon/2 \qquad \forall n \geq n_0.$$

The latter two estimates yield that

$$\left|\varphi_{n^{-1/2}\sum_{t=1}^n X_t}(u) - \varphi_Z(u)\right| \leq \epsilon \qquad \forall n \geq n_0,$$

which proves (1.4.4). This completes the proof. □

## Exercise

**Ex. 1.4.3** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \text{IID}(0, \sigma_\varepsilon^2)$ and let $X_0$ be an arbitrary random variable. The stochastic process $(X_t)_{t\in\mathbb{N}_0}$ is defined recursively by

$$X_t = \alpha X_{t-1} + \varepsilon_t \qquad \forall t \in \mathbb{N},$$

where $|\alpha| < 1$.

(i) Show that $\frac{1}{\sqrt{n}}\sum_{t=1}^n (X_t - \widetilde{X}_t) \stackrel{P}{\longrightarrow} 0$, where $\widetilde{X}_t = \sum_{k=0}^\infty \alpha^k \varepsilon_{t-k}$.

(ii) Does $Z_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n X_t$ converge in distribution to a normally distributed random variable $Z$? If so, what is the variance of $Z$?

## Nonparametric estimation of the mean and the autocovariance function

Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a weakly stationary (real-valued) process with mean $\mu$ and autocovariance function $\gamma$. Assume that realizations $x_1,\ldots,x_n$ of $X_1,\ldots,X_n$ are observed. If nothing is known about the time series, besides that it is (weakly) stationary, then natural estimators for the parameters $\mu$ and $\gamma(k)$ are

$$\widehat{\mu}_n \;=\; \bar{X}_n \;=\; \frac{1}{n}\sum_{t=1}^{n} X_t$$

and

$$\widehat{\gamma}_n(k) \;:=\; \frac{1}{n}\sum_{t=1}^{n-|k|} \big(X_{t+|k|}-\bar{X}_n\big)\big(X_t-\bar{X}_n\big). \qquad (|k|\le n-1)$$

For $|k|\ge n$, our sample does not provide information about $\gamma(k)$ and we simply set $\widehat{\gamma}_n(k)=0$. These estimators are called **nonparametric** because they are not based on a model described by a finite-dimensional parameter. Their advantage is that they work for (almost) every stationary time series. In what follows we investigate the statistical properties of these estimators.

We begin with the estimator $\widehat{\mu}_n$ of the common mean $\mu$ of the $X_t$. This estimator, or more exactly, the sequence $(\widehat{\mu}_n)_{n\in\mathbb{N}}$ is consistent under appropriate conditions. For example, if

$$X_t \;=\; \mu \,+\, \sum_{k=-\infty}^{\infty} \beta_k\varepsilon_{t-k},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}\sim \mathrm{IID}(0,\sigma_\varepsilon^2)$ and $\sum_{k=-\infty}^{\infty}|\beta_k|<\infty$, then it follows from Theorem 1.4.5 that

$$\sqrt{n}\,(\widehat{\mu}_n \,-\, \mu) \;=\; \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{k=-\infty}^{\infty} \beta_k\varepsilon_{t-k} \;\stackrel{d}{\longrightarrow}\; Z\sim N(0,v), \qquad \text{as } n\to\infty,$$

where $v \,=\, \sigma_\varepsilon^2\big(\sum_{k=-\infty}^{\infty}\beta_k\big)^2$. The statistical relevance of this result is that the sample mean is an asymptotically consistent estimator of $\mu$, with precision of the order $1/\sqrt{n}$. This result may be used in a preciser way to derive an **asymptotic confidence interval** for $\mu$. If $v>0$, then

$$\frac{\widehat{\mu}_n \,-\, \mu}{\sqrt{v/n}} \;\stackrel{d}{\longrightarrow}\; Z_0\sim N(0,1). \tag{1.4.5}$$

Let, for $\alpha\in(0,1)$, $\Phi^{-1}(1-\alpha/2)$ denote the $(1-\alpha/2)$-quantile of a standard normal distribution, i.e., for $Z_0\sim N(0,1)$, $P\big(Z_0\le \Phi^{-1}(1-\alpha/2)\big)=1-\alpha/2$. Then

$$C_\mu \;:=\; \big[\widehat{\mu}_n \,-\, \sqrt{v/n}\,\Phi^{-1}(1-\alpha/2), \widehat{\mu}_n \,+\, \sqrt{v/n}\,\Phi^{-1}(1-\alpha/2)\big]$$

is a confidence interval for $\mu$ with an asymptotic coverage probability of $1-\alpha$. Indeed, it follows from (1.4.5) that

$$\begin{aligned} P\big(\mu\in C_\mu\big) \;&=\; P\big(|\widehat{\mu}_n-\mu| \,\le\, \sqrt{v/n}\,\Phi^{-1}(1-\alpha/2)\big) \\ &=\; \underbrace{P\Big(\frac{\widehat{\mu}_n \,-\, \mu}{\sqrt{v/n}} \,\le\, \Phi^{-1}(1-\alpha/2)\Big)}_{\underset{n\to\infty}{\longrightarrow} 1-\alpha/2} \,-\, \underbrace{P\Big(\frac{\widehat{\mu}_n \,-\, \mu}{\sqrt{v/n}} \,<\, -\Phi^{-1}(1-\alpha/2)\Big)}_{\underset{n\to\infty}{\longrightarrow} \alpha/2} \\ &\underset{n\to\infty}{\longrightarrow}\; 1 \,-\, \alpha. \end{aligned} \tag{1.4.6}$$

Of course, this result is of limited value since prior knowledge of the parameter $v$ is hardly available in practice. We will see later how $v$ can be estimated. For the time being, suppose that $(\widehat{v}_n)_{n\in\mathbb{N}}$ is any **consistent** sequence of estimators of $v$, i.e.

$$\widehat{v}_n \stackrel{P}{\longrightarrow} v \qquad \text{as } n \to \infty.$$

Here, "$\stackrel{P}{\longrightarrow}$" denotes convergence in probability which means that, for all $\epsilon > 0$, $P\big(|\widehat{v}_n - v| > \epsilon\big) \underset{n\to\infty}{\longrightarrow} 0$. We conclude from (1.4.5) that[2]

$$\frac{\widehat{\mu}_n - \mu}{\sqrt{\widehat{v}_n/n}} = \underbrace{\frac{\sqrt{v}}{\sqrt{\widehat{v}_n}}}_{\stackrel{P}{\to} 1} \underbrace{\frac{\widehat{\mu}_n - \mu}{\sqrt{v/n}}}_{\stackrel{d}{\to} Z_0} \stackrel{d}{\longrightarrow} Z_0 \sim N(0,1)$$

Therefore,

$$\widehat{C}_\mu := \Big[\widehat{\mu}_n - \sqrt{\widehat{v}_n/n}\,\Phi^{-1}(1-\alpha/2), \widehat{\mu}_n + \sqrt{\widehat{v}_n/n}\,\Phi^{-1}(1-\alpha/2)\Big]$$

is a meaningful asymptotic $(1-\alpha)$-confidence interval for $\mu$.

---

[2] Here we use the result that, for two sequences $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ of real-valued random variables, $X_n \stackrel{d}{\to} X$ and $Y_n \stackrel{P}{\to} 1$ together imply that $X_n \cdot Y_n \stackrel{d}{\to} X$.
To see this, we first show that $X_n(Y_n - 1) \stackrel{P}{\to} 0$. Let $\epsilon > 0$ be arbitrary. Since $X_n \stackrel{d}{\to} X$ we have that $P(X_n \leq u) \underset{n\to\infty}{\longrightarrow} P(X \leq u)$ for all continuity points $u$ of the mapping $x \mapsto P(X \leq x)$. Since the set of discontinuity points $u$ of $x \mapsto P(X \leq x)$ is countable we find some $M = M(\epsilon) < \infty$ such that $-M$ and $M$ are continuity points and

$$P(X \leq -M \;\text{ or }\; X > M) \leq \epsilon/3,$$

which implies that

$$P(|X_n| > M) \leq P(X_n \leq -M \;\text{ or }\; X_n > M) \leq (2/3)\epsilon \qquad \forall n \geq N_1$$

and sufficiently large $N_1$. Since $Y_n \stackrel{P}{\to} 1$ we obtain, for arbitrary $K > 0$,

$$P\big(|X_n(Y_n-1)| > K\big) \leq P\big(|X_n| > M\big) + P\big(|Y_n - 1| > K/M\big) \leq \epsilon$$

for $n$ sufficiently large. Hence, $X_n(Y_n - 1) \stackrel{P}{\to} 0$.
Now we obtain

$$X_n \cdot Y_n = \underbrace{X_n}_{\stackrel{d}{\to} X} + \underbrace{X_n(Y_n-1)}_{\stackrel{P}{\to} 0} \stackrel{d}{\to} X.$$

Now we turn to the problem of estimating the autocovariance function $\gamma$. As an estimator of $\gamma(k)$, we consider

$$\widehat{\gamma}_n(k) := \begin{cases} \frac{1}{n}\sum_{t=1}^{n-|k|}\big(X_{t+|k|}-\bar{X}_n\big)\big(X_t-\bar{X}_n\big) & \text{if } |k|\le n-1, \\ 0 & \text{if } |k|\ge n. \end{cases}$$

Note that there seem to be alternatives to the choice of the factor $1/n$ in the definition of $\widehat{\gamma}_n(k)$. In the case of independent and identically distributed random variables, it is well-known that $\frac{1}{n-1}\sum_{t=1}^{n}(X_t-\bar{X}_n)^2$ is an unbiased estimator of $\gamma(0)=\mathrm{var}(X_1)$. Moreover, it also seems to make sense to divide by the number $n-|k|$ of summands in the definition of $\widehat{\gamma}_n(k)$. However, the following lemma provides a strong argument in favor of the above choice: The factor $1/n$ ensures that the function $\widehat{\gamma}_n\colon\ \mathbb{Z}\to\mathbb{R}$ has the desirable property of being the autocovariance function of an appropriate stationary process. We will therefore stick to the above definition of this estimator. To see why this holds true, recall from Theorem 1.2.5 the fact that a function $\kappa\colon\ \mathbb{Z}\to\mathbb{R}$ is the autocovariance function of a stationary process if and only if $\kappa$ is an even and non-negative definite function.

**Lemma 1.4.7.** *The function $\widehat{\gamma}_n\colon\ \mathbb{Z}\to\mathbb{R}$ is even and non-negative definite.*

*Proof.* It is obvious that $\widehat{\gamma}_n(k)=\widehat{\gamma}_n(-k)\ \forall k\in\mathbb{Z}$, that is, $\widehat{\gamma}_n$ is an even function.

To check the property of non-negative definiteness, let $t_1,\ldots,t_k\in\mathbb{Z}$, $a_1,\ldots,a_k\in\mathbb{R}$, and $k\in\mathbb{N}$ be arbitrary. We have to show that

$$\sum_{i,j=1}^{k} a_i\,\widehat{\gamma}_n(t_i-t_j)\,a_j \;\ge\; 0. \tag{1.4.7}$$

To simplify notation, we assume, w.l.o.g., that $t_1\le t_2\le\ldots\le t_k$. Note that (1.4.7) can be rewritten as

$$a^T\Gamma a \;\ge\; 0,$$

where

$$\Gamma = \begin{pmatrix} \widehat{\gamma}_n(t_1-t_1) & \ldots & \widehat{\gamma}_n(t_1-t_k) \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}_n(t_k-t_1) & \ldots & \widehat{\gamma}_n(t_k-t_k) \end{pmatrix}, \qquad a = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}.$$

However, the matrix $\Gamma$ can be represented as

$$\Gamma \;=\; M^T M,$$

where

$$M = \frac{1}{\sqrt{n}} \begin{pmatrix} X_1-\bar{X}_n & 0_{t_2-t_1} & & 0_{t_k-t_1} \\ \vdots & X_1-\bar{X}_n & & \\ X_n-\bar{X}_n & \vdots & \ldots & X_1-\bar{X}_n \\ & X_n-\bar{X}_n & & \vdots \\ 0_{t_k-t_1} & 0_{t_k-t_2} & & X_n-\bar{X}_n \end{pmatrix}.$$

Now we can see that

$$\sum_{i,j=1}^{k} a_i\,\widehat{\gamma}_n(t_i-t_j)\,a_j \;=\; a^T\Gamma a \;=\; a^T M^T M a \;=\; \|Ma\|^2 \;\ge\; 0,$$

that is, the function $\widehat{\gamma}_n$ is non-negative definite. □

It may seem tempting to replace the factor $1/n$ in the definition of $\widehat{\gamma}_n(k)$ by $1/(n-|k|)$, because there are $n-|k|$ terms in the sum. While with our advocated choice of the factor $1/n$ the function $\widehat{\gamma}_n$ is guaranteed to be non-negative definite, this is not true in general when we use the alternative factors $1/(n-|k|)$. To see this, we consider the following counterexample. Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables such that $P(X_t = 1) = P(X_t = -1) = P(X_t = 0) = 1/3$. Let $n \geq 3$. Then $X_1 = 1$, $X_n = -1$, and $X_2 = \ldots = X_{n-1} = 0$ hold with a probability of $(1/3)^n$. We obtain, for

$$\widetilde{\gamma}_n(k) \;=\; \begin{cases} \frac{1}{n-|k|}\sum_{t=1}^{n-|k|}\big(X_{t+|k|}-\bar{X}_n\big)\big(X_t-\bar{X}_n\big) & \text{if } |k| \leq n-1, \\ 0 & \text{if } |k| \geq n, \end{cases}$$

that

$$\widetilde{\Gamma}_n \;=\; \begin{pmatrix} \widetilde{\gamma}_n(0) & \widetilde{\gamma}_n(-1) & \ldots & \widetilde{\gamma}_n(2-n) & \widetilde{\gamma}_n(1-n) \\ \widetilde{\gamma}_n(1) & \ddots & \ddots & & \widetilde{\gamma}_n(2-n) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \widetilde{\gamma}_n(n-2) & & \ddots & \ddots & \widetilde{\gamma}_n(-1) \\ \widetilde{\gamma}_n(n-1) & \widetilde{\gamma}_n(n-2) & \ldots & \widetilde{\gamma}_n(1) & \widetilde{\gamma}_n(0) \end{pmatrix} \;=\; \begin{pmatrix} \frac{2}{n} & 0 & \ldots & 0 & -1 \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \ddots & 0 \\ -1 & 0 & \ldots & 0 & \frac{2}{n} \end{pmatrix}.$$

Let $c = \big(1, \underbrace{0,\ldots,0}_{n-2 \text{ times}}, 1\big)^T$. Then

$$c^T\widetilde{\Gamma}_n c \;=\; \frac{4}{n} \,-\, 2 \;<\; 0,$$

i.e., the matrix $\widetilde{\Gamma}_n$ is not non-negative definite. Therefore, the function $k \mapsto \widetilde{\gamma}_n(k)$ cannot be the autocovariance function of a stationary process.

In contrast to the case of the mean, an analysis of the asymptotic behavior of our estimator $\widehat{\gamma}_n$ of the autocovariance function requires more effort. We begin with the bias of $\widehat{\gamma}_n(k)$.

**Lemma 1.4.8.** *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a weakly stationary process with autocovariance function $\gamma$, where $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$.*
*Then, for fixed $k \in \mathbb{Z}$,*

$$E\widehat{\gamma}_n(k) \;=\; \gamma(k) \,+\, O\big(n^{-1}\big).$$

*Proof of Lemma 1.4.8.* Let $\mu = EX_t$. Then

$$\begin{aligned}
E\widehat{\gamma}_n(k) &= \frac{1}{n}\sum_{t=1}^{n-|k|} E\Big[\big(X_{t+|k|}-\mu+\mu-\bar{X}_n\big)\big(X_t-\mu+\mu-\bar{X}_n\big)\Big] \\
&= \frac{1}{n}\sum_{t=1}^{n-|k|} E\Big[\big(X_{t+|k|}-\mu\big)\big(X_t-\mu\big)\Big] \\
&\quad - \frac{1}{n}\sum_{t=1}^{n-|k|}\Big\{E\Big[\big(X_{t+|k|}-\mu\big)\big(\bar{X}_n-\mu\big)\Big] + E\Big[\big(X_t-\mu\big)\big(\bar{X}_n-\mu\big)\Big]\Big\} \\
&\quad + \frac{n-|k|}{n}\, E\Big[\big(\bar{X}_n-\mu\big)^2\Big] \\
&= \frac{n-|k|}{n}\,\gamma(k) \\
&\quad - \frac{1}{n^2}\sum_{t=1}^{n-|k|}\sum_{l=1}^{n}\big\{\gamma(t+|k|-l) + \gamma(t-l)\big\} \\
&\quad + \frac{n-|k|}{n^3}\sum_{s,t=1}^{n}\gamma(s-t).
\end{aligned}$$

Since $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$ we obtain that

$$E\widehat{\gamma}_n(k) \;=\; \gamma(k) \,+\, O\big(n^{-1}\big).$$

□

If $\gamma$ is the autocovariance function of a stationary process $(X_t)_{t\in\mathbb{Z}}$, where $X_t = \mu + \sum_{k=-\infty}^{\infty}\beta_k\varepsilon_{t-k}$, $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma_\varepsilon^2)$, and $\sum_{k=-\infty}^{\infty}|\beta_k| < \infty$, then $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$. Indeed, by Proposition 1.4.2, the autocovariance function of a linear process is given by

$$\gamma(h) \;=\; \sum_{j,k=-\infty}^{\infty}\beta_j\beta_k\,\mathrm{cov}(\varepsilon_{h-j},\varepsilon_{-k}).$$

Since $\mathrm{cov}(\varepsilon_{h-j},\varepsilon_k) = \sigma_\varepsilon^2$ if $k = j-h$ and zero otherwise, this reduces to $\gamma(h) = \sigma_\varepsilon^2\sum_{j=-\infty}^{\infty}\beta_j\beta_{j-h}$. Therefore,

$$\begin{aligned}
\sum_{h=-\infty}^{\infty}\big|\gamma(h)\big| &\le \sigma_\varepsilon^2\sum_{j,h}\big|\beta_j\big|\big|\beta_{j-h}\big| \\
&= \sigma_\varepsilon^2\Big(\sum_{j=-\infty}^{\infty}\big|\beta_j\big|\Big)^2 \,<\, \infty.
\end{aligned}$$

In order to show that the quadratic risk of $\widehat{\gamma}_n(k)$ tends to 0 as $n\to\infty$, we still have to estimate the variance of this estimator. To this end, we represent $\widehat{\gamma}_n(k)$ as a quadratic form and make use of the following lemma.

**Lemma 1.4.9.** *Suppose that $Y_1, \ldots, Y_n$ are real-valued random variables on $(\Omega, \mathcal{F}, P)$ such that $EY_t = 0$ and $EY_t^4 < \infty$ for all $t$. Furthermore, let $M$ be a symmetric $(n \times n)$-matrix and $Y = (Y_1, \ldots, Y_n)^T$.*
*Then*

$$\operatorname{var}\big(Y^T M Y\big) \;=\; \sum_{s,t,u,v=1}^{n} M_{st}\, M_{uv}\, \operatorname{cum}(Y_s, Y_t, Y_u, Y_v) \;+\; 2 \operatorname{tr}\big(M\Sigma M\Sigma\big),$$

*where*

$$\operatorname{cum}(Y_s, Y_t, Y_u, Y_v) \;:=\; E\big[Y_s Y_t Y_u Y_v\big] - E\big[Y_s Y_t\big] E\big[Y_u Y_v\big] - E\big[Y_s Y_u\big] E\big[Y_t Y_v\big] - E\big[Y_s Y_v\big] E\big[Y_t Y_u\big]$$

*denotes the* **joint cumulant** *of $Y_s$, $Y_t$, $Y_u$, and $Y_v$, $\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}$ the trace of an $(n \times n)$-matrix $A$, and $\Sigma = \operatorname{Cov}(Y)$ the covariance matrix of the random vector $Y$.*

*Proof.* Using $Y^T M Y = \sum_{s,t=1}^{n} M_{st} Y_s Y_t$ we obtain

$$\begin{aligned}
\operatorname{var}\big(Y^T M Y\big) &= E\big[Y^T M Y\, Y^T M Y\big] \;-\; E\big[Y^T M Y\big]\; E\big[Y^T M Y\big] \\
&= \sum_{s,t,u,v=1}^{n} M_{st} M_{uv} \Big( E\big[Y_s Y_t Y_u Y_v\big] \;-\; E\big[Y_s Y_t\big]\, E\big[Y_u Y_v\big] \Big) \\
&= \sum_{s,t,u,v=1}^{n} M_{st} M_{uv} \operatorname{cum}(Y_s, Y_t, Y_u, Y_v) \\
&\quad + \sum_{t=1}^{n} \sum_{s,u,v=1}^{n} \underbrace{M_{st}}_{= M_{ts}} \underbrace{E\big[Y_s Y_u\big]}_{= \Sigma_{su}} M_{uv} \underbrace{E\big[Y_v Y_t\big]}_{= \Sigma_{vt}} \\
&\quad + \sum_{t=1}^{n} \sum_{s,u,v=1}^{n} \underbrace{M_{st}}_{= M_{ts}} \underbrace{E\big[Y_s Y_v\big]}_{= \Sigma_{sv}} \underbrace{M_{uv}}_{= M_{vu}} \underbrace{E\big[Y_u Y_t\big]}_{= \Sigma_{ut}} \\
&= \sum_{s,t,u,v=1}^{n} M_{st} M_{uv} \operatorname{cum}(Y_s, Y_t, Y_u, Y_v) \\
&\quad + \; 2 \underbrace{\sum_{t=1}^{n} \big(M\Sigma M\Sigma\big)_{tt}}_{= \operatorname{tr}(M\Sigma M\Sigma)} .
\end{aligned}$$

□

Before we use Lemma 1.4.9 to derive an upper estimate for the variance of $\widehat{\gamma}_n(k)$, we state a few useful properties of cumulants which show in particular that a typical assumption on sums of cumulants will be satisfied under weak conditions. Note that the following elementary properties follow directly from the definition of cumulants. Let $Y_1, \ldots, Y_5$ be real-valued random variables with finite fourth moments and let $\alpha \in \mathbb{R}$. Then

(i) $\operatorname{cum}(Y_1, Y_2, Y_3, Y_4) = \operatorname{cum}(Y_{\pi(1)}, Y_{\pi(2)}, Y_{\pi(3)}, Y_{\pi(4)})$, for any permutation $\pi\colon \{1, \ldots, 4\} \to \{1, \ldots, 4\}$,

(ii) $\operatorname{cum}(Y_1 + Y_2, Y_3, Y_4, Y_5) = \operatorname{cum}(Y_1, Y_3, Y_4, Y_5) + \operatorname{cum}(Y_2, Y_3, Y_4, Y_5)$,

(iii) $\operatorname{cum}(\alpha Y_1, Y_2, Y_3, Y_4) = \alpha \operatorname{cum}(Y_1, Y_2, Y_3, Y_4)$.

## Exercise

**Ex. 1.4.4** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0, \sigma^2)$ and $E\varepsilon_t^4 < \infty \ \forall t \in \mathbb{Z}$. Show that

(i) $\operatorname{cum}\big(\varepsilon_s, \varepsilon_t, \varepsilon_u, \varepsilon_v\big) = \begin{cases} E[\varepsilon_0^4] - 3\sigma^4 & \text{if } s = t = u = v, \\ 0 & \text{otherwise} \end{cases}$.

(ii) Let $X_t = \beta_0 \varepsilon_t + \cdots + \beta_q \varepsilon_{t-q}$. Show that, for arbitrary $s \in \mathbb{Z}$,

$$\sum_{t,u,v=-\infty}^{\infty} \big|\operatorname{cum}\big(X_s, X_t, X_u, X_v\big)\big| \le \big|E\big[\varepsilon_0^4\big] - 3\sigma^4\big| \left(\sum_{k=0}^{q} |\beta_k|\right)^4 < \infty.$$

**Lemma 1.4.10.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a centered Gaussian process, i.e., $EX_t = 0 \ \forall t \in \mathbb{Z}$. Then*

$$\operatorname{cum}\big(X_s, X_t, X_u, X_v\big) = 0 \qquad \forall s, t, u, v \in \mathbb{Z}.$$

*Proof.* Let $Z_1, \ldots, Z_4 \sim N(0, 1)$ be independent. Since $E[Z_i^4] = 3 = 3(E[Z_i^2])^2$ we obtain by (i) of Exercise 1.7 that

$$\operatorname{cum}\big(Z_i, Z_j, Z_k, Z_l\big) = 0 \qquad \forall i, j, k, l \in \{1, \ldots, 4\}.$$

Let now $s, t, u, v \in \mathbb{Z}$ be arbitrary. Since $(X_t)_{t\in\mathbb{Z}}$ is Gaussian, the vector $X = (X_s, X_t, X_u, X_v)^T$ has a multivariate normal distribution, $X \sim N_4(0_4, \Sigma)$, for some symmetric and non-negative definite matrix $\Sigma$. Let $M := \Sigma^{1/2}$ be the square root of $\Sigma$. Then the vector $X$ has the same distributions as $MZ$, where $Z = (Z_1, Z_2, Z_3, Z_4)^T$, and it follows that

$$\begin{aligned} \operatorname{cum}\big(X_s, X_t, X_u, X_v\big) &= \operatorname{cum}\Big(\sum_i M_{1i} Z_i, \sum_j M_{2j} Z_j, \sum_k M_{3k} Z_k, \sum_l M_{4l} Z_l\Big) \\ &= \sum_{i,j,k,l=1}^{4} M_{1i} M_{2j} M_{3k} M_{4l} \underbrace{\operatorname{cum}\big(Z_i, Z_j, Z_k, Z_l\big)}_{=0} = 0. \end{aligned}$$

□

Now we are in a position to show that the variance of $\widehat{\gamma}_n(k)$ converges to zero as the sample size $n$ tends to infinity. This yields, in conjunction with Lemma 1.4.8 that the squared error risk of $\widehat{\gamma}_n(k)$ tends to zero which in turn means that the sequence of estimators $(\widehat{\gamma}_n(k))_{n\in\mathbb{N}}$ is consistent.

**Lemma 1.4.11.** *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary process with mean $\mu$ and autocovariance function $\gamma$. We assume that $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$, $E[X_0^4] < \infty$, and $\sum_{t,u,v=-\infty}^{\infty}\big|\operatorname{cum}\big(X_0-\mu, X_t-\mu, X_u-\mu, X_v-\mu\big)\big| < \infty$. Then*

$$E\big[\big(\widehat{\gamma}_n(k) \;-\; \gamma(k)\big)^2\big] \;=\; O\big(n^{-1}\big).$$

*Proof.* In order to employ Lemma 1.4.9, we represent $\widehat{\gamma}_n(k)$ as a quadratic form:

$$\begin{aligned}
\widehat{\gamma}_n(k) \;&=\; \frac{1}{n}\sum_{t=1}^{n-|k|}\big(X_{t+|k|}-\mu+\mu-\bar{X}_n\big)\big(X_t-\mu+\mu-\bar{X}_n\big) \\
&=\; \frac{1}{n}\sum_{t=1}^{n-|k|}\big(X_{t+|k|}-\mu\big)\big(X_t-\mu\big) \\
&\quad -\frac{1}{n}\sum_{t=1}^{n-|k|}\Big[\big(X_t-\mu\big)\big(\bar{X}_n-\mu\big) \;+\; \big(X_{t+|k|}-\mu\big)\big(\bar{X}_n-\mu\big)\Big] \\
&\quad +\frac{n-|k|}{n}\big(\bar{X}_n-\mu\big)^2 \\
&=\; Y^T M Y,
\end{aligned}$$

where $Y=(X_1-\mu,\ldots,X_n-\mu)^T$ and $M$ being an appropriate symmetric $(n\times n)$-matrix. The entries of $M$ are such that, for $k\neq 0$,

$$M_{st} \;=\; \begin{cases} \frac{1}{2n}+O(\frac{1}{n^2}) & \text{if } |s-t|=k, \\ O(\frac{1}{n^2}) & \text{if } |s-t|\neq k \end{cases}$$

and, for $k=0$,

$$M_{st} \;=\; \begin{cases} \frac{1}{n}+O(\frac{1}{n^2}) & \text{if } s=t, \\ O(\frac{1}{n^2}) & \text{if } s\neq t \end{cases}.$$

In either case, we obtain from Lemma 1.4.9

$$\begin{aligned}
\operatorname{var}\big(\widehat{\gamma}_n(k)\big) \;&=\; \sum_{s,t,u,v=1}^{n} M_{st}M_{uv}\operatorname{cum}\big(X_s-\mu, X_t-\mu, X_u-\mu, X_v-\mu\big) \\
&\quad + 2\sum_{s=1}^{n}\sum_{t,u,v=1}^{n} M_{st}\Sigma_{tu}M_{uv}\Sigma_{vs} \\
&\leq\; \underbrace{\max_{s,t}\{|M_{st}|\}}_{=O(n^{-1})}\;\underbrace{\max_{u,v}\{|M_{uv}|\}}_{=O(n^{-1})}\;\sum_{s=1}^{n}\underbrace{\sum_{t,u,v=1}^{n}\big|\operatorname{cum}\big(X_s-\mu, X_t-\mu, X_u-\mu, X_v-\mu\big)\big|}_{=O(1)} \\
&\quad + \underbrace{\max_{u,v}\{|M_{uv}|\}}_{=O(n^{-1})}\;\underbrace{\max_{s,t}\Big\{\sum_{u,v}|\Sigma_{tu}|\,|\Sigma_{vs}|\Big\}}_{=O(1)}\;\underbrace{\sum_{s,t=1}^{n}|M_{st}|}_{=O(1)} \\
&=\; O\big(n^{-1}\big).
\end{aligned}$$

□

## 1.5 ARMA processes

In this section we introduce an important class of processes $(X_t)_{t\in\mathbb{Z}}$ which are defined in terms of linear difference equations with coefficients that are constant over time.

**Definition.** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$. The process $(X_t)_{t\in\mathbb{Z}}$ is said to be an **autoregressive moving average process** of order $p,q$ (ARMA$(p,q)$ process) if for every $t\in\mathbb{Z}$

$$X_t \,-\, \alpha_1 X_{t-1} \,-\, \cdots \,-\, \alpha_p X_{t-p} \,=\, \varepsilon_t \,+\, \beta_1\varepsilon_{t-1} \,+\, \cdots \,+\, \beta_q\varepsilon_{t-q}. \qquad (1.5.1)$$

We say that $(X_t)_{t\in\mathbb{Z}}$ is an ARMA$(p,q)$ process with mean $\mu$ if $(X_t-\mu)_{t\in\mathbb{Z}}$ is an ARMA$(p,q)$ process.

Note that we do not require by definition that an ARMA process be stationary. As we will see below, the existence of a stationary solution to the systems of ARMA equations follows under some extra condition on the coefficients. Under such conditions, requiring stationarity is one way of making this solution unique.

The class of ARMA processes includes the following special cases.

- (**Moving-average process**)
  If $p=0$, then the system of model equations (1.5.1) reduces to

  $$X_t \,=\, \varepsilon_t \,+\, \beta_1\varepsilon_{t-1} \,+\, \cdots \,+\, \beta_q\varepsilon_{t-q} \qquad \forall t\in\mathbb{Z}. \qquad (1.5.2)$$

  The process $(X_t)_{t\in\mathbb{Z}}$ is said to be a **moving-average process** of order $q$ (or MA$(q)$ process). If $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$, then it follows from Proposition 1.4.3 that $(X_t)_{t\in\mathbb{Z}}$ is a stationary process and we obtain (defining $\beta_0=1$)

  $$EX_t \,=\, E\varepsilon_0\Big(\sum_{k=0}^{q}\beta_k\Big) \,=\, 0$$

  and

  $$\mathrm{cov}(X_{t+h},X_t) \,=\, \begin{cases} \sigma^2\sum_{j=0}^{q-|h|}\beta_j\beta_{j+|h|} & \text{if } |h|\le q, \\ 0 & \text{if } |h|<q. \end{cases}$$

- (**Autoregressive process**)
  If $q=0$, then we obtain the following system of model equations:

  $$X_t \,=\, \alpha_1 X_{t-1} \,+\, \cdots \,+\, \alpha_p X_{t-p} \,+\, \varepsilon_t \qquad \forall t\in\mathbb{Z}. \qquad (1.5.3)$$

  Such a process is called **autoregressive process** of order $p$ (or AR$(p)$ process). Note that, in contrast to the case of moving-average processes, the random variables $X_t$ are not explicitly given. The question whether or not there is a stationary solution $(X_t)_{t\in\mathbb{Z}}$ to the system of equations (1.5.3) will be discussed in what follows.

The imposition of the above additional structure leads to a class of models, the **autoregressive moving average** or **ARMA** processes, which are described by a finite number of parameters. Nevertheless, this class is quite flexible in matching a given autocovariance function. Indeed, for any autocovariance function $\gamma$ such that $\lim_{k\to\infty}\gamma(k)=0$, and for any integer $p>0$, it is possible to find an autoregressive process $(X_t)_{t\in\mathbb{Z}}$ of order $p$ with autocovariance function $\gamma_X$ such that $\gamma_X(h) = \gamma(h)$, for $h = 0, \ldots, p$. Another nice feature of these ARMA processes is that they can be represented as linear processes which makes an immediate application of tools, which are originally derived in the context of linear processes, possible.

As some sort of warm-up, we consider first the case of an autoregressive process of order 1. Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 > 0$. The system of model equations is

$$X_t \,=\, \alpha_1 X_{t-1} \,+\, \varepsilon_t \qquad \forall t \in \mathbb{Z}. \tag{1.5.4}$$

Iterating (1.5.4) we obtain that

$$\begin{aligned} X_t &= \varepsilon_t \,+\, \alpha_1 X_{t-1} \\ &= \varepsilon_t \,+\, \alpha_1 \varepsilon_{t-1} \,+\, \alpha_1^2 X_{t-2} \\ &= \ldots \\ &= \varepsilon_t \,+\, \alpha_1 \varepsilon_{t-1} \,+\, \cdots \,+\, \alpha_1^k \varepsilon_{t-k} \,+\, \alpha_1^{k+1} X_{t-k-1}, \end{aligned}$$

which leads us to a guess of a solution. Indeed, if $|\alpha_1| < 1$, then we obtain from Proposition 1.4.2 that the series $\sum_{k=0}^\infty \alpha_1^k \varepsilon_{t-k}$ is both mean square convergent and absolutely convergent with probability one. Furthermore, $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ with $\widetilde{X}_t = \sum_{k=0}^\infty \alpha_1^k \varepsilon_{t-k}$ is by Proposition 1.4.3 a stationary process and we obtain that

$$\widetilde{X}_t \,=\, \sum_{k=0}^\infty \alpha_1^k \varepsilon_{t-k} \,=\, \varepsilon_t \,+\, \alpha_1 \sum_{t=0}^\infty \alpha_1^k \varepsilon_{t-1-k} \,=\, \alpha_1 \widetilde{X}_{t-1} \,+\, \varepsilon_t$$

holds for all $t \in \mathbb{Z}$. Therefore, $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ is a stationary and **causal** (since only $\varepsilon_s$ with $s \leq t$ are involved in the definition of $\widetilde{X}_t$) solution to (1.5.4). The issues of stationarity of an AR(1) process are summarized in the following proposition.

**Proposition 1.5.1.** *Let* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ *and* $\alpha_1 \in \mathbb{R}$ *with* $|\alpha_1| < 1$.

*(i)* *A stationary solution to (1.5.4) is given by* $(\widetilde{X}_t)_{t\in\mathbb{Z}}$*, where*

$$\widetilde{X}_t \,=\, \sum_{k=0}^\infty \alpha_1^k \, \varepsilon_{t-k}. \tag{1.5.5}$$

*(ii)* *a)* $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ *is the unique weakly stationary solution which satisfies (1.5.4) for all* $t \in \mathbb{Z}$.

*b)* *Let* $(X_t)_{t\in\mathbb{N}_0}$ ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) *be an arbitrary process which satisfies (1.5.4) for all* $t \in \mathbb{N}$. *Then*

$$X_t \,-\, \widetilde{X}_t \,=\, \alpha_1^t \big(X_0 - \widetilde{X}_0\big)$$

*and*

$$\big|X_t - \widetilde{X}_t\big| \;\stackrel{a.s.}{\longrightarrow}\; 0 \qquad \textit{as } t \to \infty.$$

*If* $(X_t)_{t\in\mathbb{N}_0}$ *is weakly stationary, then*

$$E\Big[\big(X_t - \widetilde{X}_t\big)^2\Big] \,\leq\, 2\alpha_1^{2t} \big(EX_0^2 \,+\, E\widetilde{X}_0^2\big).$$

*Proof.* After the preceding discussion, it only remains to prove part (ii).

a) Let $(X_t)_{t\in\mathbb{Z}}$ be an arbitrary solution to (1.5.4). Then

$$X_t - \widetilde{X}_t = \alpha_1\big(X_{t-1} - \widetilde{X}_{t-1}\big) = \ldots = \alpha_1^k\big(X_{t-k} - \widetilde{X}_{t-k}\big)$$

holds for all $k \in \mathbb{N}$. Since both $(X_t)_{t\in\mathbb{Z}}$ and $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ are stationary processes we obtain hat

$$\begin{aligned} E\big|X_t - \widetilde{X}_t\big| &\le \big|\alpha_1\big|^k\, E\Big[\big|X_{t-k}\big| + \big|\widetilde{X}_{t-k}\big|\Big] \\ &\le \big|\alpha_1\big|^k \left\{\sqrt{E\big[X_{t-k}^2\big]} + \sqrt{E\big[\widetilde{X}_{t-k}^2\big]}\right\} \underset{k\to\infty}{\longrightarrow} 0. \end{aligned}$$

This means that $E|X_t - \widetilde{X}_t| = 0$, which implies that

$$P\big(X_t \neq \widetilde{X}_t\big) = 0.$$

b) Let $(X_t)_{t\in\mathbb{N}_0}$ be an arbitrary process which satisfies (1.5.4) for all $t \in \mathbb{N}$. Then we conclude as above that

$$X_t - \widetilde{X}_t = \alpha_1^t\big(X_0 - \widetilde{X}_0\big) \overset{a.s.}{\longrightarrow} 0 \qquad \text{as } t \to \infty.$$

If in addition $(X_t)_{t\in\mathbb{N}_0}$ is weakly stationary, then

$$E\big(X_t - \widetilde{X}_t\big)^2 \le 2\alpha_1^{2t}\,\big(EX_0^2 + E\widetilde{X}_0^2\big).$$

□

**Remark 1.5.2.** *For a one-sided AR(1) process $(X_t)_{t\in\mathbb{N}_0}$ it is only required that (1.5.4) is satisfied for all $t \ge 1$. In this case, a stationary solution to (1.5.4) may not be unique. Indeed, suppose that the underlying probability space allows the construction of two independent processes $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0,\sigma^2)$ and $(\varepsilon'_t)_{t\in\mathbb{Z}} \sim WN(0,\sigma^2)$. Let*

$$\bar{\varepsilon}_t := \begin{cases} \varepsilon_t & \text{if } t \ge 1, \\ \varepsilon'_t & \text{if } t \le 0. \end{cases}$$

*Then*

$$\widetilde{X}_t = \sum_{k=0}^{\infty} \alpha_1^k \varepsilon_{t-k}$$

*and*

$$\bar{X}_t = \sum_{k=0}^{\infty} \alpha_1^k \bar{\varepsilon}_{t-k}$$

*are both solutions to (1.5.4) for all $t \in \mathbb{N}$. However, unless $\sigma^2 = 0$, these processes are not equal.*

Before we investigate autoregressive processes of order $p \geq 1$ we stick to $p = 1$ and take a brief look at the case of $|\alpha_1| \geq 1$.

**The case of** $|\alpha_1| > 1$
Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 > 0$. We seek again a stationary solution to the system of equations (1.5.4), for $t \in \mathbb{Z}$. Since $\sigma^2 > 0$ it is obvious that the series $\sum_{k=0}^{\infty} \alpha_1^k \varepsilon_{t-k}$ does not converge in mean square. On the other hand, the following system of equations is equivalent to (1.5.4):

$$X_{t-1} \;=\; (1/\alpha_1)X_t \;-\; (1/\alpha_1)\varepsilon_t \qquad \forall t \in \mathbb{Z}. \tag{1.5.6}$$

To guess a solution to (1.5.5) we iterate this equation and obtain

$$\begin{aligned} X_t &= -(1/\alpha_1)\varepsilon_{t+1} \;+\; (1/\alpha_1)X_{t+1} \\ &= -(1/\alpha_1)\varepsilon_{t+1} \;-\; (1/\alpha_1)^2\varepsilon_{t+2} \;+\; (1/\alpha_1)^2 X_{t+2} \\ &= \ldots \\ &= -\sum_{k=1}^{\infty} (1/\alpha_1)^k \varepsilon_{t+k}. \end{aligned} \tag{1.5.7}$$

It follows from Proposition 1.4.2 that the series $\sum_{k=1}^{\infty}(1/\alpha_1)^k\varepsilon_{t+k}$ converges both in mean square and with probability one. Furthermore, $(X_t)_{t\in\mathbb{Z}}$ with $X_t = -\sum_{k=1}^{\infty}(1/\alpha_1)^k\varepsilon_{t+k}$ is by Proposition 1.4.3 a stationary process. This process satisfies (1.5.6) and, therefore, (1.5.4) as well.

**The case of** $\alpha_1 = \pm 1$
In this case, a stationary solution to (1.5.4) does not exist. Suppose again that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 > 0$. If, for example, $\alpha_1 = 1$, then any stationary solution $(X_t)_{t\in\mathbb{Z}}$ has to fulfill

$$X_t \;=\; \varepsilon_t \;+\; X_{t-1} \;=\; \ldots \;=\; \varepsilon_t \;+\; \cdots \;+\; \varepsilon_{t-(k-1)} \;+\; X_{t-k} \quad \forall k \in \mathbb{N}.$$

Now we have $\mathrm{var}(X_t) = \mathrm{var}(X_{t-k}) < \infty$ but

$$\mathrm{var}\left(\varepsilon_t \;+\; \cdots \;+\; \varepsilon_{t-(k-1)}\right) \;=\; k\,\sigma^2 \underset{k\to\infty}{\longrightarrow} \infty,$$

which leads to a contradiction.

We would like to note that the stationary solution (1.5.7) is frequently regarded as unnatural. When autoregressive processes are employed to model real-world phenomena such as the evolution of stock prices, the $\varepsilon_t$ usually describe the effect of external shocks which influence the further evolution of the stock price. Since such shocks are usually unforeseeable it does not make sense to include $\varepsilon_s$ for any $s > t$ in the definition of $X_t$. It is customary therefore when modelling stationary time series to restrict attention to AR(1) processes with $|\alpha_1| < 1$ for which the unique stationary solution has the representation (1.5.5) in terms of $(\varepsilon_s)_{s\leq t}$. This also applies to autoregressive processes of higher order which will be investigated below.

## Exercises

**Ex. 1.5.1** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 > 0$. Show that there is no stationary solution $(X_t)_{t\in\mathbb{Z}}$ to (1.5.4) if $\alpha_1 = -1$.

**Ex. 1.5.2** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 > 0$ and that $\alpha_1 \in \mathbb{R}$ such that $|\alpha_1| > 1$.
Show that the (non-causal) solution (1.5.7) also satisfies the AR(1) equations

$$X_t \;=\; (1/\alpha_1)X_{t-1} \,+\, \widetilde{\varepsilon}_t \quad \forall t \in \mathbb{Z},$$

for a suitably chosen process $(\widetilde{\varepsilon}_t)_{t\in\mathbb{Z}}$.
Show that $(\widetilde{\varepsilon}_t)_{t\in\mathbb{Z}}$ is a white noise and determine $\mathrm{var}(\widetilde{\varepsilon}_t)$.

Now we turn to autoregressive processes of higher order. Assume again that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$. The system of model equations is given by (1.5.3). Again we would like to know if (1.5.3) has a causal stationary solution. Suppose for the time being that $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ is such a solution. We obtain by a repeated iteration that the following equation has to be satisfied for all $t \in \mathbb{Z}$.

$$\begin{aligned}
\widetilde{X}_t &= \varepsilon_t + \alpha_1 \widetilde{X}_{t-1} + \cdots + \alpha_p \widetilde{X}_{t-p} \\
&= \varepsilon_t + \alpha_1\Big(\varepsilon_{t-1} + \alpha_1 \underbrace{\widetilde{X}_{t-2}}_{=\varepsilon_{t-2}+\cdots} + \cdots + \alpha_p \widetilde{X}_{t-p-1}\Big) + \alpha_2 \underbrace{\widetilde{X}_{t-2}}_{=\varepsilon_{t-2}+\cdots} + \alpha_3 \widetilde{X}_{t-3} + \cdots + \alpha_p \widetilde{X}_{t-p} \\
&= \ldots \\
&= \varepsilon_t + \sum_{k=1}^{\infty} \Big( \sum_{r=1}^{k} \sum_{(k_1,\ldots,k_r):\, k_1+\cdots+k_r=k} \alpha_{k_1}\cdots\alpha_{k_r} \Big) \varepsilon_{t-k}. \qquad (1.5.8)
\end{aligned}$$

At this point we want to find a condition on the coefficients $\alpha_1, \ldots, \alpha_p$ which ensures that the series on the right-hand side of (1.5.8) actually converges. After that we will check that the corresponding process solves (1.5.3). Suppose that

$$|\alpha_1| + \cdots + |\alpha_p| < 1. \qquad (1.5.9)$$

Then

$$\begin{aligned}
1 + \sum_{k=1}^{\infty}\sum_{r=1}^{k} \sum_{(k_1,\ldots,k_r):\, k_1+\cdots+k_r=k} |\alpha_{k_1}|\cdots|\alpha_{k_r}| &= 1 + \sum_{r=1}^{\infty}\sum_{k\geq r} \sum_{(k_1,\ldots,k_r):\, k_1+\cdots+k_r=k} |\alpha_{k_1}|\cdots|\alpha_{k_r}| \\
&= \sum_{r=0}^{\infty} \big(|\alpha_1| + \cdots + |\alpha_p|\big)^r \\
&= \frac{1}{1-|\alpha_1|-\cdots-|\alpha_p|} < \infty.
\end{aligned}$$

Therefore, it follows from Proposition 1.4.2 that the series on the right-hand side of (1.5.8) converges absolutely with probability one. This allows us in particular to alter the order of summation and we obtain that

$$\begin{aligned}
\widetilde{X}_t &= \varepsilon_t + \sum_{k=1}^{\infty} \Big( \sum_{r\leq k,\, k_1+\cdots+k_r=k} \alpha_{k_1}\cdots\alpha_{k_r} \Big) \varepsilon_{t-k} \\
&= \varepsilon_t + \sum_{k_1=1}^{p} \alpha_{k_1} \underbrace{\Big( \varepsilon_{t-k_1} + \sum_{k>k_1} \Big( \sum_{r\leq k,\, k_2+\cdots+k_r=k-k_1} \alpha_{k_2}\cdots\alpha_{k_r} \Big) \varepsilon_{(t-k_1)-k_2-\cdots-k_r} \Big)}_{=\widetilde{X}_{t-k_1}}.
\end{aligned}$$

This shows that $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ solves (1.5.3). We can also write this solution in a more compact form. Let $\beta_0 = 1$ and, for $k \in \mathbb{N}$,

$$\beta_k = \sum_{r\leq k,\, k_1+\cdots+k_r=k} \alpha_{k_1}\cdots\alpha_{k_r}. \qquad (1.5.10)$$

Then

$$\widetilde{X}_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}.$$

Since $\sum_{k=0}^{\infty} |\beta_k| < 1 + \sum_{k=1}^{\infty} \sum_{r\leq k,\, k_1+\cdots+k_r=k} |\alpha_{k_1}|\cdots|\alpha_{k_r}| < \infty$ it is clear that the conditions of Propositions 1.4.2 and 1.4.3 are fulfilled. The process $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ is therefore stationary. We will see in the following that condition (1.5.9) is stronger than necessary.

In the following we intend to relax condition (1.5.9) which guarantees that the system of linear difference equations (1.5.3) has a causal stationary solution $\sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$ with $\sum_{k=0}^{\infty} |\beta_k| < \infty$. (The latter condition ensures that the infinite series converges both in mean square and absolutely with probability one.) Recall that, for a given sequence of innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, the following equations have to be solved simultaneously for all $t \in \mathbb{Z}$.

$$X_t \,-\, \alpha_1 X_{t-1} \,-\, \cdots \,-\, \alpha_p X_{t-p} \,=\, \varepsilon_t \tag{1.5.11a}$$

and

$$X_t \,=\, \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k} \tag{1.5.11b}$$

To avoid this somewhat bulky notation, we introduce the so-called **backward shift operator** $B$ as

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

Powers of the operator are defined in the obvious way, i.e. $B^0 X_t = X_t$ and $B^k X_t = X_{t-k}$, $k \geq 1$. We can rewrite (1.5.11a) and (1.5.11b) in a more compact way,

$$\alpha(B) X_t \,=\, \varepsilon_t \tag{1.5.12a}$$

and

$$X_t \,=\, \beta(B) \varepsilon_t, \tag{1.5.12b}$$

where $\alpha(B) = B^0 - \alpha_1 B^1 - \cdots - \alpha_p B^p$ and $\beta(B) = \sum_{k=0}^{\infty} \beta_k B^k$. Plugging in equation (1.5.12b) into (1.5.12a) we obtain that

$$\alpha(B)\beta(B)\varepsilon_t \,=\, \varepsilon_t. \tag{1.5.13}$$

This is actually fulfilled if a comparison of coefficients yields that

$$\alpha(B)\beta(B) \,=\, B^0 \tag{1.5.14}$$

and, of course, if the infinite series on the left-hand side of (1.5.13) converges absolutely. The latter requirement if fulfilled if $\sum_{k=0}^{\infty} |\beta_k| < \infty$, what we keep in mind in what follows. For given $\alpha_1, \ldots, \alpha_p$, it **remains to solve** (1.5.14). But this could be equally well done by solving an equation with polynomials with real arguments. (1.5.14) is equivalent to

$$\alpha(z)\beta(z) \,=\, 1 \qquad \forall z \in I, \tag{1.5.15}$$

where $I \subseteq \mathbb{R}$ is some non-empty interval, $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p$, $\beta(z) = \sum_{k=0}^{\infty} \beta_k z^k$. Coefficients $(\beta_k)_{k\in\mathbb{N}}$ which solve (1.5.15) are candidates for a possible solution to (1.5.14). Moreover, it will turn out that $\sum_{k=0}^{\infty} |\beta_k| < \infty$ follows if the polynomial $\alpha$ has all of its zeroes outside the unit circle. The following lemma provides a sufficient and necessary condition for the existence of a solution to (1.5.15).

**Lemma 1.5.3.** *Let* $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p$*, where* $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$*. Then there exist* $(\beta_k)_{k\in\mathbb{N}}$ *such that* $\sum_{k=0}^{\infty} |\beta_k| < \infty$ *and*

$$\alpha(z) \sum_{k=0}^{\infty} \beta_k z^k \,=\, 1 \qquad \forall z \in \mathbb{C} \text{ with } |z| \leq 1$$

*if and only if*

$$\alpha(z) \,\neq\, 0 \qquad \forall z \in \mathbb{C} \text{ with } |z| \leq 1.$$

*In this case, the coefficients* $\beta_0, \beta_1, \ldots$ *are real.*

*Proof.* ($\Longleftarrow$) Suppose that

$$\alpha(z) \neq 0 \qquad \forall z \in \mathbb{C} \text{ with } |z| \leq 1.$$

The polynomial $\alpha$ can be written as

$$\alpha(z) = \big(1 - z/c_1\big) \cdots \big(1 - z/c_p\big),$$

where $c_1, \ldots, c_p$ are the zeroes of $\alpha$, according to their multiplicities. Since $1/(1-z/c) = \sum_{k=0}^{\infty} (z/c)^k$ holds for all $z \in \mathbb{C}$ with $|z/c| < 1$ we obtain, for these values of $z$,

$$\begin{aligned} \frac{1}{\alpha(z)} &= \frac{1}{1 - z/c_1} \cdots \frac{1}{1 - z/c_p} \\ &= \Big( \sum_{k_1=0}^{\infty} c_1^{-k_1} z^{k_1} \Big) \cdots \Big( \sum_{k_p=0}^{\infty} c_p^{-k_p} z^{k_p} \Big), \end{aligned} \tag{1.5.16}$$

where all series on the right-hand side converge absolutely. In order to obtain that $\alpha(z)\beta(z) = 1$ we choose $(\beta_k)_{k \in \mathbb{N}_0}$ such that

$$\sum_{k=0}^{\infty} \beta_k z^k = \Big( \sum_{k_1=0}^{\infty} c_1^{-k_1} z^{k_1} \Big) \cdots \Big( \sum_{k_p=0}^{\infty} c_p^{-k_p} z^{k_p} \Big).$$

A comparison of coefficients reveals that this is achieved by the choice $\beta_k = \sum_{k_1,\ldots,k_p \geq 0:\ k_1+\cdots+k_p=k} c_1^{-k_1} \cdots c_p^{-k_p}$. Furthermore, since the power series on the right-hand side converge absolutely for $z = 1$ we obtain $\sum_{k=0}^{\infty} |c_i^{-k}| < \infty$ and therefore $\sum_{k=0}^{\infty} |\beta_k| \leq \big( \sum_{k_1=0}^{\infty} |c_1^{-k_1}| \big) \cdots \big( \sum_{k_p=0}^{\infty} |c_p^{-k_p}| \big) < \infty$.
To see that the coefficients $\beta_0, \beta_1, \ldots$ are real, write $\beta_k = \beta_k^R + i\beta_k^I$, where $\beta_k^R, \beta_k^I \in \mathbb{R}$. Then

$$\alpha(z) \Big( \sum_{k=0}^{\infty} \beta_k^I z^k \Big) = \Im \Big( \underbrace{\alpha(z) \sum_{k=0}^{\infty} \beta_k z^k}_{=1} \Big) = 0 \qquad \forall z \in \mathbb{C}, |z| \leq 1$$

which implies that $\sum_{k=0}^{\infty} \beta_k^I z^k = 0\ \forall z \in \mathbb{C},\ |z| \leq 1$, and, hence, $\beta_k^I = 0\ \forall k \geq 0$.

($\Longrightarrow$) This direction is trivial. Indeed, if $\alpha(z) \sum_{k=0}^{\infty} \beta_k z^k = 1$ for all $z \in \mathbb{C}$, $|z| \leq 1$, then it is clear that $\alpha(z) \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$. □

It is an immediate corollary of Proposition 1.4.2 that operators such as $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$ with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, when applied to a stationary process $(Z_t)_{t \in \mathbb{Z}}$, are not only meaningful but also inherit the algebraic properties of power series. In particular, if $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$, $\sum_{j=-\infty}^{\infty} |\beta_j| < \infty$, $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$, $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$, and $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$, where

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k},$$

then $\alpha(B)\beta(B)Z_t$ is well-defined and

$$\alpha(B)\beta(B)Z_t = \beta(B)\alpha(B)Z_t = \psi(B)Z_t.$$

Therefore, the following theorem is mainly a direct consequence of the previous Lemma 1.5.3.

**Theorem 1.5.4.** *Let* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *be a stationary process and suppose that*

$$\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0 \qquad \forall z \in \mathbb{C}, |z| \leq 1. \tag{1.5.17}$$

*(i)* *The system of equations*

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \qquad \forall t \in \mathbb{Z} \tag{1.5.18}$$

*has a stationary solution* $(\widetilde{X}_t)_{t\in\mathbb{Z}}$, *where* $\widetilde{X}_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$, $\beta_k = \sum_{k_1,\dots,k_p\geq 0:\ k_1+\cdots+k_p=k} c_1^{-k_1}\cdots c_p^{-k_p}$, *and* $c_1,\dots,c_p$ *are the zeroes of* $\alpha$.

*(ii)* *If* $(X_t)_{t\in\mathbb{Z}}$ *is an arbitrary stationary solution to (1.5.18), then*

$$P\big(X_t = \widetilde{X}_t\big) = 1.$$

*Proof.* (i) It follows from Lemma 1.5.3 that $\sum_{k=0}^{\infty} |\beta_k| < \infty$. Therefore, $\sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$ converges both in mean square and absolutely with probability one which means that $\widetilde{X}_t$ is well-defined. According to Proposition 1.4.3, $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ inherits the property of stationarity from the underlying innovation process $(\varepsilon_t)_{t\in\mathbb{Z}}$. Furthermore, $\alpha(z)\beta(z) = 1\ \forall |z| \leq 1$ is equivalent to $\alpha(B)\beta(B) = B^0$. Hence,

$$\alpha(B)\widetilde{X}_t = \underbrace{\alpha(B)\beta(B)}_{=B^0}\varepsilon_t = \varepsilon_t,$$

i.e., $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ solves (1.5.18).

(ii) We have that $\big(\alpha(B)\widetilde{X}_t\big)_{t\in\mathbb{Z}}$ and $\big(\alpha(B)X_t\big)_{t\in\mathbb{Z}}$ are both stationary processes. Therefore $\big(\beta(B)\alpha(B)\widetilde{X}_t\big)_{t\in\mathbb{Z}}$ and $\big(\beta(B)\alpha(B)X_t\big)_{t\in\mathbb{Z}}$ are also stationary and it follows from $\alpha(B)\widetilde{X}_t = \varepsilon_t = \alpha(B)X_t$ that

$$\underbrace{\beta(B)\alpha(B)\widetilde{X}_t}_{=\widetilde{X}_t} = \beta(B)\varepsilon_t = \underbrace{\beta(B)\alpha(B)X_t}_{=X_t} \qquad P-a.s.$$

Hence, $X_t$ and $\widetilde{X}_t$ are equal with probability one. □

To conclude these considerations, we want to clarify how the regularity conditions (1.5.9) and (1.5.17) are related to each other.

**Remark 1.5.5.** *(i) If $|\alpha_1| + \cdots + |\alpha_p| < 1$, then $\alpha(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. The converse statement is not true in general.*

*(ii) If $\alpha_1, \ldots, \alpha_p$ are non-negative, then (1.5.9) and (1.5.17) are equivalent.*

*Proof.* (i) Suppose that $|\alpha_1| + \cdots + |\alpha_p| < 1$ and $z \in \mathbb{C}$, $|z| \leq 1$. Then

$$\left|\alpha_1 z^1 + \cdots + \alpha_p z^p\right| < 1,$$

which implies that

$$\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0.$$

To disprove the converse, we consider a simple counterexample ($p = 2$). Let $c_1$ and $c_2$ be the zeroes of the autoregressive polynomial. Then

$$\alpha(z) = \left(1 - \frac{z}{c_1}\right)\left(1 - \frac{z}{c_2}\right) = 1 - \underbrace{\left(\frac{1}{c_1} + \frac{1}{c_2}\right)}_{=\alpha_1} z + \frac{1}{c_1\, c_2} z^2.$$

For $c_1, c_2 \in \mathbb{R}$ with $1 < c_i < 2$, we obtain that $|\alpha_1| > 1$ but (1.5.17) is satisfied
(ii) The second statement is trivial. □

In the following we show how flexible the class of autoregressive processes is in matching a given structure of the autocovariances. We show in particular that, for any autocovariance function $\gamma$ such that $\gamma(0) > 0$ and $\gamma(k) \longrightarrow_{k\to\infty} 0$, we can find a causal stationary AR($p$) process with autocovariance function $\gamma_X$ such that $\gamma_X(k) = \gamma(k)$ for $k = 0, \ldots, p$. To this end, we establish first an important relation between the parameters of an AR($p$) process and the autocovariances.

Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a zero-mean stationary and causal autoregressive process of order $p$ obeying

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \quad \forall t \in \mathbb{Z},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$ and $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$. It follows from the uniqueness of the stationary solution stated in Theorem 1.5.4 that $X_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$, where the sequence of coefficients $(\beta_k)_{k\in\mathbb{N}_0}$ is determined by $\sum_{k=0}^{\infty} \beta_k z^k = 1/\alpha(z)$, $|z| \leq 1$. We have in particular $\beta_0 = 1$ and $\sum_{k=0}^{\infty} |\beta_k| < \infty$. It follows from continuity of the inner product (see Lemma 1.3.4) that

$$E[\varepsilon_t X_0] = \lim_{m\to\infty} E\big[\varepsilon_t \big(\sum_{k=0}^{m} \beta_k \varepsilon_{-k}\big)\big] = \begin{cases} 0 & \text{if } t > 0, \\ \sigma^2 & \text{if } t = 0. \end{cases}$$

Let $\gamma$ be the autocovariance function of the process $(X_t)_{t\in\mathbb{Z}}$. For $t = 0, 1, \ldots, p$, we obtain the following equations which are called **Yule-Walker equations**:

$$\begin{aligned} \gamma(t) &= E\big[X_t\, X_0\big] = E\Big[\big(\sum_{k=1}^{p} \alpha_k X_{t-k} + \varepsilon_t\big) X_0\Big] \\ &= \sum_{k=1}^{p} \alpha_k \gamma(t-k) + E[\varepsilon_t\, X_0]. \end{aligned}$$

The Yule-Walker equations can be condensed in matrix/vector notation as

$$\underbrace{\begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}}_{=:\gamma_p} = \underbrace{\begin{pmatrix} \gamma(1-1) & \dots & \gamma(1-p) \\ \vdots & \ddots & \vdots \\ \gamma(p-1) & \dots & \gamma(p-p) \end{pmatrix}}_{=:\Gamma_p} \underbrace{\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}}_{=:\alpha} \tag{1.5.19}$$

and

$$\gamma(0) \;=\; \gamma_p^T \alpha \,+\, \sigma^2. \tag{1.5.20}$$

These equations can be used to determine $\gamma(0), \dots, \gamma(p)$ from $\sigma^2$ and $\alpha_1, \dots, \alpha_p$. On the other hand, if we replace the autocovariances $\gamma(j)$, $j = 0, \dots, p$, appearing in (1.5.19) and (1.5.20) by the corresponding sample autocovariances $\widehat{\gamma}(j)$, we obtain a set of equations for the so-called **Yule-Walker estimators** $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ and $\widehat{\sigma}^2$ of $\alpha_1, \dots, \alpha_p$ and $\sigma^2$, respectively. And finally, as shown below, these equations may also be used to find, for a given set of autocovariances $\gamma(0), \dots, \gamma(p)$, parameters $\alpha_1, \dots, \alpha_p$ and $\sigma^2$ for an AR($p$) process with these autocovariances. To this end, we state first a result that guarantees that the matrix $\Gamma_p = \big(\gamma(i-j)\big)_{i,j=1,\dots,p}$ is regular which means that, for given $\gamma(0), \dots, \gamma(p)$, (1.5.19) has always a solution.

**Lemma 1.5.6.** *Let $\gamma$ be the autocovariance function of a weakly stationary, real-valued process $(X_t)_{t\in\mathbb{Z}}$.*
*If $\gamma(0) > 0$ and $\gamma(k) \longrightarrow_{k\to\infty} 0$, then the covariance matrix $\Gamma_n = \big(\gamma(i-j)\big)_{i,j=1,\dots,n}$ is regular for all $n \in \mathbb{N}$.*

*Proof.* We prove this result by contradiction. Let $\widetilde{X}_t := X_t - EX_t$. It is clear that $\Gamma_1 = \big(\gamma(0)\big)$ is regular. Assume that $\Gamma_n$ is regular and $\Gamma_{n+1}$ singular. Then since $E\widetilde{X}_t = 0$ there exists $a = (a_1, \dots, a_{n+1})^T \neq 0_{n+1}$ such that

$$0 \;=\; a^T \Gamma_{n+1} a \;=\; a^T E\left[\begin{pmatrix} \widetilde{X}_1 \\ \vdots \\ \widetilde{X}_{n+1} \end{pmatrix} \Big(\widetilde{X}_1 \cdots \widetilde{X}_n\Big)\right] a \;=\; E\left[\left(a^T \begin{pmatrix} \widetilde{X}_1 \\ \vdots \\ \widetilde{X}_{n+1} \end{pmatrix}\right)^2\right],$$

which implies that $\sum_{i=1}^{n+1} a_i \widetilde{X}_i \;=\; 0$ holds $P$-almost surely. Moreover, we have that $a_{n+1} \neq 0$ since otherwise $\Gamma_n$ would be singular. Hence,

$$\widetilde{X}_{n+1} \;=\; \sum_{i=1}^{n} \underbrace{(-a_i/a_{n+1})}_{=:\,d_i} \widetilde{X}_i \qquad P-a.s.,$$

which means in particular that $E\big[\big(\widetilde{X}_{n+1} \,-\, \sum_{i=1}^n d_i \widetilde{X}_i\big)^2\big] = 0$. By stationarity we then have $E\big[\big(\widetilde{X}_{n+k+1} \,-\, \sum_{i=1}^n d_i \widetilde{X}_{i+k}\big)^2\big] = 0$, i.e.

$$\widetilde{X}_{n+k+1} \;=\; \sum_{i=1}^{n} d_i \widetilde{X}_{i+k} \qquad P-a.s., \qquad \text{for all } k \in \mathbb{N}.$$

Consequently, for all $m \geq n+1$, there exist constants $d_1^{(m)}, \dots, d_n^{(m)}$ such that

$$\widetilde{X}_m \;=\; \sum_{i=1}^{n} d_i^{(m)} \widetilde{X}_i \qquad P-a.s.$$

We have

$$\begin{aligned}
\gamma(0) &= \operatorname{var}(\bar{X}_m) = \operatorname{var}\big(\sum_{i=1}^n d_i^{(m)} \widetilde{X}_i\big) \\
&= (d_1^{(m)}, \dots, d_n^{(m)}) \Gamma_n \begin{pmatrix} d_1^{(m)} \\ \vdots \\ d_n^{(m)} \end{pmatrix} \\
&\geq \lambda_{\min}(\Gamma_n) \sum_{i=1}^n (d_i^{(m)})^2,
\end{aligned}$$

where $\lambda_{\min}(\Gamma_n)$ denotes the smallest eigenvalue of $\Gamma_n$. Since $\Gamma_n$ is regular and, as a covariance matrix, non-negative definite, it follows that $\lambda_{\min}(\Gamma_n)$ is strictly positive. This shows that, for each fixed $i$, $d_i^{(m)}$ is a bounded function of $m$. On the other hand,

$$\begin{aligned}
0 < \gamma(0) &= \operatorname{cov}\big(\widetilde{X}_m, \sum_{i=1}^n d_i^{(m)} \widetilde{X}_i\big) \\
&\leq \sum_{i=1}^n |d_i^{(m)}| |\gamma(m-i)| \underset{m\to\infty}{\longrightarrow} 0.
\end{aligned}$$

This is a contradiction and our assumption that $\Gamma_n$ is singular must be wrong. This completes the proof. □

**Corollary 1.5.7.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary and causal AR(p) process,*

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \qquad \forall t \in \mathbb{Z},$$

*where $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0,\sigma^2)$ with $\sigma^2 > 0$, $\alpha(z) \neq 0$ for all $|z| \leq 1$.*
*Then the autocovariance function $\gamma_X$ of this process fulfills $\gamma_X(0) > 0$ and $\sum_{k=0}^\infty |\gamma_X(k)| < \infty$, which implies that $\gamma_X(k) \longrightarrow_{k\to\infty} 0$. Therefore, the corresponding covariance matrices $\Gamma_n$ are regular for all $n \in \mathbb{N}$.*

Now we are in a position to prove that an arbitrary autocovariance function can be approximated by the autocovariance function of a suitable autoregressive process.

**Theorem 1.5.8.** *Suppose that $\gamma$ is the autocovariance function of a stationary (real-valued) process such that $\gamma(0) > 0$ and $\gamma(k) \longrightarrow_{k\to\infty} 0$.*
*Then there exists a causal stationary AR(p) process $(X_t)_{t\in\mathbb{Z}}$ with autocovariance function $\gamma_X$ such that $\gamma_X(k) = \gamma(k)$ for all $k = 0, 1, \dots, p$.*

*Proof.* We show that there exists a causal stationary process $(X_t)_{t\in\mathbb{Z}}$ such that

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \qquad \forall t \in \mathbb{Z},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$ and $\operatorname{cov}(X_{t+k}, X_t) = \gamma(k)$, $k = 0, 1, \dots, p$.
If there exists such a process at all, then it follows from the Yule-Walker equations (1.5.19) and (1.5.20) that

$$\big(\alpha_1, \dots, \alpha_p\big)^T = \Gamma_p^{-1} \gamma_p$$

and

$$\sigma^2 \;=\; \gamma(0) \;-\; \gamma_p^T \alpha,$$

where $\Gamma_p = \big(\gamma(i-j)\big)_{i,j=1}^p$, $\gamma_p = (\gamma(1),\ldots,\gamma(p))^T$, and $\alpha = (\alpha_1,\ldots,\alpha_p)^T$. It only remains to show that

$$\alpha(z) \;=\; 1 \;-\; \alpha_1 z^1 \;-\; \cdots \;-\; \alpha_p z^p \;\neq\; 0 \qquad \forall z \in \mathbb{C}, |z| \leq 1. \tag{1.5.21}$$

We prove this by contradiction and assume that the autoregressive polynomial $\alpha$ has a zero inside the unit circle, i.e.

$$\alpha(z) \;=\; \big(1 \;-\; z/c\big)\,\xi(z),$$

for some $c \in \mathbb{C}$, $|c| \leq 1$. Then the polynomial $\xi$ has the form $\xi(z) = 1 - \sum_{j=1}^{p-1} b_j z^j$, for some $b_1,\ldots,b_{p-1} \in \mathbb{C}$.
Let $\big(Y_1,\ldots,Y_{p+1}\big)^T \sim N\big(0_{p+1}, \Gamma_{p+1}\big)$. We define $Z_j := \xi(B) Y_j$ and

$$\rho \;:=\; \frac{\langle Z_{p+1}, Z_p\rangle}{\|Z_{p+1}\|\,\|Z_p\|}.$$

Note that the random variable $Z_j$ is not necessarily real-valued, it could also be complex-valued. We define the polynomial

$$\gamma(B) \;:=\; \big(B^0 \;-\; \rho B\big)\xi(B),$$

which can also be written as

$$\gamma(B) \;=\; B^0 \;-\; \sum_{j=1}^p \widetilde{\alpha}_j B^j,$$

for some $\widetilde{\alpha}_1,\ldots,\widetilde{\alpha}_p \in \mathbb{C}$. We have

$$\begin{aligned} E\Big|Y_{p+1} \;-\; \sum_{j=1}^p \widetilde{\alpha}_j Y_{p+1-j}\Big|^2 \;&=\; E\Big|\big(B^0 \;-\; \rho B\big)\xi(B) Y_{p+1}\Big|^2 \\ &=\; E\Big|\big(B^0 \;-\; \rho B\big) Z_{p+1}\Big|^2 \\ &=\; E\Big|Z_{p+1} \;-\; \rho\, Z_p\Big|^2 \end{aligned} \tag{1.5.22}$$

and

$$\begin{aligned} E\Big|Y_{p+1} \;-\; \sum_{j=1}^p \alpha_j Y_{p+1-j}\Big|^2 \;&=\; E\Big|\big(B^0 \;-\; (1/c) B\big)\xi(B) Y_{p+1}\Big|^2 \\ &=\; E\Big|Z_{p+1} \;-\; (1/c)\, Z_p\Big|^2. \end{aligned} \tag{1.5.23}$$

It follows from the projection theorem (Theorem 1.3.6) that $\sum_{j=1}^p \alpha_j Y_{p+1-j}$ is the projection of $Y_{p+1}$ onto the subspace $\mathcal{M} = \{\sum_{i=1}^p c_i Y_{p+1-i} \colon\; c_1,\ldots,c_p \in \mathbb{C}\}$. Therefore the left-hand side of (1.5.23) is smaller than or equal to the left-hand side of (1.5.22). On the other hand, it follows again from the projection theorem that $\rho Z_p$ is the projection of $Z_{p+1}$ onto $\{c Z_p \colon\; c \in \mathbb{C}\}$. Therefore, the right-hand side of (1.5.22) is smaller than or equal to the right-hand side of (1.5.23). Hence we conclude that

$$E\Big|Z_{p+1} \;-\; \rho Z_p\Big|^2 \;=\; E\Big|Z_{p+1} \;-\; (1/c) Z_p\Big|^2,$$

which means that $(1/c)Z_p$ is also a projection of $Z_{p+1}$ onto $\{cZ_p\colon\ c \in \mathbb{C}\}$. Since the projection is unique we conclude that

$$0 = E\big|\rho Z_p - (1/c)Z_p\big|^2 = \big(\rho - (1/c)\big)^2 \underbrace{\|Z_p\|^2}_{\neq 0},$$

i.e. $\rho = 1/c$. It follows from the Cauchy-Schwarz inequality that $|\rho| \le 1$. However, if $|\rho| = 1$ then a simple computation reveals that

$$Z_{p+1} = \rho Z_p \qquad P - a.s.$$

Since $\xi(B) = B^0 - \sum_{i=1}^{p-1} b_j B^j$ we conclude that

$$0 = Z_{p+1} - \rho Z_p = Y_{p+1} - \sum_{j=1}^{p-1} b_j Y_{p+1-j} - \rho \xi(B) Y_p \qquad P - a.s.$$

This, however, contradicts the regularity of $\Gamma_{p+1}$, which in turn follows from Lemma 1.5.9. Therefore, our assumption that $\alpha$ has a zero inside the unit circle was wrong and the proof is complete. □

## Exercise

**Ex. 1.5.3** Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a causal stationary AR(2) process obeying

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, $\sigma^2 > 0$, and $\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2 \neq 0\ \forall z \in \mathbb{C}$, $|z| \le 1$. Compute the autocorrelations $\rho(1)$ and $\rho(2)$, where $\rho(k) = \gamma(k)/\gamma(0)$ and $\gamma(k) = \operatorname{cov}(X_{t+k}, X_t)$.

## A central limit theorem for sums of martingale differences

In the following we consider two popular methods of estimating the parameters of an autoregressive process, the least squares method and the method of moments, which is based on a sample version of the Yule-Walker equations. We will also investigate the asymptotic behavior of the least squares estimator when the sample size $n$ tends to infinity. It will be shown that this estimator, properly normalized, is asymptotically normally distributed. Such a result can be used to construct confidence sets for the unknown parameters, where a prescribed coverage probability is asymptotically guaranteed. As a prerequisite to such a result, we derive a suitable central limit theorem. It will turn out that we are faced with sums of dependent random variables which have the particular structure of martingale differences. Next we state and prove an appropriate extension of the Lindeberg-Feller central limit theorem to martingales.

Recall that the characteristic function $\varphi$ of a standard normal distribution with zero mean and variance $\sigma^2$ is given by $\varphi(t) = e^{-t^2\sigma^2/2}$ $\forall t \in \mathbb{R}$. In our proof of the central limit theorem, we make use of the following lemma which shows that the characteristic function of an **arbitrary** random variable $X$ with $EX = 0$ and $\mathrm{var}(X) = \sigma^2$ can be approximated, for small values of $\sigma^2$, by the characteristic function of a normal distribution with the same first two moments. Recall that the characteristic function $\varphi$ of a standard normal distribution with zero mean and variances $\sigma^2$ is given by $\varphi(t) = e^{-t^2\sigma^2/2}$ $\forall t \in \mathbb{R}$.

**Lemma 1.5.9.** *Let $X$ be a real-valued random variable such that $EX = 0$ and $E\big[X^2\big] = \sigma^2 < \infty$. Then*

$$\varphi_X(t) := Ee^{itX} = e^{-t^2\sigma^2/2} + r(t),$$

*where, for all $\epsilon > 0$,*

$$\big|r(t)\big| \leq \epsilon\frac{|t|^3\,\sigma^2}{2} + t^2 E\big[X^2\,\mathbb{1}\big(|X| > \epsilon\big)\big] + \frac{t^4\,\sigma^4}{8} \qquad \forall t \in \mathbb{R}.$$

*Proof.* Since $E\big[X^2\big] < \infty$, the characteristic function $\varphi_X$ is two times continuously differentiable and

$$\varphi_X'(t) = E\big[iX\,e^{itX}\big], \qquad \varphi_X''(t) = E\big[(iX)^2\,e^{itX}\big].$$

Therefore, we obtain by a Taylor series expansion

$$\begin{aligned}\big|\varphi_X(t) - e^{-t^2\sigma^2/2}\big| &= \Big|\underbrace{\varphi_X(0)}_{=1} + t\,\underbrace{\varphi_X'(0)}_{=0} + \frac{t^2}{2}\underbrace{\varphi_X''(0)}_{=-\sigma^2} + \frac{t^2}{2}\left[\varphi_X''(\xi) - \varphi_X''(0)\right] - e^{-t^2\sigma^2/2}\Big| \\ &\leq \left|\left(1 - \frac{t^2\sigma^2}{2}\right) - e^{-t^2\sigma^2/2}\right| + \frac{t^2}{2}\Big|\varphi_X''(\xi) - \varphi_X''(0)\Big|,\end{aligned}$$

for some $\xi$ between 0 and $t$. We have, again by a Taylor series expansion, that $e^{-u} = 1 - u + \frac{u^2}{2}e^{-\eta}$, for all $u \geq 0$ and suitable $\eta \in (0, u)$. This implies that

$$\left|\left(1 - \frac{t^2\sigma^2}{2}\right) - e^{-t^2\sigma^2/2}\right| \leq \frac{t^4\sigma^4}{8}.$$

Furthermore,

$$\begin{aligned}\left|\varphi_X''(\xi) - \varphi_X''(0)\right| &= \left|E\big[(iX)^2\big(e^{i\xi X} - e^{i0X}\big)\big]\right| \\ &\le E\Big[X^2 \underbrace{\left|e^{i\xi X} - e^{i0X}\right|}_{\le \epsilon|\xi| \le \epsilon|t|} \mathbb{1}\big(|X| \le \epsilon\big)\Big] + E\Big[X^2 \underbrace{\left|e^{i\xi X} - e^{i0X}\right|}_{\le 2} \mathbb{1}\big(|X| > \epsilon\big)\Big] \\ &\le \epsilon\,|t|\,E\big[X^2\big] + 2\,E\big[X^2\,\mathbb{1}\big(|X| > \epsilon\big)\big],\end{aligned}$$

which completes the proof. □

It will turn out that our proof of the central limit theorem for sums of martingale differences is rather complex. Therefore, we first take a look at the simpler proof of the classical central limit theorem for independent, but not necessarily identically distributed random variables, which is named after the Finnish mathematician Jarl Waldemar Lindeberg and the Croatian-American mathematician William Feller.

**Theorem 1.5.10.** *For $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent random variables on respective probability spaces $\big(\Omega_n, \mathcal{F}_n, P_n\big)$. Suppose that*

*(i)* $EX_{n,k} = 0 \qquad \forall k = 1, \ldots, k_n$,

*(ii) for* $\sigma_{n,k}^2 = E\big[X_{n,k}^2\big]$,

$$\sigma_n^2 := \sum_{k=1}^{k_n} \sigma_{n,k}^2 \underset{n\to\infty}{\longrightarrow} \sigma^2 \in [0,\infty),$$

*(iii) for all* $\epsilon > 0$,

$$L_n(\epsilon) := \sum_{k=1}^{k_n} E\Big[X_{n,k}^2 \mathbb{1}\big(|X_{n,k}| > \epsilon\big)\Big] \underset{n\to\infty}{\longrightarrow} 0.$$

*Then*

$$S_n := X_{n,1} + \cdots + X_{n,k_n} \overset{d}{\longrightarrow} Y \sim N(0,\sigma^2).$$

**Remark 1.5.11.** *Condition (iii) is the so-called Lindeberg condition. Since, for arbitrarily small $\epsilon > 0$,*

$$\sigma_{n,k}^2 = E\big[X_{n,k}^2\big] \le \epsilon^2 + E\Big[X_{n,k}^2 \mathbb{1}\big(|X_{n,k}| > \epsilon\big)\Big] \le \epsilon^2 + L_n(\epsilon)$$

*we conclude that*

$$\max_{1\le k\le k_n} \big\{\sigma_{n,k}^2\big\} \underset{n\to\infty}{\longrightarrow} 0. \tag{1.5.24}$$

*In other words, the Lindeberg condition guarantees that the contribution of any individual random variable $X_{n,k}$ ($1 \le k \le k_n$) to the variance $\sigma_n^2$ of the sum is arbitrarily small, for sufficiently large values of $n$.*

*Proof of Theorem 1.5.10.* It follows from **Lévy's continuity theorem**, named after the French mathematician Paul Lévy, that convergence in distribution of a sequence of random variables is equivalent to pointwise convergence of the corresponding characteristic functions. Let $\varphi_S$ denote the characteristic function of a generic random variable $S$. We have to show that

$$\varphi_{S_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_Y(t) \;=\; e^{-t^2\sigma^2/2} \qquad \forall t \in \mathbb{R}. \tag{1.5.25}$$

Since $\sigma_n^2 \underset{n\to\infty}{\longrightarrow} \sigma^2$ it suffices to show that

$$\left|\varphi_{S_n}(t) \;-\; e^{-t^2\sigma_n^2/2}\right| \underset{n\to\infty}{\longrightarrow} 0.$$

We know from Lemma 1.5.9 that the characteristic function of a zero mean random variable $X$ can be well approximated by the characteristic function of a normal distribution, provided the variance of $X$ is small. According to (1.5.24), the Lindeberg condition ensures that the variances $\sigma_{n,k}^2$ of the individual random variables $X_{n,k}$ are arbitrarily small, for large values of $n$. To bring the smallness of $\varphi_{X_{n,k}}(t) - e^{-t^2\sigma_{n,k}^2/2}$ into play, we split up

$$\begin{aligned}
&\varphi_{S_n}(t) \;-\; e^{-t^2\sigma_n^2/2} \\
&\quad= E\Big[\prod_{k=1}^{k_n}\varphi_{X_{n,k}}(t) \;-\; \prod_{k=1}^{k_n} e^{-t^2\sigma_{n,k}^2/2}\Big] \\
&\quad= \sum_{k=1}^{k_n}\Big[\prod_{j=1}^{k-1}\varphi_{X_{n,j}}(t)\,\Big(\varphi_{X_{n,k}}(t) \;-\; e^{-t^2\sigma_{n,k}^2/2}\Big)\prod_{j=k+1}^{k_n} e^{-t^2\sigma_{n,j}^2/2}\Big]
\end{aligned} \tag{1.5.26}$$

Since $\left|\prod_{j=1}^{k-1}\varphi_{X_{n,j}}(t)\right| \le 1$ and $\left|\prod_{j=k+1}^{k_n} e^{-t^2\sigma_{n,j}^2/2}\right| \le 1$ we obtain by Lemma 1.5.9 that

$$\begin{aligned}
\left|\varphi_{S_n}(t) \;-\; e^{-t^2\sigma_n^2/2}\right| \;&\le\; \sum_{k=1}^{k_n}\left|\varphi_{X_{n,k}}(t) \;-\; e^{-t^2\sigma_{n,k}^2/2}\right| \\
&\le\; \sum_{k=1}^{k_n}\Big\{\epsilon\frac{|t|^3\sigma_{n,k}^2}{2} \;+\; t^2\,E\big[X_{n,k}^2\mathbb{1}\big(|X_{n,k}|>\epsilon\big)\big] \;+\; \frac{t^4\sigma_{n,k}^4}{8}\Big\} \\
&\le\; \epsilon\,\frac{|t|^3\sigma_n^2}{2} \;+\; t^2\,L_n(\epsilon) \;+\; \max_{1\le j\le k_n}\big\{\sigma_{n,j}^2\big\}\,\frac{t^4\sigma_n^2}{8} \\
&=:\; R_{n,1}(\epsilon) \;+\; R_{n,2}(\epsilon) \;+\; R_{n,3},
\end{aligned} \tag{1.5.27}$$

say. Let $\delta > 0$ be arbitrary. Since $\sigma_n^2 \underset{n\to\infty}{\longrightarrow} \sigma^2 < \infty$ we have that $\sup_{n\in\mathbb{N}}\{\sigma_n^2\} < \infty$ and therefore

$$\left|R_{n,1}(\epsilon)\right| \;\le\; \delta/3 \qquad \forall n \in \mathbb{N} \tag{1.5.28a}$$

if $\epsilon = \epsilon(\delta)$ is sufficiently small. It follows from the Lindeberg condition (iii) that

$$\left|R_{n,2}(\epsilon)\right| \;\le\; \delta/3 \qquad \forall n \ge N_1, \tag{1.5.28b}$$

for sufficiently large $N_1$. We obtain from (1.5.24) that

$$\left|R_{n,3}\right| \;\le\; \delta/3 \qquad \forall n \ge N_2, \tag{1.5.28c}$$

also for sufficiently large $N_2$. Finally, it follows from (1.5.27) and (1.5.28a) to (1.5.28c) that

$$\left|\varphi_{S_n}(t) \;-\; e^{-t^2\sigma_n^2/2}\right| \;\le\; \delta \qquad \forall n \ge \max\{N_1, N_2\},$$

which completes the proof. □

Now we generalize the classical Lindeberg-Feller central limit theorem to the case of sums of martingale differences.

**Theorem 1.5.12.** *For $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be random variables on respective probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$. Let $\mathcal{F}_k^{(n)}$ be $\sigma$-algebras ($k = 0, 1, \ldots, k_n$) such that*

$$\sigma\big(X_{n,1}, \ldots, X_{n,k}\big) \subseteq \mathcal{F}_k^{(n)} \subseteq \mathcal{F}_{k+1}^{(n)} \subseteq \mathcal{F}_n.$$

*Moreover, suppose that*

*(i)* $E\big(X_{n,k} \big| \mathcal{F}_{k-1}^{(n)}\big) = 0 \quad P_n - a.s. \qquad \forall k = 1, \ldots, k_n,$

*(ii)* *for* $\sigma_{n,k}^2 := E\big(X_{n,k}^2 \big| \mathcal{F}_{k-1}^{(n)}\big)$,

$$\sigma_n^2 := \sum_{k=1}^{k_n} \sigma_{n,k}^2 \xrightarrow{P} \sigma^2 < \infty \qquad (\text{as } n \to \infty),$$

*(iii)* *for all* $\epsilon > 0$,

$$L_n(\epsilon) := \sum_{k=1}^{k_n} E\Big(X_{n,k}^2 \, \mathbb{1}\big(|X_{n,k}| > \epsilon\big) \Big| \mathcal{F}_{k-1}^{(n)}\Big) \xrightarrow{P} 0.$$

*Then*

$$S_n := X_{n,1} + \cdots + X_{n,k_n} \xrightarrow{d} Z \sim N(0, \sigma^2).$$

*Proof.* We start out with some preparatory considerations. It follows from the **conditional Lindeberg condition** (iii) that

$$\begin{aligned} \sigma_{n,k}^2 &\leq \epsilon^2 + E\big(X_{n,k}^2 \, \mathbb{1}\big(|X_{n,k}| > \epsilon\big) \big| \mathcal{F}_{k-1}^{(n)}\big) \\ &\leq \epsilon^2 + L_n(\epsilon), \end{aligned}$$

which implies that

$$\max_{1 \leq k \leq k_n} \big\{\sigma_{n,k}^2\big\} \xrightarrow{P} 0. \tag{1.5.29}$$

In order to make the transition from the not necessarily normally distributed random variables $X_{n,1}, \ldots, X_{n,k_n}$ to the Gaussian case we apply Lemma 1.5.9 to the **conditional distributions** of the $X_{n,k}$ and obtain the following estimate:

$$\begin{aligned} &\sum_{k=1}^{k_n} \Big| E\Big(e^{itX_{n,k}} - e^{-t^2 \sigma_{n,k}^2/2} \Big| \mathcal{F}_{k-1}^{(n)}\Big) \Big| \\ &\leq \sum_{k=1}^{k_n} \Big\{ \epsilon \frac{|t|^3 \, \sigma_{n,k}^2}{2} + t^2 \, E\Big(X_{n,k}^2 \, \mathbb{1}\big(|X_{n,k}| > \epsilon\big) \Big| \mathcal{F}_{k-1}^{(n)}\Big) + \frac{t^4}{8} \max_{1 \leq j \leq k_n} \big\{\sigma_{n,j}^2\big\} \, \sigma_{n,k}^2 \Big\}, \end{aligned}$$

where $\epsilon > 0$ is arbitrary. Using this we obtain from the conditional Lindeberg condition (iii) and (1.5.29) that

$$\sum_{k=1}^{k_n} \Big| E\Big(e^{itX_{n,k}} - e^{-t^2 \sigma_{n,k}^2/2} \Big| \mathcal{F}_{k-1}^{(n)}\Big) \Big| \xrightarrow{P} 0.$$

Therefore, there exists a null sequence $(\epsilon_n)_{n\in\mathbb{N}}$ such that

$$P\left(\sum_{k=1}^{k_n}\left|E\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right| > \epsilon_n\right) \le \epsilon_n. \tag{1.5.30}$$

Now we begin with the main part of the proof. We show that the characteristic function $\varphi_{S_n}$ of $S_n$ converges to that of a normal distribution with zero mean and variance $\sigma^2$. Since the behavior of $S_n$ is closely connected with the sum of the conditional variances we split up

$$\left|\varphi_{S_n}(t) - e^{-t^2\sigma^2/2}\right| \le \left|Ee^{itS_n} - Ee^{-t^2\sigma_n^2/2}\right| + \left|E\left[e^{-t^2\sigma_n^2/2} - e^{-t^2\sigma^2/2}\right]\right|. \tag{1.5.31}$$

Since $\sigma_n^2 \xrightarrow{P} \sigma^2$ as $n \to \infty$ we obtain that

$$\left|E\left[e^{-t^2\sigma_n^2/2} - e^{-t^2\sigma^2/2}\right]\right| \underset{n\to\infty}{\longrightarrow} 0. \tag{1.5.32}$$

To estimate the first term on the right-hand side of (1.5.31), it is tempting to adapt an approach often used for proving a central limit theorem for sums of independent random variables. Here is a natural attempt:

$$\begin{aligned}
&E\left[e^{itS_n} - e^{-t^2\sigma_n^2/2}\right] \\
&\quad = E\left[\prod_{k=1}^{k_n} e^{itX_{n,k}} - \prod_{k=1}^{k_n} e^{-t^2\sigma_{n,k}^2/2}\right] \\
&\quad = \sum_{k=1}^{k_n} E\left[e^{it(X_{n,1}+\cdots+X_{n,k-1})}\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\right)e^{-t^2(\sigma_{n,k+1}^2+\cdots+\sigma_{n,k_n}^2)/2}\right] \\
&\quad = \sum_{k=1}^{k_n} E\left[E\left(e^{it(X_{n,1}+\cdots+X_{n,k-1})}\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\right)e^{-t^2(\sigma_{n,k+1}^2+\cdots+\sigma_{n,k_n}^2)/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right] \\
&\quad = \sum_{k=1}^{k_n} E\left[e^{it(X_{n,1}+\cdots+X_{n,k-1})}\, E\left(\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\right)e^{-t^2(\sigma_{n,k+1}^2+\cdots+\sigma_{n,k_n}^2)/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right].
\end{aligned}$$

Now it seems that we have achieved what we want: The term $e^{it(X_{n,1}+\cdots+X_{n,k-1})}$ is bounded in absolute value by 1. The term $e^{-t^2(\sigma_{n,k+1}^2+\cdots+\sigma_{n,k_n}^2)/2}$ is also bounded by 1, and (1.5.30) provides a useful estimate for the sum of the remaining terms, $\sum_{k=1}^{k_n}\left|E\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right|$. Nevertheless, we are at a dead end here since the term $e^{-t^2(\sigma_{n,k+1}^2+\cdots+\sigma_{n,k_n}^2)/2}$ cannot be taken out of the conditional expectation. To get out of this deadlock we could multiply $e^{itS_n} - e^{-t^2\sigma_n^2/2}$ by $e^{t^2\sigma_n^2/2}$, which leads to

$$\begin{aligned}
&E\left[\left(e^{itS_n} - e^{-t^2\sigma_n^2/2}\right)e^{t^2\sigma_n^2/2}\right] \\
&\quad = \ldots = \sum_{k=1}^{k_n} E\left[e^{it(X_{n,1}+\cdots+X_{n,k-1})}\, E\left(\left(e^{itX_{n,k}} - e^{-t^2\sigma_{n,k}^2/2}\right)e^{t^2(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2)/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right].
\end{aligned}$$

Now it follows from $\sigma_{n,j}^2 = E\left(X_{n,j}^2\middle|\mathcal{F}_{j-1}^{(n)}\right)$ that $e^{t^2(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2)/2}$ is $\mathcal{F}_{k-1}^{(n)}$-measurable. Therefore, the term $e^{t^2(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2)/2}$ can be taken out of the conditional expectation and we can hope to make progress. There is, however, one more obstacle: Although we have $e^{t^2(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2)/2} \leq e^{t^2\sigma_n^2/2} \xrightarrow{P} e^{t^2\sigma^2/2}$, we cannot find an upper bound for the **expectation** of $e^{t^2(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2)/2}$. Therefore, we use a typical truncation argument and define

$$\widetilde{X}_{n,k} := \begin{cases} X_{n,k} & \text{if } \sigma_{n,1}^2+\cdots+\sigma_{n,k}^2 \leq 2\sigma^2 \text{ and } \sum_{j=1}^k \left|E\left(e^{itX_{n,j}} - e^{-t^2\sigma_{n,j}^2/2}\middle|\mathcal{F}_{j-1}^{(n)}\right)\right| \leq \epsilon_n, \\ 0 & \text{otherwise} \end{cases}$$

and, accordingly,

$$\begin{aligned} \widetilde{\sigma}_{n,k}^2 &:= E\big(\widetilde{X}_{n,k}^2\big|\mathcal{F}_{k-1}^{(n)}\big) \\ &= E\Big(X_{n,k}^2\, \mathbb{1}\Big(\sigma_{n,1}^2+\cdots+\sigma_{n,k}^2 \leq 2\sigma^2 \quad \text{and} \quad \sum_{j=1}^k \left|E\left(e^{itX_{n,j}} - e^{-t^2\sigma_{n,j}^2/2}\middle|\mathcal{F}_{j-1}^{(n)}\right)\right| \leq \epsilon_n\Big)\Big|\mathcal{F}_{k-1}^{(n)}\Big) \\ &= \mathbb{1}(\ldots)\, E\big(X_{n,k}^2\big|\mathcal{F}_{k-1}^{(n)}\big) \\ &= \begin{cases} \sigma_{n,k}^2 & \text{if } \sigma_{n,1}^2+\cdots+\sigma_{n,k}^2 \leq 2\sigma^2 \quad \text{and} \quad \sum_{j=1}^k \left|E\left(e^{itX_{n,j}} - e^{-t^2\sigma_{n,j}^2/2}\middle|\mathcal{F}_{j-1}^{(n)}\right)\right| \leq \epsilon_n, \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note that we still have

$$E\big(\widetilde{X}_{n,k}\big|\mathcal{F}_{k-1}^{(n)}\big) \;=\; E\big(X_{n,k}\mathbb{1}(\ldots)\big|\mathcal{F}_{k-1}^{(n)}\big) \;=\; \mathbb{1}(\ldots)E\big(X_{n,k}\big|\mathcal{F}_{k-1}^{(n)}\big) \;=\; 0 \quad P-a.s.$$

It follows from the above definitions that

$$\widetilde{\sigma}_n^2 \;:=\; \widetilde{\sigma}_{n,1}^2 \,+\, \cdots \,+\, \widetilde{\sigma}_{n,k_n}^2 \;\leq\; 2\,\sigma^2 \tag{1.5.33}$$

and

$$\sum_{k=1}^{k_n} \left|E\left(e^{it\widetilde{X}_{n,k}} - e^{-t^2\widetilde{\sigma}_{n,k}^2/2}\middle|\mathcal{F}_{k-1}^{(n)}\right)\right| \;\leq\; \epsilon_n. \tag{1.5.34}$$

We obtain from (1.5.30) and $P\big(\sigma_n^2 > 2\sigma^2\big) \underset{n\to\infty}{\longrightarrow} 0$ that

$$P\big(X_{n,k} \neq \widetilde{X}_{n,k} \quad \text{for at least one } k \leq k_n\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{1.5.35}$$

We define

$$\widetilde{S}_n \;:=\; \widetilde{X}_{n,1} \,+\, \cdots \,+\, \widetilde{X}_{n,k_n}, \qquad \widetilde{\sigma}_n^2 \;:=\; \widetilde{\sigma}_{n,1}^2 \,+\, \cdots \,+\, \widetilde{\sigma}_{n,k_n}^2.$$

Now we are prepared to derive the missing upper estimate for the first term on the right-hand side of (1.5.31). We have

$$\begin{aligned} &\left|Ee^{itS_n} \,-\, Ee^{-t^2\sigma_n^2/2}\right| \\ &\leq\; \left|Ee^{itS_n} \,-\, Ee^{it\widetilde{S}_n}\right| \,+\, \left|Ee^{it\widetilde{S}_n} \,-\, Ee^{-t^2\widetilde{\sigma}_n^2/2}\right| \,+\, \left|Ee^{-t^2\widetilde{\sigma}_n^2/2} \,-\, Ee^{-t^2\sigma_n^2/2}\right| \\ &=:\; T_{n,1} \,+\, T_{n,2} \,+\, T_{n,3}, \end{aligned} \tag{1.5.36}$$

say. It follows immediately from (1.5.35) that

$$T_{n,1} \;\leq\; 2\,P\big(S_n \neq \widetilde{S}_n\big) \underset{n\to\infty}{\longrightarrow} 0 \tag{1.5.37}$$

as well as

$$T_{n,3} \;\leq\; P\big(\sigma_n^2 \neq \widetilde{\sigma}_n^2\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{1.5.38}$$

It remains to derive an upper estimate for $T_{n,2}$. Using that

$$\begin{aligned} e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2} &= \prod_{k=1}^{k_n} e^{it\widetilde{X}_{n,k}} - \prod_{k=1}^{k_n} e^{-t^2\widetilde{\sigma}_{n,k}^2/2} \\ &= \sum_{k=1}^{k_n} e^{it(\widetilde{X}_{n,1}+\cdots+\widetilde{X}_{n,k-1})}\Big(e^{it\widetilde{X}_{n,k}} - e^{-t^2\widetilde{\sigma}_{n,k}^2/2}\Big)e^{-t^2(\widetilde{\sigma}_{n,k+1}^2+\cdots+\widetilde{\sigma}_{n,k_n}^2)/2} \end{aligned}$$

we have

$$\begin{aligned} &\Big|E\Big[\Big(e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2}\Big)e^{t^2\widetilde{\sigma}_n^2/2}\Big]\Big| \\ &\le \Big|\sum_{k=1}^{k_n} E\Big[E\Big(\underbrace{e^{it(\widetilde{X}_{n,1}+\cdots+\widetilde{X}_{n,k-1})}}_{|\ldots|\le 1}\Big(e^{it\widetilde{X}_{n,k}} - e^{-t^2\widetilde{\sigma}_{n,k}^2/2}\Big)\underbrace{e^{t^2(\widetilde{\sigma}_{n,1}^2+\cdots+\widetilde{\sigma}_{n,k}^2)/2}}_{\le e^{t^2\sigma^2} \text{ by (1.5.33)}}\Big|\mathcal{F}_{k-1}^{(n)}\Big)\Big]\Big| \\ &\le e^{t^2\sigma^2}\, E\Big[\underbrace{\sum_{k=1}^{n}\Big|E\Big(e^{it\widetilde{X}_{n,k}} - e^{-t^2\widetilde{\sigma}_{n,k}^2/2}\Big|\mathcal{F}_{k-1}^{(n)}\Big)\Big|}_{\le \epsilon_n \text{ by (1.5.34)}}\Big] \\ &\le e^{t^2\sigma^2}\,\epsilon_n. \end{aligned}$$

Therefore, we obtain that

$$\begin{aligned} T_{n,2} &= \Big|E\Big[e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2}\Big]\Big| \\ &= \Big|E\Big[e^{-t^2\widetilde{\sigma}_n^2/2}\Big(e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2}\Big)e^{t^2\widetilde{\sigma}_n^2/2}\Big]\Big| \\ &\le \Big|e^{-t^2\sigma^2/2}\underbrace{E\Big[\Big(e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2}\Big)e^{t^2\widetilde{\sigma}_n^2/2}\Big]}_{\underset{n\to\infty}{\longrightarrow} 0}\Big| \\ &\quad + \Big|E\Big[\underbrace{\Big(e^{-t^2\widetilde{\sigma}_n^2/2} - e^{-t^2\sigma^2/2}\Big)}_{\overset{P}{\longrightarrow} 0}\underbrace{\Big(e^{it\widetilde{S}_n} - e^{-t^2\widetilde{\sigma}_n^2/2}\Big)}_{|\ldots|\le 2}\underbrace{e^{t^2\widetilde{\sigma}_n^2/2}}_{\le e^{t^2\sigma^2}}\Big]\Big| \\ &\underset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{1.5.39}$$

(1.5.36) to (1.5.39) yield that the first term on the right-hand side of (1.5.31) tends to zero, which completes the proof. □

## Exercise

**Ex. 1.5.4** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0,\sigma^2)$, $\sigma^2 > 0$, and $X_t = \sum_{k=1}^{q} \beta_k \varepsilon_{t-k}$ $((\beta_1,\ldots,\beta_q)^T \neq 0_q)$. Consider the following linear regression model with **dependent** explanatory variables:

$$Y_t \,=\, \alpha\, X_t \,+\, \varepsilon_t, \qquad t = 1,\ldots,n$$

and let $\widetilde{\alpha}_n \in \arg\min_{\alpha\in\mathbb{R}} \sum_{t=1}^{n} \big(Y_t - \alpha X_t\big)^2$ be the least squares estimator of $\alpha$, based on $(X_1,Y_1),\ldots,(X_n,Y_n)$.

(i) Show that

$$\frac{1}{n}\sum_{t=1}^{n} \varepsilon_{t-k}\varepsilon_{t-l} \stackrel{a.s.}{\longrightarrow} E[\varepsilon_{-k}\varepsilon_{-l}] \quad \text{as } n\to\infty \quad (k,l = 1,\ldots,q)$$

and conclude that

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 \stackrel{P}{\longrightarrow} EX_0^2 > 0 \qquad \text{as } n\to\infty.$$

(ii) Show that $\widetilde{\alpha}_n = \big(\sum_{t=1}^{n} X_t^2\big)^{-1} \sum_{t=1}^{n} X_t Y_t$ if $\sum_{t=1}^{n} X_t^2 > 0$.

(iii) Show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t\varepsilon_t \stackrel{d}{\longrightarrow} Z_0 \sim N(0, v_0)$$

and determine $v_0$.
(*Hint: Choose* $\mathcal{F}_t^{(n)} = \sigma(\varepsilon_{1-q},\ldots,\varepsilon_t)$ *and use Theorem 1.5.12.*)

(iv) Show that

$$\sqrt{n}\big(\widetilde{\alpha}_n \,-\, \alpha\big) \stackrel{d}{\longrightarrow} Z \sim N(0, v)$$

and determine $v$.

## Parameter estimation for autoregressive processes

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary autoregressive process satisfying

$$X_t \;=\; \alpha_1 X_{t-1} \,+\, \cdots \,+\, \alpha_p X_{t-p} \,+\, \varepsilon_t \qquad \forall t\in\mathbb{Z},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$ and $\alpha(z) = 1-\alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0\ \forall z\in\mathbb{C},\ |z|\leq 1$. We assume that realizations $x_1,\ldots,x_n$ of the random variables $X_1,\ldots,X_n$ are observed. Our aim is to find estimators of the coefficient vector $\alpha=(\alpha_1,\ldots,\alpha_p)^T$ and the white noise variance $\sigma^2$. We will briefly consider two popular methods, the least squares method and the method of moments which is based on a sample version of the Yule-Walker equations.

1) **The least squares estimator**

Based on $X_1,\ldots,X_n$, the least squares estimator of the vector $\alpha$ is given by

$$\widetilde{\alpha}_n \;=\; \big(\widetilde{\alpha}_{n,1},\ldots,\widetilde{\alpha}_{n,p}\big)^T \;\in\; \arg\min_{\alpha\in\mathbb{R}^p} \sum_{t=p+1}^{n} \Big(X_t \,-\, \sum_{k=1}^{p} \alpha_k X_{t-k}\Big)^2.$$

To simplify our presentation we prefer to use matrix/vector notation, i.e.

$$\widetilde{\alpha}_n \;\in\; \arg\min_{\alpha} \big\|Y \,-\, X\alpha\big\|^2, \tag{1.5.40}$$

where

$$X \;=\; \begin{pmatrix} X_{p+1-1} & \ldots & X_{p+1-p} \\ \vdots & \ddots & \vdots \\ X_{n-1} & \ldots & X_{n-p} \end{pmatrix}, \qquad Y \;=\; \begin{pmatrix} X_{p+1} \\ \vdots \\ X_n \end{pmatrix},$$

and $\|\cdots\|$ denotes the Euclidean norm on $\mathbb{R}^{n-d}$. Recall that the projection theorem (Theorem 1.3.6) guarantees that a solution to the optimization problem (1.5.40) exists. Indeed, $\mathcal{M} := \{Xb\colon\ b\in\mathbb{R}^p\}$ is a closed subspace of $\mathcal{H}=\mathbb{R}^{n-p}$ and $X\widetilde{\alpha}_n$ is therefore the unique orthogonal projection of $Y$ onto $\mathcal{M}$. Part (ii) of Theorem 1.3.6 helps us to identify a solution. Any solution $\widetilde{\alpha}_n$ has to satisfy

$$\underbrace{\langle Y - X\widetilde{\alpha}_n,\, Xb\rangle}_{=b^T X^T (Y - X\widetilde{\alpha}_n)} \;=\; 0 \qquad \forall b\in\mathbb{R}^p,$$

which is is equivalent to

$$X^T Y \;=\; X^T X \widetilde{\alpha}_n.$$

Hence, $\widetilde{\alpha}_n$ is a solution to the so-called **normal equation**. If $X^T X$ is regular, then

$$\widetilde{\alpha}_n \;=\; \big(X^T X\big)^{-1} X^T Y$$

is the unique solution. Otherwise, there exist infinitely many solutions, although $X\widetilde{\alpha}_n$ is the same for all solutions $\widetilde{\alpha}_n$.

2) **The Yule-Walker estimator**

The Yule-Walker estimator is based on the relation between the parameters $\alpha_1, \dots, \alpha_p$ and $\sigma^2$, and the covariances $\gamma_0, \dots, \gamma_p$. This connection is expressed by the Yule-Walker equations (1.5.19) and (1.5.20). If we replace the autocovariances by their corresponding sample versions,

$$\widehat{\gamma}_n(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{t+|k|} - \bar{X}_n)(X_t - \bar{X}_n),$$

we obtain the equations

$$\widehat{\gamma}_n^{(p)} = \widehat{\Gamma}_{n,p} \, \widehat{\alpha}_n \tag{1.5.41}$$

and

$$\widehat{\gamma}_n(0) = \widehat{\gamma}_p^T \widehat{\alpha}_n + \widehat{\sigma}_n^2, \tag{1.5.42}$$

where $\widehat{\gamma}_n^{(p)} = \big(\widehat{\gamma}_n(1), \dots, \widehat{\gamma}_n(p)\big)^T$, $\widehat{\Gamma}_{n,p} = \big(\widehat{\gamma}_n(i-j)\big)_{i,j=1}^p$. $\widehat{\alpha}_n = \big(\widehat{\alpha}_{n,1}, \dots, \widehat{\alpha}_{n,p}\big)^T$ and $\widehat{\sigma}_n^2$ are the **Yule-Walker estimators** of $\alpha$ and $\sigma^2$, respectively. Recall that the choice of the factor of $1/n$ in the definition of $\widehat{\gamma}_n(k)$ ensures that $\widehat{\gamma}_n \colon \mathbb{Z} \to \mathbb{R}$ is an even and non-negative definite function; see Lemma 1.4.7. Theorem **??** yields then that $\widehat{\gamma}_n$ is the autocovariance function of an appropriate stationary process. If $\widehat{\gamma}_n(0) > 0$, it follows from Lemma 1.5.9 that the matrix $\widehat{\Gamma}_{n,p}$ is regular. Therefore, equation (1.5.41) has a unique solution. Furthermore, Theorem 1.5.8 shows that there exists a stationary AR($p$) process $(Y_t)_{t\in\mathbb{Z}}$ obeying

$$Y_t = \widehat{\alpha}_{n,1} Y_{t-1} + \cdots + \widehat{\alpha}_{n,p} Y_{t-p} + Z_t,$$

where $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \widehat{\sigma}_n^2)$ and $\mathrm{cov}(Y_{t+k}, Y_t) = \widehat{\gamma}_n(k)$, for $k = 0, \dots, p$. The case of $\widehat{\gamma}_n(0) = 0$ is of minor interest since we know that $\widehat{\gamma}_n(0)$ converges to $\gamma(0)$ as $n \to \infty$ and $\gamma(0)$ means that $X_t = 0$ with probability one. Nonetheless, we mention that $\widehat{\alpha}_n = 0_p^T$ and $\widehat{\sigma}_n^2 = 0$ is a trivial solution to (1.5.41) and (1.5.42) in the latter case.

It is easy to see that $\big(\widehat{\alpha}_n\big)_{n\in\mathbb{N}}$ and $\big(\widehat{\sigma}_n^2\big)_{n\in\mathbb{N}}$ are **consistent** sequences of estimators of $\alpha$ and $\sigma^2$, respectively. Indeed, under the conditions of Lemma 1.4.11 we have that

$$E\Big[\big(\widehat{\gamma}_n(k) - \gamma(k)\big)^2\Big] \underset{n\to\infty}{\longrightarrow} 0,$$

which implies

$$\widehat{\gamma}_n(k) \xrightarrow{P} \gamma(k) \qquad \forall k \in \mathbb{Z}.$$

Therefore, $\widehat{\gamma}_n^{(p)} \xrightarrow{P} \gamma_p$ and, if $\Gamma_p$ is regular, $\widehat{\Gamma}_{n,p}^{-1} \xrightarrow{P} \Gamma_p^{-1}$. This implies that

$$\widehat{\alpha}_n = \widehat{\Gamma}_{n,p}^{-1} \widehat{\gamma}_n^{(p)} \xrightarrow{P} \Gamma_p^{-1} \gamma_p = \alpha$$

and

$$\widehat{\sigma}_n^2 = \widehat{\gamma}_n(0) - \widehat{\alpha}_n^T \widehat{\gamma}_n^{(p)} \xrightarrow{P} \gamma(0) - \alpha^T \gamma_p = \sigma^2.$$

In the following, we investigate the asymptotic behavior of the least squares estimator $\widetilde{\alpha}_n$, as the sample size $n$ tends to infinity. We suppose that $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary AR($p$) process,

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \qquad \forall t \in \mathbb{Z},$$

where $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1$. In view of an application of the central limit theorem for sums of martingale differences (Theorem 1.5.12) we tighten our assumption on the sequence of innovations and assume that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 > 0$. We further assume that realizations $x_1, \ldots, x_n$ of the random variables $X_1, \ldots, X_n$ are observed and we denote the least squares estimator based on these random variables by $\widetilde{\alpha}_n$. Since we are interested in an asymptotic result, we use the index $n$ to indicate the corresponding sample size Recall that $\widetilde{\alpha}_n$ is given by

$$\widetilde{\alpha}_n \in \arg\min_{\alpha\in\mathbb{R}^p} \left\| Y - X\alpha \right\|^2$$

where

$$Y = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_n \end{pmatrix}, \qquad X = \begin{pmatrix} X_{p+1-1} & \ldots & X_{p+1-p} \\ \vdots & \ddots & \vdots \\ X_{n-1} & \ldots & X_{n-p} \end{pmatrix}.$$

The next theorem shows that $\widetilde{\alpha}_n$, properly normalized, is asymptotically normally distributed. Note that, if $X^TX$ is regular, then $\widetilde{\alpha}_n = (X^TX)^{-1}X^TY$ and

$$\sqrt{n}\big(\widetilde{\alpha}_n - \alpha\big) = \sqrt{n}\Big[\big(X^TX\big)^{-1}X^T\big(X\alpha + \varepsilon\big) - \alpha\Big] = \Big(\frac{1}{n}X^TX\Big)^{-1}\frac{1}{\sqrt{n}}X^T\varepsilon,$$

where $\varepsilon = (\varepsilon_{p+1}, \ldots, \varepsilon_n)^T$.

**Theorem 1.5.13.** *Suppose that the above conditions are fulfilled. Then,*

*(i)* *if $\gamma$ denotes the autocovarianec function of $(X_t)_{t\in\mathbb{Z}}$,*

$$\frac{1}{n}X^TX \xrightarrow{P} \Gamma_p = \begin{pmatrix} \gamma(1-1) & \ldots & \gamma(1-p) \\ \vdots & \ddots & \vdots \\ \gamma(p-1) & \ldots & \gamma(p-p) \end{pmatrix},$$

*(ii)* $\frac{1}{\sqrt{n}}X^T\varepsilon \xrightarrow{d} Z_0 \sim N\big(0_p, \sigma_\varepsilon^2\,\Gamma_p\big)$,

*(iii)* $\sqrt{n}\big(\widetilde{\alpha}_n - \alpha\big) \xrightarrow{d} Z \sim N\big(0_p, \sigma_\varepsilon^2\,\Gamma_p^{-1}\big)$.

*Proof.* (i) We show that

$$\Big(\frac{1}{n}X^TX\Big)_{i,j} = \frac{1}{n}\sum_{t=p+1}^{n} X_{t-i}X_{t-j} \xrightarrow{P} \big(\Gamma_p\big)_{i,j} = \gamma(i-j). \tag{1.5.43}$$

Recall that $(X_t)_{t\in\mathbb{Z}}$ has a representation as a linear process,

$$X_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k},$$

where $\sum_{k=0}^{\infty} |\beta_k| < \infty$. Define

$$X_{t,m} := \sum_{k=0}^{m} \beta_k \varepsilon_{t-k}.$$

It follows from the strong law of large numbers that

$$\frac{1}{n} \sum_{t=p+1}^{n} X_{t-i,m} X_{t-j,m} \xrightarrow{a.s.} \gamma^{(m)}(i-j) := E\big[X_{t-i,m} X_{t-j,m}\big]. \tag{1.5.44}$$

Indeed, we have that

$$\begin{aligned}
&\frac{1}{n} \sum_{t=p+1}^{n} X_{t-i,m} X_{t-j,m} \\
&\quad = \frac{1}{n} \sum_{t=p+1}^{n} \sum_{k_1,k_2=0}^{m} \beta_{k_1} \varepsilon_{t-i-k_1} \, \beta_{k_2} \varepsilon_{t-j-k_2} \\
&\quad = \sum_{k_1,k_2=0}^{m} \beta_{k_1} \beta_{k_2} \frac{1}{n} \sum_{t=p+1}^{n} \varepsilon_{t-i-k_1} \varepsilon_{t-j-k_2} \\
&\quad = \sum_{k_1,k_2=0}^{m} \beta_{k_1} \beta_{k_2} \sum_{l=0}^{\Delta(k_1,k_2)} \underbrace{\frac{1}{n} \sum_{s:\, p+1 \le s(\Delta(k_1,k_2)+1)+l \le n} \varepsilon_{s(\Delta(k_1,k_2)+1)+l-i-k_1} \varepsilon_{s(\Delta(k_1,k_2)+1)+l-j-k_2}}_{\xrightarrow{a.s} E[\varepsilon_{-i-k_1} \varepsilon_{-j-k_2}]/(\Delta(k_1,k_2)+1)} \\
&\quad \xrightarrow{a.s.} \sum_{k_1,k_2=0}^{m} \beta_{k_1} \beta_{k_2} E[\varepsilon_{-i-k_1} \varepsilon_{-j-k_2}] = E\big[X_{t-i,m} X_{t-j,m}\big],
\end{aligned}$$

where $\Delta(k_1, k_2) = |i + k_1 - j - k_2|$ is chosen such that the summands in the inner sum on the fourth line of this display are independent. This allows us to apply the strong law of large numbers which yields almost sure convergence. Furthermore, since

$$\begin{aligned}
&X_{t-i,m} X_{t-j,m} - X_{t-i} X_{t-j} \\
&\quad = \big(X_{t-i,m} - X_{t-i}\big)\big(X_{t-j,m} - X_{t-j}\big) + X_{t-i}\big(X_{t-j,m} - X_{t-j}\big) + \big(X_{t-i,m} - X_{t-i}\big) X_{t-j}
\end{aligned}$$

we obtain by $\|X_{t,m} - X_t\| \underset{m\to\infty}{\longrightarrow} 0$ that

$$\begin{aligned}
E\big|X_{t-i,m} X_{t-j,m} - X_{t-i} X_{t-j}\big|
&\le \sqrt{E\big(X_{t-i,m} - X_{t-i}\big)^2} \sqrt{E\big(X_{t-j,m} - X_{t-j}\big)^2} \\
&\quad + \sqrt{E X_{t-i}^2} \sqrt{E\big(X_{t-j,m} - X_{t-j}\big)^2} \\
&\quad + \sqrt{E\big(X_{t-i,m} - X_{t-i}\big)^2} \sqrt{E X_{t-j}^2} \\
&\underset{m\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Therefore,

$$\sup_n \left\{ E\left| \frac{1}{n} \sum_{t=p+1}^{n} X_{t-i,m} X_{t-j,m} - \frac{1}{n} \sum_{t=p+1}^{n} X_{t-i} X_{t-j} \right| \right\} \underset{m\to\infty}{\longrightarrow} 0,$$

which implies by Markov's inequality that, for all $n \geq p+1$ and arbitrary $\epsilon > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{t=p+1}^{n} X_{t-i,m}X_{t-j,m} - \frac{1}{n}\sum_{t=p+1}^{n} X_{t-i}X_{t-j}\Big| \geq \epsilon\Big) \leq \epsilon, \tag{1.5.45}$$

if $m$ is sufficiently large. Finally, again by $\|X_{t,m} - X_t\| \underset{m\to\infty}{\longrightarrow} 0$, it follows from continuity of the inner product that

$$\gamma^{(m)}(i-j) \underset{m\to\infty}{\longrightarrow} \gamma(i-j). \tag{1.5.46}$$

From (1.5.44) to (1.5.46) we obtain (1.5.43).

(ii) Let $Z_n = \frac{1}{\sqrt{n}}X^T\varepsilon$. By the Cramer-Wold device, the relation $Z_n \stackrel{d}{\longrightarrow} Z_0$ is equivalent to

$$c^T Z_n \stackrel{d}{\longrightarrow} c^T Z_0 \sim N\big(0, \sigma_\varepsilon^2\, c^T\Gamma_p c\big) \qquad \forall c \in \mathbb{R}^p. \tag{1.5.47}$$

Let $c \in \mathbb{R}^p$ be arbitrary. We have that

$$c^T Z_n = \sum_{t=p+1}^{n} \underbrace{\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{p} c_i X_{t-i}\Big)\varepsilon_t}_{=:\, Z_{n,t}}.$$

We will show that the triangular array of random variables $Z_{n,t}$ ($t = p+1, \ldots, n$, $n \geq p+1$) satisfies the conditions of Theorem 1.5.12. Let $\mathcal{F}_t^{(n)} := \sigma\big(\varepsilon_t, \varepsilon_{t-1}, \ldots\big)$. Since $X_t = \sum_{k=0}^\infty \beta_k\varepsilon_{t-k}$ we have that

$$\sigma\big(X_1, \ldots, X_{t-1}\big) \subseteq \mathcal{F}_{t-1}^{(n)}.$$

This implies that

$$\begin{aligned} E\big(Z_{n,t}\big|\mathcal{F}_{t-1}^{(n)}\big) &= E\Big(\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{p} c_i X_{t-i}\Big)\varepsilon_t\Big|\mathcal{F}_{t-1}^{(n)}\Big) \\ &= \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{p} c_i X_{t-i}\Big)\ \underbrace{E\big(\varepsilon_t\big|\mathcal{F}_{t-1}^{(n)}\big)}_{=\,E\varepsilon_t\,=\,0\ a.s.} = 0 \qquad \text{a.s.} \end{aligned}$$

and, analogously,

$$E\big(Z_{n,t}^2\big|\mathcal{F}_{t-1}^{(n)}\big) = \frac{1}{n}\Big(\sum_{i=1}^{p} c_i X_{t-i}\Big)^2\ \underbrace{E\big(\varepsilon_t^2\big|\mathcal{F}_{t-1}^{(n)}\big)}_{=\,E[\varepsilon_t^2]\,=\,\sigma_\varepsilon^2\ a.s.} = \frac{1}{n}\Big(\sum_{i=1}^{p} c_i X_{t-i}\Big)^2\sigma_\varepsilon^2 \qquad \text{a.s.}$$

The latter equation implies in conjunction with (i) that

$$\sum_{t=p+1}^{n} E\big(Z_{n,t}^2\big|\mathcal{F}_{t-1}^{(n)}\big) = \frac{\sigma_\varepsilon^2}{n}\, c^T X^T X c \stackrel{P}{\longrightarrow} \sigma_\varepsilon^2\, c^T\Gamma_p c.$$

It remains to check the (conditional) Lindeberg condition, that is, for all $\epsilon > 0$,

$$L_n(\epsilon) = \sum_{t=p+1}^{n} E\Big(Z_{n,t}^2\mathbb{1}(|Z_{n,t}| > \epsilon)\Big|\mathcal{F}_{t-1}^{(n)}\Big) \stackrel{P}{\longrightarrow} 0. \tag{1.5.48}$$

We will actually show that $E\big[L_n(\epsilon)\big] \underset{n\to\infty}{\longrightarrow} 0$, which yields (1.5.48) by Markov's inequality. We have that

$$
\begin{aligned}
& E\big(Z_{n,t}^2 \mathbb{1}(|Z_{n,t}| > \epsilon) \big| \mathcal{F}_{t-1}^{(n)}\big) \\
&= \frac{1}{n}\big(c_1 X_{t-1} + \cdots + c_p X_{t-p}\big)^2 \, E\Big(\varepsilon_t^2 \mathbb{1}\Big(\frac{|c_1 X_{t-1} + \cdots + c_p X_{t-p}|\, |\varepsilon_t|}{\sqrt{n}} > \epsilon\Big) \Big| \mathcal{F}_{t-1}^{(n)}\Big).
\end{aligned}
$$

Therefore, by strict stationarity and dominated convergence,

$$
E\big[L_n(\epsilon)\big] = \frac{n-p}{n} E\Big[\big(c_1 X_p + \cdots + c_p X_1\big)^2 \, \varepsilon_{p+1}^2 \, \mathbb{1}\big(|c_1 X_p + \cdots + c_p X_1|\, |\varepsilon_{p+1}| > \sqrt{n}\,\epsilon\big)\Big] \underset{n\to\infty}{\longrightarrow} 0.
$$

Indeed, a dominating integrable random variable exists since

$$
\begin{aligned}
E\Big[\big(c_1 X_p + \cdots + c_p X_1\big)^2 \, \varepsilon_{p+1}^2\Big] &= E\Big[E\Big(\big(c_1 X_p + \cdots + c_p X_1\big)^2 \, \varepsilon_{p+1}^2 \Big| \mathcal{F}_p^{(n)}\Big)\Big] \\
&= E\Big[\big(c_1 X_p + \cdots + c_p X_1\big)^2 \, \underbrace{E\big(\varepsilon_{p+1}^2 \big| \mathcal{F}_p^{(n)}\big)}_{= \sigma_\varepsilon^2 \ a.s.}\Big] \\
&= \sigma_\varepsilon^2 \, E\Big[\big(c_1 X_p + \cdots + c_p X_1\big)^2\Big] < \infty.
\end{aligned}
$$

To summarize, the conditions of Theorem 1.5.12 are fulfilled by the triangular array of random variables $Z_{n,t}$ $(t = p+1, \ldots, n,\ n \geq p+1)$ and we obtain that (1.5.47) holds true.

(iii) We split up

$$
\begin{aligned}
\sqrt{n}\big(\widetilde{\alpha}_n - \alpha\big) &= \Gamma_p^{-1} \frac{1}{\sqrt{n}} X^T \varepsilon + \Big\{\sqrt{n}(\widetilde{\alpha}_n - \alpha) - \Gamma_p^{-1} \frac{1}{\sqrt{n}} X^T \varepsilon\Big\} \\
&= \Gamma_p^{-1} \frac{1}{\sqrt{n}} X^T \varepsilon \\
&\quad + \Big[\Big(\frac{1}{n} X^T X\Big)^{-1} - \Gamma_p^{-1}\Big] \frac{1}{\sqrt{n}} X^T \varepsilon \, \mathbb{1}\big(X^T X \quad \text{regular}\big) \\
&\quad + \Big[\sqrt{n}(\widetilde{\alpha}_n - \alpha) - \Gamma_p^{-1} \frac{1}{\sqrt{n}} X^T \varepsilon\Big] \mathbb{1}\big(X^T X \quad \text{singular}\big) \\
&= \Gamma_p^{-1} Z_n + R_{n,1} + R_{n,2},
\end{aligned}
$$

say. It follows from (ii) that

$$
\Gamma_p^{-1} Z_n \stackrel{d}{\longrightarrow} Z \sim N\big(0_p, \sigma_\varepsilon^2 \Gamma_p^{-1}\big).
$$

From $\big\|(n^{-1} X^T X)^{-1} - \Gamma_p^{-1}\big\| \mathbb{1}\big(X^T X \ \text{ regular}\big) \stackrel{P}{\longrightarrow} 0$ and $n^{-1/2} X^T \varepsilon \stackrel{d}{\longrightarrow} Z_0$ we conclude

$$
\big\|R_{n,1}\big\| \stackrel{P}{\longrightarrow} 0.
$$

Finally, it follows from $P\big(X^T X \ \text{ singular}\big) \underset{n\to\infty}{\longrightarrow} 0$ that

$$
\big\|R_{n,2}\big\| \stackrel{P}{\longrightarrow} 0.
$$

This completes the proof of (iii). □

Based on the asymptotic result for the least squares estimator $\widetilde{\alpha}_n$, we can also show asymptotic normality of the Yule-Walker estimator $\widehat{\alpha}_n$. Suppose that the conditions of Theorem 1.5.13 are fulfilled. We have in particular that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0, \sigma_\varepsilon^2)$, where $\sigma_\varepsilon^2 > 0$. This implies that $EX_t = 0$ which allows us to use the following estimator of $\gamma(k)$:

$$\widehat{\gamma}_n(k) \;=\; \frac{1}{n}\sum_{t=1}^{n-|k|} X_{t+|k|}X_t,$$

i.e., there is no need to center by $\bar{X}_n$. The least squares estimator has the explicit form

$$\widetilde{\alpha}_n \;=\; \widetilde{\Gamma}_{n,p}^{-1}\,\widetilde{\gamma}_n^{(p)}, \tag{1.5.49}$$

where

$$\begin{aligned}
\widetilde{\Gamma}_{n,p} \;&=\; \frac{1}{n}X^TX \;=\; \Big(\Big(\frac{1}{n}\sum_{t=p+1}^{n} X_{t-i}X_{t-j}\Big)\Big)_{i,j=1,\dots,p},\\
\widetilde{\gamma}_n^{(p)} \;&=\; \frac{1}{n}X^TY \;=\; \Big(\frac{1}{n}\sum_{t=p+1}^{n} X_{t-1}X_t, \dots, \frac{1}{n}\sum_{t=p+1}^{n} X_{t-p}X_t\Big)^T,
\end{aligned}$$

provided the matrix $\widetilde{\Gamma}_{n,p}$ is regular. We have a (very) similar representation of the Yule-Walker estimator:

$$\widehat{\alpha}_n \;=\; \widehat{\Gamma}_{n,p}^{-1}\,\widehat{\gamma}_n^{(p)}, \tag{1.5.50}$$

where

$$\begin{aligned}
\widehat{\Gamma}_{n,p} \;&=\; \Big(\Big(\widehat{\gamma}_n(i-j)\Big)\Big)_{i,j=1,\dots,p} \;=\; \Big(\Big(\frac{1}{n}\sum_{t=\max\{i,j\}+1}^{n} X_{t-i}X_{t-j}\Big)\Big)_{i,j=1,\dots,p},\\
\widehat{\gamma}_n^{(p)} \;&=\; \Big(\widehat{\gamma}_n(1), \dots, \widehat{\gamma}_n(p)\Big)^T \;=\; \Big(\frac{1}{n}\sum_{t=2}^{n} X_{t-1}X_t, \dots, \frac{1}{n}\sum_{t=p+1}^{n} X_{t-p}X_t\Big)^T,
\end{aligned}$$

provided the matrix $\widehat{\Gamma}_{n,p}$ is regular. We will see that the difference between $\widehat{\alpha}_n$ and $\widetilde{\alpha}_n$ is of smaller order than the critical $1/\sqrt{n}$. Such rates for the convergence of sequences of random variables are most conveniently expressed by the **stochastic Landau symbol** $O_P$. For a sequence $(Y_n)_{n\in\mathbb{N}}$ of random variables and a sequence $(r_n)_{n\in\mathbb{N}}$ of positive reals we write

$$Y_n \;=\; O_P\big(r_n\big)$$

if $(Y_n/r_n)_{n\in\mathbb{N}}$ is **bounded in probability**, i.e., for all $\epsilon > 0$ there exists some $M(\epsilon) < \infty$ such that

$$P\big(|Y_n/r_n| > M(\epsilon)\big) \;\le\; \epsilon.$$

Note that $E|Y_n| = O\big(r_n\big)$ implies by Markov's inequality that $Y_n = O_P\big(r_n\big)$. Hence, it follows from

$$E\big|\widehat{\gamma}_n(k) \,-\, \widetilde{\gamma}_n(k)\big| \;=\; O\big(1/n\big)$$

that

$$\big\|\widehat{\gamma}_n^{(p)} \,-\, \widetilde{\gamma}_n^{(p)}\big\| \;=\; O_P\big(1/n\big).$$

Likewise we obtain that

$$\big\|\widehat{\Gamma}_{n,p} \,-\, \widetilde{\Gamma}_{n,p}\big\| \;=\; O_P\big(1/n\big).$$

Since $\widetilde{\Gamma}_{n,p} \xrightarrow{P} \Gamma_p$ and $\Gamma_p$ is regular we see that $\widetilde{\Gamma}_{n,p}$ and $\widehat{\Gamma}_{n,p}$ are also regular with a probability tending to 1 as $n \to \infty$. Let $A_n$ denote the event that both $\widetilde{\Gamma}_{n,p}$ and $\widehat{\Gamma}_{n,p}$ are regular. We obtain from $\widehat{\Gamma}_{n,p}^{-1} - \widetilde{\Gamma}_{n,p}^{-1} = \widehat{\Gamma}_{n,p}^{-1}\Big(\widetilde{\Gamma}_{n,p} - \widehat{\Gamma}_{n,p}\Big)\widetilde{\Gamma}_{n,p}^{-1}$ that

$$\big\|\widehat{\Gamma}_{n,p}^{-1} - \widetilde{\Gamma}_{n,p}^{-1}\big\| \, \mathbb{1}_{A_n} \leq \underbrace{\big\|\widehat{\Gamma}_{n,p}^{-1}\big\|}_{=O_P(1)} \underbrace{\big\|\widetilde{\Gamma}_{n,p} - \widehat{\Gamma}_{n,p}\big\|}_{=O_P(1/n)} \underbrace{\big\|\widetilde{\Gamma}_{n,p}^{-1}\big\|}_{=O_P(1)} \, \mathbb{1}_{A_n} = O_P\big(1/n\big).$$

This implies that

$$\begin{aligned} \Big(\widehat{\alpha}_n - \widetilde{\alpha}_n\Big) \, \mathbb{1}_{A_n} &= \Big(\widehat{\Gamma}_{n,p}^{-1}\widehat{\gamma}_n^{(p)} - \widetilde{\Gamma}_{n,p}^{-1}\widetilde{\gamma}_n^{(p)}\Big) \, \mathbb{1}_{A_n} \\ &= \underbrace{\widehat{\Gamma}_{n,p}^{-1}}_{=O_P(1)} \underbrace{\big(\widehat{\gamma}_n^{(p)} - \widetilde{\gamma}_n^{(p)}\big)}_{=O_P(1/n))} \, \mathbb{1}_{A_n} + \underbrace{\Big(\widehat{\Gamma}_{n,p}^{-1} - \widetilde{\Gamma}_{n,p}^{-1}\Big)}_{=O_P(1/n)} \underbrace{\widetilde{\gamma}_n^{(p)}}_{=O_P(1)} \, \mathbb{1}_{A_n} \\ &= O_P\big(1/n\big) \end{aligned}$$

and therefore

$$\sqrt{n}\big(\widehat{\alpha}_n - \alpha\big) = \sqrt{n}\big(\widetilde{\alpha}_n - \alpha\big) + O_P\big(1/\sqrt{n}\big) \xrightarrow{d} Z \sim N(0, \sigma_\varepsilon^2 \, \Gamma_p^{-1}).$$

## Exercise

**Ex. 1.5.5** Let $(Y_n)_{n\in\mathbb{Z}}$ and $(Z_n)_{n\in\mathbb{Z}}$ be sequences of random variables and $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ be sequences of positive real numbers such that

$$Y_n = O_P\big(r_n\big) \qquad \text{and} \qquad Z_n = O_P\big(s_n\big).$$

Show that

(i) $Y_n + Z_n = O_P\big(\max\{r_n, s_n\}\big)$,

(ii) $Y_n Z_n = O_P\big(r_n s_n\big)$.

At the end of this section, we briefly consider the case of autoregressive moving average (ARMA) processes. We recall the definition: A process $(X_t)_{t\in\mathbb{Z}}$ is said to be an **autoregressive moving average process** of order $p,q$ (ARMA$(p,q)$ process) if for every $t\in\mathbb{Z}$

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_q\varepsilon_{t-q}. \tag{1.5.51}$$

Here and in the following we assume that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$. Using the backward shift operator $B$ we can rewrite (1.5.51) as

$$\alpha(B)X_t = \beta(B)\varepsilon_t, \qquad \forall t\in\mathbb{Z},$$

where

$$\begin{aligned}\alpha(B) &= B^0 - \alpha_1 B^1 - \cdots - \alpha_p B^p,\\ \beta(B) &= B^0 + \beta_1 B^1 + \cdots + \beta_q B^q.\end{aligned}$$

As already done for the special case of autoregressive processes, we want to find sufficient conditions for the existence and uniqueness of a stationary solution to (1.5.51). An immediate answer to these questions is provided by Theorem 1.5.4. We can rewrite (1.5.51) as

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = \widetilde{\varepsilon}_t \qquad \forall t\in\mathbb{Z},$$

where $\widetilde{\varepsilon}_t := \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_p\varepsilon_{t-q}$. Since $(\varepsilon_t)_{t\in\mathbb{Z}}$ is weakly stationary, the process $(\widetilde{\varepsilon}_t)_{t\in\mathbb{Z}}$ is weakly stationary as well. Therefore, it follows from Theorem 1.5.4 that the condition

$$\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p \neq 0 \qquad \forall z\in\mathbb{C},\, |z|\leq 1$$

implies that there exists a unique stationary solution $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ to (1.5.51), where

$$\widetilde{X}_t = \sum_{k=0}^{\infty}\widetilde{\gamma}_k\widetilde{\varepsilon}_{t-k}.$$

The sequence $(\widetilde{\gamma}_k)_{k\in\mathbb{N}_0}$ is determined by the power series expansion of $1/\alpha(z)$, i.e.

$$\sum_{k=0}^{\infty}\widetilde{\gamma}_k z^k = \frac{1}{\alpha(z)} \qquad \forall z\in\mathbb{C},\, |z|\leq 1;$$

see also Lemma 1.5.3. Since the coefficients $\widetilde{\gamma}_k$ are absolutely summable we obtain that

$$\begin{aligned}\widetilde{X}_t &= \sum_{k=0}^{\infty}\widetilde{\gamma}_k\big(\varepsilon_{t-k} + \beta_1\varepsilon_{t-k-1} + \cdots + \beta_q\varepsilon_{t-k-q}\big)\\ &= \sum_{k=0}^{\infty}\underbrace{\big(\widetilde{\gamma}_k + \widetilde{\gamma}_{k-1}\beta_1 + \cdots + \widetilde{\gamma}_{k-q}\beta_q\big)}_{=:\,\gamma_k}\varepsilon_{t-k}\\ &= \beta(B)\,\widetilde{\gamma}(B)\,\varepsilon_t.\end{aligned}$$

(On the second line of this display we set $\widetilde{\gamma}_k = 0$, for $k<0$.)

If the polynomials $\alpha(\cdot)$ and $\beta(\cdot)$ have common roots, then the condition of $\alpha(z) \neq 0$ $\forall z \in \mathbb{C},\ |z| \leq 1$ can be (slightly) relaxed:

**Theorem 1.5.14.** *Suppose that the polynomials $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p$ and $\beta(z) = 1 + \beta_1 z^1 + \cdots + \beta_q z^q$ can be represented as*

$$\alpha(z) = \xi(z)\,\alpha_0(z) \qquad \text{and} \qquad \beta(z) = \xi(z)\,\beta_0(z)$$

*and that*

$$\alpha_0(z) \neq 0 \qquad \forall z \in \mathbb{C},\ |z| \leq 1.$$

*(i) There exists an absolutely summable sequence $(\gamma_k)_{k\in\mathbb{Z}}$ such that*

$$\gamma(z) = \sum_{k=0}^{\infty} \gamma_k z^k = \frac{\beta_0(z)}{\alpha_0(z)} \qquad \forall z \in \mathbb{C},\ |z| \leq 1. \tag{1.5.52}$$

*(ii) If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a weakly stationary process, then there exists a causal stationary solution $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ to (1.5.51) such that*

$$\widetilde{X}_t = \sum_{k=0}^{\infty} \gamma_k \varepsilon_{t-k}.$$

*If $\alpha(z) \neq 0$ $\forall z \in \mathbb{C},\ |z| \leq 1$, then $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ is the unique stationary solution.*

*Proof.* (i) It follows from Lemma 1.5.3 that there exists an absolutely summable sequence $(\widetilde{\gamma}_k)_{k\in\mathbb{N}_0}$ such that

$$\widetilde{\gamma}(z) = \sum_{k=0}^{\infty} \widetilde{\gamma}_k z^k = \frac{1}{\alpha_0(z)} \qquad \forall z \in \mathbb{C},\ |z| \leq 1.$$

The polynomial $\beta_0(\cdot)$ can be written as $\beta_{00} + \beta_{01} z^1 + \cdots + \beta_{0l} z^l$, for some $l \leq q$. Let $\gamma_k = \widetilde{\gamma}_k \beta_{00} + \widetilde{\gamma}_{k-1}\beta_{01} + \cdots + \widetilde{\gamma}_{k-q}\beta_{0l}$ ($\widetilde{\gamma}_k = 0$, for $k < 0$). Then the sequence $(\gamma_k)_{k\in\mathbb{N}_0}$ is also absolutely summable and a comparison of coefficients reveals that

$$\gamma(z) = \sum_{k=0}^{\infty} \gamma_k z^k = \widetilde{\gamma}(z)\,\beta_0(z) = \frac{\beta_0(z)}{\alpha_0(z)} \qquad \forall z \in \mathbb{C},\ |z| \leq 1.$$

(ii) Note that $\widetilde{\gamma}(B)\varepsilon_t = \sum_{k=0}^{\infty} \widetilde{\gamma}_k \varepsilon_{t-k}$ converges absolutely with probability one. Since $\widetilde{X}_t = \sum_{k=0}^{\infty} \gamma_k \varepsilon_{t-k} = \beta_0(B)\widetilde{\gamma}(B)\varepsilon_t$ we obtain that

$$\begin{aligned}
\alpha_0(B)\widetilde{X}_t &= \alpha_0(B)\beta_0(B)\widetilde{\gamma}(B)\varepsilon_t \\
&= \beta_0(B)\underbrace{\alpha_0(B)\widetilde{\gamma}(B)}_{=B^0}\varepsilon_t = \beta_0(B)\varepsilon_t
\end{aligned}$$

and therefore

$$\underbrace{\xi(B)\alpha_0(B)}_{=\alpha(B)}\widetilde{X}_t = \underbrace{\xi(B)\beta_0(B)}_{=\beta(B)}\varepsilon_t.$$

Hence, $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ is a causal stationary solution to (1.5.51).

Assume now that $\alpha(z) \neq 0 \ \forall z \in \mathbb{C}, \ |z| \leq 1$ and that $(X_t)_{t\in\mathbb{Z}}$ is an arbitrary stationary solution to (1.5.51). Then

$$\alpha(B)\widetilde{X}_t \ = \ \beta(B)\varepsilon_t \ =: \ \widetilde{\varepsilon}_t$$

and

$$\alpha(B)X_t \ = \ \beta(B)\varepsilon_t \ = \ \widetilde{\varepsilon}_t.$$

It follows from Theorem 1.5.4(ii), applied to the innovation sequence $(\widetilde{\varepsilon}_t)_{t\in\mathbb{Z}}$, that

$$P\big(X_t \ = \ \widetilde{X}_t\big) \ = \ 1.$$

□

**Remark 1.5.15.** *If at least one of the common zeroes of $\alpha(\cdot)$ and $\beta(\cdot)$ lies on the unit circle, then the ARMA equations may have more than one stationary solution.*

*Example:*
*Suppose that $\alpha(1) = \beta(1) = 0$. Then*

$$\alpha(z) \ = \ \big(1 \ - \ z\big)\alpha_0(z) \qquad \textit{and} \qquad \beta(z) \ = \ \big(1 \ - \ z\big)\beta_0(z).$$

*Assume in addition that $\alpha_0(z) \neq 0 \ \forall z \in \mathbb{C}, \ |z| \leq 1$ and that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$. Then, a causal stationary solution $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ to (1.5.51) is given by*

$$\widetilde{X}_t \ = \ \sum_{k=0}^{\infty} \gamma_k \varepsilon_{t-k},$$

*where $\gamma(z) \ = \ \beta_0(z)/\alpha_0(z)$. Let $Z$ be an arbitrary random variable. Then, for $X_t := \widetilde{X}_t + Z$,*

$$\alpha(B)X_t \ = \ \alpha(B)\widetilde{X}_t \ + \ \alpha_0(B)\underbrace{\big(B^0 \ - \ B^1\big)Z}_{\equiv 0} \ = \ \beta(B)\varepsilon_t,$$

*i.e. $(X_t)_{t\in\mathbb{Z}}$ is also a solution to (1.5.51). If $Z$ is independent of $(\widetilde{X}_t)_{t\in\mathbb{Z}}$, $EZ^2 < \infty$, then $(X_t)_{t\in\mathbb{Z}}$ is also weakly stationary.*

## Exercise

**Ex. 1.5.6** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$ and let $(\widetilde{X}_t)_{t\in\mathbb{Z}}$ be a causal stationary solution to

$$X_t \ - \ \alpha X_{t-1} \ = \ \varepsilon_t \ + \ \beta\varepsilon_{t-1} \qquad \forall t \in \mathbb{Z},$$

where $|\alpha| < 1$.

Determine the coefficients $\gamma_k$ such that $\widetilde{X}_t = \sum_{k=0}^{\infty} \gamma_k \varepsilon_k$. How do these coefficients look like if $\beta = -\alpha$?

## 1.6 GARCH processes

As we have seen so far, ARMA processes are used to model the conditional mean given the past. However, within this class of models, the conditional variance given the past is constant. In contrast, financial time series (stock returns etc.) can often be modeled with white noise processes but with conditional variances that depend on past values of the process. In particular, it is frequently observed that periods with large conditional variances ("high volatility") alternate with periods with small conditional variances ("low volatility"). In 1982 the American economist and statistician Robert F. Engle introduced so-called ARCH processes, where ARCH is an acronym meaning **A**uto**R**egressive **C**onditional **H**eteroskedasticity. Four years later, the Danish economist Tim Bollerslev, who was at that time a PhD student of Engle, generalized this approach and introduced GARCH (**G**eneralized **ARCH**) processes. These models became quite popular in financial mathematics. In 2003 Robert F. Engle was awarded the (shared) Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, commonly termed Nobel Prize, "for methods of analyzing time series with time-varying volatility (ARCH)".

There are many types of GARCH processes. In this course, we restrain ourselves to **linear** GARCH processes which are still the most popular ones and may also be viewed as **the** GARCH processes. Here is a possible definition of this class of models:

**Definition.** A process $(X_t)_{t\in\mathbb{Z}}$ on a probability space $(\Omega, \mathcal{F}, P)$ is called **GARCH process** of order $p$ and $q$ (GARCH($p$,$q$)) if

$$X_t = \sigma_t \, \varepsilon_t \qquad \forall t \in \mathbb{Z} \tag{1.6.1a}$$

and

$$\sigma_t^2 = \alpha + \phi_1 \sigma_{t-1}^2 + \cdots + \phi_p \sigma_{t-p}^2 + \theta_1 X_{t-1}^2 + \cdots + \theta_q X_{t-q}^2 \qquad \forall t \in \mathbb{Z}, \tag{1.6.1b}$$

where $\alpha > 0$, $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \geq 0$ and $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0,1)$.

If the coefficients $\phi_1, \ldots, \phi_p$ all vanish, then $(X_t)_{t\in\mathbb{Z}}$ is an **ARCH process** of order $q$ (ARCH($q$)=GARCH(0,$q$)).

At this point we can already conjecture why GARCH processes are quite popular in financial mathematics. Let $P_t$ be the price of a financial asset (such as a stock) at day $t$. Then the return $X_t$ of "buying yesterday and selling today" is given by

$$X_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Financial data show that, to a good approximation,

$$E\big(X_t \mid \text{"past"}\big) = 0.$$

On the other hand, the market becomes volatile whenever big news comes (e.g. unexpected quarter results or a profit warning), and it takes several periods for the market to fully digest the news. This feature is obviously captured by a GARCH model: The conditional variance of $X_t$ given the past is equal to $\sigma_t^2$, which does depend on past values of both the squared return process $(X_t^2)_{t\in\mathbb{Z}}$ and the process $(\sigma_t^2)_{t\in\mathbb{Z}}$. Since the coefficients $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are non-negative we see from (1.6.1b) that large values of $\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2$ and $X_{t-1}^2, \ldots, X_{t-q}^2$ are followed by a large value of $\sigma_t^2$. This effect is also called volatility clustering in financial mathematics.

As in the case of ARMA processes, the random variables $X_t$ appear both on the left- and the right-hand sides of the model equations (1.6.1a) and (1.6.1b). Although the underlying innovation process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be strictly stationary, it is not obvious if there exists, for a given process $(\varepsilon_t)_{t\in\mathbb{Z}}$, a stationary solution to these equations. We will see that a sufficient condition for the existence and uniqueness of a strictly stationary solution is given by

$$\phi_1 + \cdots + \phi_p + \theta_1 + \cdots + \theta_q \; < \; 1. \tag{1.6.2}$$

First we intend to guess a possible solution $\big((\widetilde{X}_t, \widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$. Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a given sequence of innovations and suppose that (1.6.2) is fulfilled. Let, for simplicity of notation, $p = q$. (Otherwise, we set $\phi_{p+1} = \ldots = \phi_q := 0$ if $p < q$ or $\theta_{q+1} = \ldots = \theta_p := 0$ if $q < p$.) If there exists a (stationary or non-stationary) solution $\big((X_t, \sigma_t)\big)_{t\in\mathbb{Z}}$ at all, then it follows from (1.6.1a) and (1.6.1b) that

$$\begin{aligned}
\sigma_t^2 \;&=\; \alpha + \sum_{k=1}^p \Big(\phi_k \sigma_{t-k}^2 \,+\, \theta_k \underbrace{X_{t-k}^2}_{=\varepsilon_{t-k}^2\sigma_{t-k}^2}\Big) \\
&=\; \alpha + \sum_{k=1}^p \big(\phi_k + \theta_k\varepsilon_{t-k}^2\big) \underbrace{\sigma_{t-k}^2}_{=\alpha+\sum_{j=1}^p(\phi_j+\theta_j\varepsilon_{t-k-j}^2)\sigma_{t-k-j}^2} \\
&=\; \alpha\Big\{1 \,+\, \big(\phi_1 + \theta_1\varepsilon_{t-1}^2\big) \,+\, \cdots \,+\, \big(\phi_p + \theta_p\varepsilon_{t-p}^2\big) \,+\, \big(\phi_1 + \theta_1\varepsilon_{t-1}^2\big)\big(\phi_1 + \theta_1\varepsilon_{t-1-1}^2\big) \,+\, \cdots\Big\}.
\end{aligned}$$

In view of this, a reasonable **guess** for a possible solution to (1.6.1a) and (1.6.1b) is given by

$$\widetilde{\sigma}_t^2 \;:=\; \alpha\left\{1 + \sum_{r=1}^\infty \sum_{k_1,\ldots,k_r=1}^p \big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\right\} \tag{1.6.3a}$$

and

$$\widetilde{X}_t \;:=\; \widetilde{\sigma}_t\varepsilon_t. \tag{1.6.3b}$$

The next theorem shows that $\big((\widetilde{X}_t, \widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$ actually solves our system of model equations.

**Theorem 1.6.1.** *Suppose that* $\phi_1+\cdots+\phi_p+\theta_1+\cdots+\theta_q \;<\; 1$ *and that* $\big(\varepsilon_t\big)_{t\in\mathbb{Z}} \sim IID(0,1)$. *Then the system of equations (1.6.1a) and (1.6.1b) has a unique strictly stationary solution* $\big((\widetilde{X}_t, \widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$ *which is given by (1.6.3a) and (1.6.3b).*

*Proof.* First of all, we show that the infinite series on the right-hand side of (1.6.3a) converges with probability 1. Since

$$\begin{aligned}
&E\Big[\big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\Big] \\
&\quad=\; E\Big[E\Big(\big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\Big|\varepsilon_{t-k_1-1}, \varepsilon_{t-k_1-2}, \ldots\Big)\Big] \\
&\quad=\; E\Big[\big(\phi_{k_2} + \theta_{k_2}\varepsilon_{t-k_1-k_2}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big) \,\times \\
&\qquad\qquad \times \underbrace{E\Big(\big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\Big|\varepsilon_{t-k_1-1}, \varepsilon_{t-k_1-2}, \ldots\Big)}_{=E[\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2]=\phi_{k_1}+\theta_{k_1}}\Big] \\
&\quad=\; \ldots \;=\; \big(\phi_{k_1} + \theta_{k_1}\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\big)
\end{aligned}$$

we obtain

$$E\left[1 + \sum_{r=1}^{\infty} \sum_{k_1,\ldots,k_r=1}^{p} \big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\right]$$

$$= 1 + \sum_{r=1}^{\infty} \sum_{k_1,\ldots,k_r=1}^{p} E\Big[\big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\Big]$$

$$= 1 + \sum_{r=1}^{\infty} \sum_{k_1,\ldots,k_r=1}^{p} \big(\phi_{k_1} + \theta_{k_1}\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\big)$$

$$= 1 + \sum_{r=1}^{\infty} \Big(\big(\phi_1 + \theta_1\big) + \cdots + \big(\phi_p + \theta_p\big)\Big)^r$$

$$= \frac{1}{1 - (\phi_1 + \cdots + \phi_p + \theta_1 + \cdots \theta_q)} < \infty.$$

Hence, the infinite series on the right-hand side of (1.6.3a) converges with probability 1.

Next we show strict stationarity of the process $\big((\widetilde{X}_t, \widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$. To this end, we first consider the truncated versions,

$$\widetilde{\sigma}_{t,m}^2 = \alpha \left\{1 + \sum_{r=1}^{m} \sum_{k_1,\ldots,k_r=1}^{p} \big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-k_2-\cdots-k_r}^2\big)\right\}$$

and

$$\widetilde{X}_{t,m} = \widetilde{\sigma}_{t,m}\varepsilon_t.$$

Then $\widetilde{\sigma}_{t,m}^2 = g_m\big(\varepsilon_{t-1}, \ldots, \varepsilon_{t-mp}\big)$ and $\widetilde{X}_{t,m} = h_m\big(\varepsilon_{t-1}, \ldots, \varepsilon_{t-mp}\big)$, for suitable functions $h_m \colon \mathbb{R}^{mp} \to [0,\infty)$ and $g_m \colon \mathbb{R}^{mp+1} \to \mathbb{R}$. Since the underlying process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary we obtain that

$$P^{(\widetilde{X}_{t_1,m},\widetilde{\sigma}_{t_1,m}),\ldots,(\widetilde{X}_{t_k,m},\widetilde{\sigma}_{t_k,m})} = P^{(\widetilde{X}_{t+t_1,m},\widetilde{\sigma}_{t+t_1,m}),\ldots,(\widetilde{X}_{t+t_k,m},\widetilde{\sigma}_{t+t_k,m})} \qquad \forall t, t_1, \ldots, t_k \in \mathbb{Z}, \forall k \in \mathbb{N},$$

i.e. the finite-dimensional distributions of $\big((\widetilde{X}_{t,m}, \widetilde{\sigma}_{t,m})\big)_{t\in\mathbb{Z}}$ are shift-invariant. Since $\widetilde{\sigma}_{t,m} \xrightarrow{a.s.} \widetilde{\sigma}_t$ and $\widetilde{X}_{t,m} \xrightarrow{a.s.} \widetilde{X}_t$ we conclude that the finite-dimensional distributions of the process $\big((\widetilde{X}_t, \widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$ are shift-invariant as well which means that this process is strictly stationary.

We can easily see that $\big(\widetilde{\sigma}_t^2\big)_{t\in\mathbb{Z}}$ solves the system of equations (1.6.1b). Indeed, we have that

$$\alpha + \sum_{k=1}^{p} \phi_k \widetilde{\sigma}_{t-k}^2 + \theta_k \widetilde{X}_{t-k}^2$$

$$= \alpha + \sum_{k=1}^{p} \big(\phi_k + \theta_k \varepsilon_{t-k}^2\big)\widetilde{\sigma}_{t-k}^2$$

$$= \alpha + \sum_{k=1}^{p} \big(\phi_k + \theta_k \varepsilon_{t-k}^2\big)\, \alpha \left\{1 + \sum_{r=1}^{\infty} \sum_{k_1,\ldots,k_r=1}^{p} \big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k-k_1-\cdots-k_r}^2\big)\right\}$$

$$= \alpha \left\{1 + \sum_{r=1}^{\infty} \sum_{k_1,\ldots,k_r=1}^{p} \big(\phi_{k_1} + \theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r} + \theta_{k_r}\varepsilon_{t-k_1-\cdots-k_r}^2\big)\right\} = \widetilde{\sigma}_t^2.$$

Uniqueness of the strictly stationary solution can be most easily seen in the special case of a GARCH(1,1) process. Let $\big((\widehat{X}_t,\widehat{\sigma}_t)\big)_{t\in\mathbb{Z}}$ be any arbitrary strictly stationary solution to (1.6.1a) and (1.6.1b). Using the model equation we obtain

$$\begin{aligned}
\widetilde{\sigma}_t^2 &= \alpha + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big) \underbrace{\widetilde{\sigma}_{t-1}^2}_{=\alpha+\big(\phi_1+\theta_1\varepsilon_{t-2}\big)\widetilde{\sigma}_{t-2}^2} \\
&= \ldots = \alpha\Big\{1 + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big) + \cdots + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m}^2\big)\Big\} \\
&\qquad + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m-1}^2\big)\widetilde{\sigma}_{t-m-1}^2
\end{aligned}$$

and, analogously,

$$\begin{aligned}
\widehat{\sigma}_t^2 &= \alpha\Big\{1 + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big) + \cdots + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m}^2\big)\Big\} \\
&\qquad + \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m-1}^2\big)\widehat{\sigma}_{t-m-1}^2.
\end{aligned}$$

Therefore,

$$\big|\widehat{\sigma}_t^2 - \widetilde{\sigma}_t^2\big| = \big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m-1}^2\big)\big|\widehat{\sigma}_{t-m-1}^2 - \widetilde{\sigma}_{t-m-1}^2\big|. \tag{1.6.4}$$

Since $\big(\phi_1+\theta_1\varepsilon_{t-1}^2\big)\cdots\big(\phi_1+\theta_1\varepsilon_{t-m-1}^2\big) \xrightarrow{P} 0$ and since $\big(\widetilde{\sigma}_t^2\big)_{t\in\mathbb{Z}}$ and $\big(\widehat{\sigma}_t^2\big)_{t\in\mathbb{Z}}$ are both sequences of identically distributed random variables we conclude that the right-hand side of (1.6.4) converges in probability to 0 as $m\to\infty$. This, however, implies that

$$P\Big(\widehat{\sigma}_t^2 \neq \widetilde{\sigma}_t^2\Big) = 0$$

and

$$P\Big(\widehat{X}_t^2 \neq \widetilde{X}_t^2\Big) = P\Big(\widehat{\sigma}_t^2\varepsilon_t^2 \neq \widetilde{\sigma}_t^2\varepsilon_t^2\Big) = 0.$$

The proof of uniqueness in the general case is similar but the corresponding calculations are more cumbersome. Applying again the model equations (1.6.1a) and (1.6.1b) to $\widetilde{\sigma}_t^2$ we replace successively the terms $\widetilde{\sigma}_{t-j}^2$ on the right-hand side and we stop replacing when some of the factors $(\phi_1+\theta_1\varepsilon_{t-m-1}^2)\widetilde{\sigma}_{t-m-1}^2,\ldots,(\phi_p+\theta_p\varepsilon_{t-m-p}^2)\widetilde{\sigma}_{t-m-p}^2$ pop up. We obtain, for $m\in\mathbb{N}$,

$$\begin{aligned}
\widetilde{\sigma}_t^2 &= \alpha + \sum_{k=1}^p \big(\phi_k+\theta_k\varepsilon_{t-k}^2\big) \underbrace{\widetilde{\sigma}_{t-k}^2}_{=\alpha+\sum_{j=1}^p(\phi_j+\theta_j\varepsilon_{t-j}^2)\widetilde{\sigma}_{t-k-j}^2} \\
&= \alpha\Big\{1 + \sum_{r=1}^m \sum_{(k_1,\ldots,k_r):\ k_1+\cdots+k_r\le m} \big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\cdots-k_r}^2\big)\Big\} \\
&\qquad + \sum_{j=1}^p\sum_{r=1}^{m+j} \sum_{(k_1,\ldots,k_r):\ k_1+\cdots+k_r=m+j} \big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\cdots-k_r}^2\big)\widetilde{\sigma}_{t-m-j}^2
\end{aligned}$$

and, analogously,

$$\begin{aligned}
\widehat{\sigma}_t^2 &= \alpha\Big\{1 + \sum_{r=1}^m \sum_{(k_1,\ldots,k_r):\ k_1+\cdots+k_r\le m} \big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\cdots-k_r}^2\big)\Big\} \\
&\qquad + \sum_{j=1}^p\sum_{r=1}^{m+j} \sum_{(k_1,\ldots,k_r):\ k_1+\cdots+k_r=m+j} \big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\cdots-k_r}^2\big)\widehat{\sigma}_{t-m-j}^2.
\end{aligned}$$

This implies that

$$\begin{aligned}&\big|\widehat{\sigma}_t^2 - \widetilde{\sigma}_t^2\big| \\ &\le \sum_{j=1}^{p}\sum_{r=1}^{m+j}\sum_{(k_1,\dots,k_r):\ k_1+\dots+k_r=m+j}\big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\dots-k_r}^2\big)\big|\widehat{\sigma}_{t-m-j}^2 - \widetilde{\sigma}_{t-m-j}^2\big|.\end{aligned}$$

Since $\sum_{j=1}^{p}\sum_{r=1}^{m+j}\sum_{(k_1,\dots,k_r):\ k_1+\dots+k_r=m+j}\big(\phi_{k_1}+\theta_{k_1}\varepsilon_{t-k_1}^2\big)\cdots\big(\phi_{k_r}+\theta_{k_r}\varepsilon_{t-k_1-\dots-k_r}^2\big)\xrightarrow{P}0$ as $m\to\infty$ and since $\big(\widetilde{\sigma}_t^2\big)_{t\in\mathbb{Z}}$ and $\big(\widehat{\sigma}_t^2\big)_{t\in\mathbb{Z}}$ are both sequences of identically distributed random variables we conclude that the right-hand side of the above display converges in probability to 0 as $m\to\infty$. Therefore we can obtain as above that

$$P\Big(\widehat{\sigma}_t^2 \ne \widetilde{\sigma}_t^2\Big) = 0$$

and

$$P\Big(\widehat{X}_t^2 \ne \widetilde{X}_t^2\Big) = P\Big(\widehat{\sigma}_t^2\varepsilon_t^2 \ne \widetilde{\sigma}_t^2\varepsilon_t^2\Big) = 0.$$

□

## Exercise

**Ex. 1.14** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}(0,\sigma^2)$ and let

$$\begin{aligned}\widetilde{\sigma}_t^2 &= \alpha\Big\{1 + \sum_{k=1}^{\infty}\theta^k\,\varepsilon_{t-1}^2\cdots\varepsilon_{t-k}^2\Big\},\\ \widetilde{X}_t &= \widetilde{\sigma}_t\varepsilon_t\end{aligned}$$

be such that $\big((\widetilde{X}_t,\widetilde{\sigma}_t)\big)_{t\in\mathbb{Z}}$ is the unique strictly stationary solution to

$$X_t = \sigma_t\varepsilon_t, \qquad \sigma_t^2 = \alpha + \theta X_{t-1}^2 \qquad \forall t\in\mathbb{Z},$$

where $\theta\in[0,1)$.

(i) Compute $E\big[\widetilde{\sigma}_t^2\big]$.

(ii) Compute $E\widetilde{X}_t$, $\mathrm{var}\,\big(\widetilde{X}_t^2\big)$, and $\mathrm{cov}(\widetilde{X}_{t+k},\widetilde{X}_t)$ for $k\ge 1$.

(iii) Suppose additionally that $E\big[\varepsilon_t^4\big] := \kappa < \infty$ and compute $E\big[\widetilde{\sigma}_t^4\big]$ and $E\big[\widetilde{X}_t^4\big]$.

# 2 Spectral analysis of stationary processes

## 2.1 Spectral density, spectral distribution function, spectral measure

We suppose that $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a weakly stationary process with autocovariance function $\gamma_X$, i.e. $\gamma_X(k) = \mathrm{cov}(X_{t+k}, X_t)$. For the time being we assume that

$$\sum_{k=-\infty}^{\infty} \big|\gamma_X(k)\big| \;<\; \infty.$$

The function $f_X\colon\ [-\pi,\pi] \longrightarrow \mathbb{R}$ with

$$f_X(\lambda) \;:=\; \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_X(k)e^{-ik\lambda} \;=\; \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_X(k)\cos(k\lambda)$$

is called the **spectral density** of the process $\mathbf{X}$. Because it is periodic with period $2\pi$ it suffices to consider it on an interval of length $2\pi$, which we shall take to be $[-\pi,\pi]$. In the present context the values $\lambda$ in this interval are often referred to **frequencies**, for reasons that become clear in what follows.

Before we proceed, we consider a few examples. We suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma_\varepsilon^2)$, i.e., $(\varepsilon_t)_{t\in\mathbb{Z}}$ is weakly stationary, $E\varepsilon_t = 0$, $\mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$, and $\mathrm{cov}(\varepsilon_{t+k},\varepsilon_t) = 0$ if $k \neq 0$.

1) **White noise**

Since $\gamma_\varepsilon(0) = \sigma_\varepsilon^2$ and $\gamma_\varepsilon(k) = 0$ if $k \neq 0$ we obtain that

$$f_\varepsilon(\lambda) \;=\; \frac{\sigma_\varepsilon^2}{2\pi} \qquad \forall \lambda \in [-\pi,\pi].$$

2) **Linear processes**

Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary process with an absolutely summable autocovariance function $\gamma_\varepsilon$ and a spectral density $f_\varepsilon$. Let $X_t = \sum_{k=-\infty}^{\infty} \beta_k\varepsilon_{t-k}$, where $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$.
Then the process $(X_t)_{t\in\mathbb{Z}}$ is also stationary and has a spectral density $f_X$, where

$$f_X(\lambda) \;=\; \Big|\sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda}\Big|^2 \, f_\varepsilon(\lambda) \qquad \forall \lambda \in [-\pi,\pi]. \tag{2.1.1}$$

To see this, note that it follows from Proposition 1.4.3 that $(X_t)_{t\in\mathbb{Z}}$ has an absolutely summable autocovariance function $\gamma_X$, where

$$\gamma_X(h) \;=\; \sum_{j,k=-\infty}^{\infty} \beta_j\beta_k\gamma_\varepsilon(h-j+k) \qquad \forall h \in \mathbb{Z}$$

Hence, $(X_t)_{t\in\mathbb{Z}}$ has a spectral density $f_X$ which is given by

$$\begin{aligned} f_X(\lambda) \;&=\; \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \gamma_X(h)\, e^{-ih\lambda} \\ &=\; \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\sum_{j,k=-\infty}^{\infty} \beta_j\beta_k\gamma_\varepsilon(h-j+k)\, e^{-ih\lambda}. \end{aligned}$$

Since $\sum_{h,j,k} |\beta_j||\beta_k||\gamma_\epsilon(h-j+k)| = \left(\sum_j |\beta_j|\right)^2 \sum_h |\gamma_\epsilon(h)| < \infty$ we can change the order of summation and we obtain that

$$\begin{aligned} f_X(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j,k=-\infty}^{\infty} \beta_j e^{-ij\lambda} \, \beta_k e^{ik\lambda} \, \gamma_\varepsilon(h-j+k) \, e^{-i(h-j+k)\lambda} \\ &= \Big| \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda} \Big|^2 \underbrace{\sum_{h=-\infty}^{\infty} \gamma_\epsilon(h) e^{-ih\lambda}}_{= f_\varepsilon(\lambda)} . \end{aligned}$$

We can use (2.1.1) for deriving the spectral densities of stationary ARMA processes. Suppose now that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$.

3a) **MA(q) processes**

Let

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \qquad \forall t \in \mathbb{Z}.$$

It follows from Example 1) and (2.1.1) that

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| 1 + \beta_1 e^{-i\lambda} + \cdots + \beta_q e^{-iq\lambda} \right|^2 \qquad \forall \lambda \in [-\pi, \pi].$$

3b) **AR(p) processes**

Suppose that $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p \neq 0 \ \ \forall z \in \mathbb{C}, \, |z| \leq 1$. Let $(X_t)_{t\in\mathbb{Z}}$ with

$$X_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k} \qquad \forall t \in \mathbb{Z}$$

be the unique stationary solution to

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \qquad \forall t \in \mathbb{Z}.$$

According to Lemma 1.5.3 and Theorem 1.5.4, the $\beta_k$ are absolutely summable and we obtain from (2.1.1) that

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \Big| \sum_{k=0}^{\infty} \beta_k e^{-ik\lambda} \Big|^2 \qquad \forall \lambda \in [-\pi, \pi].$$

Since $\sum_{k=0}^{\infty} \beta_k z^k = 1/\alpha(z)$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$ and since $|e^{-i\lambda}| = 1$ we can rewrite $f_X(\lambda)$ as

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{\left| 1 - \alpha_1 e^{-i\lambda} - \cdots - \alpha_p e^{-ip\lambda} \right|^2} \qquad \forall \lambda \in (-\pi, \pi].$$

Some properties of a spectral density are summarized in the following lemma.

**Lemma 2.1.1.** *Let* $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ *be a stationary real-valued process with an absolutely summable autocovariance function* $\gamma_X$*. Then*

(i) $f_X$ *is uniformly continuous on* $[-\pi, \pi]$,

(ii) $f_X(\lambda) = f_X(-\lambda) \qquad \forall \lambda \in [-\pi, \pi]$,

(iii) $f_X(\lambda) \geq 0 \qquad \forall \lambda \in [-\pi, \pi]$,

(iv) $\gamma_X(\lambda) = \int_{-\pi}^{\pi} e^{ik\lambda} f_X(\lambda)\, d\lambda \qquad \forall k \in \mathbb{Z}$.

*Proof.* (i) Let $\epsilon > 0$ be arbitrary. Then

$$\begin{aligned}
\big|f_X(\lambda) - f_X(\omega)\big| &= \Big|\frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_X(k)\big\{\cos(\lambda k) - \cos(\omega k)\big\}\Big| \\
&\leq \frac{1}{2\pi}\sum_{k=-K}^{K} \big|\gamma_X(k)\big| \underbrace{\big|\cos(\lambda k) - \cos(\omega k)\big|}_{\leq |k||\lambda-\omega|} + \frac{1}{\pi}\sum_{k:\,|k|>K} \big|\gamma_X(k)\big|.
\end{aligned}$$

The second term on the right-hand side is not greater than $\epsilon/2$ if $K = K(\epsilon)$ is sufficiently large. For such a $K$, the first term is also less than or equal to $\epsilon/2$ if $|\lambda - \omega| \leq \delta = \delta(\epsilon)$. Hence,

$$\big|f_X(\lambda) - f_X(\omega)\big| \leq \epsilon \qquad \forall \lambda, \omega \in [-\pi, \pi],\ |\lambda - \omega| \leq \delta,$$

i.e., $f_X$ is uniformly continuous.
(ii) This is an immediate consequence of $\gamma_X(k) = \gamma_X(-k)$.
(iii) Let $\mu := EX_t$. Then

$$\begin{aligned}
0 &\leq E\Big[\frac{1}{2\pi n}\Big|\sum_{t=1}^{n}(X_t - \mu)e^{-it\lambda}\Big|^2\Big] \\
&= \frac{1}{2\pi n} E\Big[\sum_{s,t=1}^{n}(X_s - \mu)(X_t - \mu)e^{-i(s-t)\lambda}\Big] \\
&= \frac{1}{2\pi n}\sum_{s,t=1}^{n} \gamma_X(s-t)\cos\big((s-t)\lambda\big) \\
&= \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1} \frac{n-|k|}{n}\gamma_X(k)\cos(k\lambda) \\
&\underset{n\to\infty}{\longrightarrow} \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_X(k)\cos(k\lambda) = f_X(\lambda).
\end{aligned}$$

Note that the convergence follows by Lebesgue's dominated convergence theorem. Hence, $f_X$ is non-negative.

(iv) We obtain by Fubini's theorem that

$$\begin{aligned}\int_{-\pi}^{\pi} e^{ik\lambda} f_X(\lambda)\, d\lambda &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \sum_{l=-\infty}^{\infty} \gamma_X(l) e^{-il\lambda}\, d\lambda \\ &= \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_X(l) \underbrace{\int_{-\pi}^{\pi} e^{i(k-l)\lambda}\, d\lambda}_{=2\pi\delta_{k,l}} = \gamma_X(k).\end{aligned}$$

Note that we can actually change the order of integration/summation since

$$\int_{-\pi}^{\pi} \sum_{l=-\infty}^{\infty} \left|e^{ik\lambda}\gamma_X(l)e^{-il\lambda}\right| d\lambda \le \int_{-\pi}^{\pi} \sum_{l=-\infty}^{\infty} \left|\gamma_X(l)\right| d\lambda = 2\pi \sum_{l=-\infty}^{\infty} \left|\gamma_X(l)\right| < \infty.$$

□

We have seen that absolute summability of the autocovariance function of a stationary process makes the definition of the spectral density possible. Furthermore, assertion (iv) of Lemma 2.1.1 shows that there is a one-to-one relation between absolutely summable autocovariance functions and the corresponding spectral densities. This means that a spectral density provides a complete description of the (second order) dependence structure.

As for probability densities, we can also define for spectral densities the corresponding counterparts of a probability distribution function and a probability measure. Suppose that $f_X$ is the spectral density of a real-valued stationary process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$. Then $F_X\colon\ [-\pi,\pi] \to [0,\infty)$ defined by

$$F_X(\lambda) := \int_{-\pi}^{\lambda} f_X(\omega)\, d\omega \qquad \forall \lambda \in [-\pi,\pi]$$

is the **spectral distribution function** of $\mathbf{X}$. Furthermore, there exists a measure $\mu_X$ on $\mathcal{B}\big|_{(-\pi,\pi]} := \left\{B \cap (-\pi,\pi]\colon\ \ B \in \mathcal{B}\right\}$ (the **trace** of $\mathcal{B}$ on $(-\pi,\pi]$ or trace $\sigma$-algebra) such that

$$\mu_X((a,b]) = F_X(b) - F_X(a) \qquad \forall a,b \in [-\pi,\pi],\ \ a \le b.$$

$\mu_X$ is said to be the **spectral measure** of the process $\mathbf{X}$. It follows from Lemma 2.1.1 that

$$\gamma_X(k) = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF_X(\lambda) = \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu_X(\lambda) \qquad \forall k \in \mathbb{Z}. \tag{2.1.2}$$

$F_X$ and $\mu_X$ inherit the property of symmetry about 0 from $f_X$. We have that $F(b) - F(a) = F(-a) - F(-b)$ for $a,b \in [-\pi,\pi]$ and $\mu_X(B) = \mu_X(-B)$ for $B \in \mathcal{B}\big|_{(-\pi,\pi]}$, $B \subseteq (-\pi,\pi)$. To summarize, if a stationary process has absolutely summable autocovariances, then these autocovariances can be described either by the spectral density, by the spectral distribution function or by the spectral measure. In what follows we intend to relax the condition of absolute summability of the autocovariances. It will turn out that there still exist a distribution function $F_X$ and a measure $\mu_X$ such that (2.1.2) is satisfied. However, without absolute summability, it could be the case that a spectral density does not exist. It could also happen that the measure $\mu_X$ is not fully symmetric about 0 if $\mu_X(\{\pi\}) > 0$. Of course, (2.1.2) remains true if we redistribute some mass of $\mu_X$ by setting $\mu'(\{-\pi\}) = \mu'(\{\pi\}) := \mu(\{\pi\})/2$. Such a restriction of $\mu_X$ to the half-open interval $(-\pi, \pi]$ is actually intended since this guarantees uniqueness of the spectral measure. Furthermore, if $\mu_X$ had a positive point mass in $-\pi$, then a corresponding spectral distribution function had to fulfill $F_X(-\pi) - F_X(-\pi - 0) = \mu_X\big(\{-\pi\}\big) > 0$ which is unintended since this requires a definition of $F_X$ beyond the interval $[-\pi, \pi]$.

## Exercise

**Ex. 2.1.1** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$ and that $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is the unique stationary solution to

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \qquad \forall t \in \mathbb{Z},$$

where $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$.

Compute the spectral density of the process $(X_t)_{t\in\mathbb{Z}}$.

In the following we generalize these results to stationary processes with an autocovariance function that is not necessarily absolutely summable. To this end we derive a few useful results from probability theory.

**Theorem 2.1.2.** *(Helly's selection theorem)*
*Let, for any $K < \infty$,*

$$V := \Big\{F\colon \mathbb{R} \to [0, K],\ \ F \text{ is monotonically nondecreasing and right-continuous}\Big\}.$$

*If $\big(F_n\big)_{n\in\mathbb{N}}$ is any sequence of functions from $V$, then there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ and a function $F \in V$ such that*

$$F_{n_k}(x) \underset{k\to\infty}{\longrightarrow} F(x) \qquad \text{for all continuity points } x \text{ of } F.$$

*Proof.* The proof of this result is split up into four steps.

(i) (*Identification of the limit*)

Since $\mathbb{Q}$ is countable we can enumerate the rational numbers by $r_1, r_2, \ldots$. Since $\big(F_n(r_1)\big)_{n\in\mathbb{N}}$ is a bounded sequence of real numbers there exists a subsequence $\big(n_k^{(1)}\big)_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that

$$F_{n_k^{(1)}}(r_1) \underset{k\to\infty}{\longrightarrow} G(r_1),$$

for some $G(r_1) \in [0, K]$. For the same reason, there exists a further subsequence $\big(n_k^{(2)}\big)_{k\in\mathbb{N}}$ of $\big(n_k^{(1)}\big)_{k\in\mathbb{N}}$ such that

$$F_{n_k^{(2)}}(r_2) \underset{k\to\infty}{\longrightarrow} G(r_2),$$

for some $G(r_2) \in [0, K]$. We proceed in the same way. In the $m$th step we can choose a subsequence $\big(n_k^{(m)}\big)_{k\in\mathbb{N}}$ of $\big(n_k^{(m-1)}\big)_{k\in\mathbb{N}}$ such that

$$F_{n_k^{(m)}}(r_m) \underset{k\to\infty}{\longrightarrow} G(r_m),$$

for some $G(r_m) \in [0, K]$. We take the "**diagonal sequence**" $\big(n_k\big)_{k\in\mathbb{N}}$, where $n_k = n_k^{(k)}$ $\forall k \in \mathbb{N}$. Then

$$F_{n_k}(r) \underset{k\to\infty}{\longrightarrow} G(r) \qquad \forall r \in \mathbb{Q}.$$

The function $G\colon \mathbb{Q} \to \mathbb{R}$ is monotonically nondecreasing and it holds $G(r) \in [0, K]$ for all $r \in \mathbb{Q}$.

(ii) (*Extension to a function on* $\mathbb{R}$)

We define

$$F(x) := \inf\big\{G(r)\colon\ r \in \mathbb{Q},\ r > x\big\} \qquad \forall x \in \mathbb{R}.$$

(We will see in step (iii) below that "$r > x$" rather than "$r \geq x$" is really important since this ensures that $F$ is right-continuous.)

(iii) (*Properties of $F$*)

It follows from the definition that $F$ is monotonically nondecreasing and that $F(x) \in [0,K]\ \forall x \in \mathbb{R}$.

As for right-continuity, let $(x_n)_{n\in\mathbb{N}}$ be a sequence of monotonically nonincreasing real numbers such that $x_n \searrow x$. In order to show that $F(x_n) \underset{n\to\infty}{\longrightarrow} F(x)$ actually holds, we choose an "**accompanying sequence**" $(s_n)_{n\in\mathbb{N}}$, $s_n \in \mathbb{Q}$ such that $x \leq x_n < s_n$ and $s_n \underset{n\to\infty}{\longrightarrow} x$. Then

$$F(x) \;\leq\; F(x_n) \leq G(s_n).$$

But since $G(s_n) \underset{n\to\infty}{\longrightarrow} F(x)$ (At this point we see that the strict inequality sign in $F(x) = \inf\left\{G(r)\colon\ r \in \mathbb{Q},\ r{>}x\right\}$ is necessary.) we conclude that

$$F(x_n) \underset{n\to\infty}{\longrightarrow} F(x).$$

Hence, $F \in V$.

(iv) (*Convergence of $F_{n_k}$ to $F$*)

Let $x$ be a continuity point of $F$. We have to show that

$$F_{n_k}(x) \underset{k\to\infty}{\longrightarrow} F(x). \tag{2.1.3}$$

To do this, we can only use the fact that $F_{n_k}(r) \underset{k\to\infty}{\longrightarrow} G(r)\ \forall r \in \mathbb{Q}$. Let $\epsilon > 0$ be arbitrary. According to the **definition** of $F(x)$, there exists some $\overline{r} \in \mathbb{Q}$ such that $x < \overline{r}$ and

$$G(\overline{r}) \;\leq\; F(x) \,+\, \epsilon.$$

This implies

$$\limsup_{k\to\infty} F_{n_k}(x) \;\leq\; \lim_{k\to\infty} F_{n_k}(\overline{r}) \;=\; G(\overline{r}) \;\leq\; F(x) \,+\, \epsilon. \tag{2.1.4}$$

On the other hand, by **continuity** of $F$ in $x$, there exists some $\underline{x} < x$ such that

$$F(\underline{x}) \;\geq\; F(x) \,-\, \epsilon.$$

Let $\underline{r} \in \mathbb{Q} \cap (\underline{x}, x]$. Then

$$\liminf_{k\to\infty} F_{n_k}(x) \;\geq\; \lim_{k\to\infty} F_{n_k}(\underline{r}) \;=\; G(\underline{r}) \;\geq\; F(\underline{x}) \;\geq\; F(x) \,-\, \epsilon. \tag{2.1.5}$$

From (2.1.4) and (2.1.5) we obtain that (2.1.3) holds true. □

**Helly's selection theorem in probability theory**

When applied to probability distribution functions or probability measures, Helly's selection theorem plays an important role in probability theory. Before we state and prove corresponding results we take a brief look at a typical example which shows what actually could happen.

For $n \in \mathbb{N}$, let $\mu_n = N(0, \sigma_n^2)$ be normal distributions with mean 0 and variance $\sigma_n^2 > 0$ and let $F_n$ be the corresponding distributions functions. Then $F_n(x) = \Phi(x/\sigma_n)\ \forall x \in \mathbb{R}$, where $\Phi$ denotes the distribution function of a standard normal distribution. If $\sigma_n \underset{n\to\infty}{\longrightarrow} \sigma$ for some $\sigma \in [0, \infty]$, then the full sequence $\big(F_n(x)\big)_{n\in\mathbb{N}}$ converges for all $x \in \mathbb{R}$, i.e., we do not have to select an appropriate subsequence $\big(n_k\big)_{k\in\mathbb{N}}$ as in Theorem 2.1.2. We can distinguish between the following cases:

a) If $\sigma \in (0, \infty)$, then

$$F_n(x) \underset{n\to\infty}{\longrightarrow} \Phi(x/\sigma) \qquad \forall x \in \mathbb{R},$$

which is the distribution function of a normal distribution with mean 0 and variance $\sigma^2$.

b) If $\sigma = 0$, then

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F_0(x) := \begin{cases} 0, & \text{if } x < 0, \\ 1/2, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases}$$

Here we have to be careful since $F_0$ is **not** a distribution function since it is not right-continuous in the point $x = 0$. On the other hand, it also holds that

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) := \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0 \end{cases} \qquad \text{for all continuity points } x \text{ of } F.$$

$F$ is the distribution function of a Dirac measure in the point 0.

c) If $\sigma = \infty$, then

$$F_n(x) \underset{n\to\infty}{\longrightarrow} \Phi(0) = 1/2 =: F(x) \qquad \forall x \in \mathbb{R},$$

which is a distribution function corresponding to the zero measure on $(\mathbb{R}, \mathcal{B})$. In this case, the limit of the probability measures $\mu_n$ is still a measure but **not** a probability measure.

What we have seen in this example can be described by the notions of **weak** and **vague convergence of probability measures**. Here is a formal definition of these two modes of convergence:

**Definition.** Let $\big(\mu_n\big)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\big(\mathbb{R}, \mathcal{B}\big)$. Then $\big(\mu_n\big)_{n\in\mathbb{N}}$ is said to

(i) **converge weakly** to a **probability measure** $\mu$ on $\big(\mathbb{R}, \mathcal{B}\big)$ ($\mu_n \Longrightarrow \mu$ or $\mu_n \overset{w}{\longrightarrow} \mu$) if

$$\int_{\mathbb{R}} f(x)\, d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\, d\mu(x) \qquad \text{for all continuous and bounded functions } f\colon\ \mathbb{R} \to \mathbb{R}.$$

(ii) **converge vaguely** to a **measure** $\mu$ on $\big(\mathbb{R}, \mathcal{B}\big)$ ($\mu_n \overset{v}{\longrightarrow} \mu$) if

$$\int_{\mathbb{R}} f(x)\, d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\, d\mu(x) \qquad \text{for all continuous, compactly supported functions } f\colon\ \mathbb{R} \to \mathbb{R}.$$

The following lemma provides the relation between convergence of distribution functions and weak or vague convergence of the corresponding measures.

**Lemma 2.1.3.** *Let $\big(F_n\big)_{n\in\mathbb{N}}$ be a sequence of probability distribution functions on $\mathbb{R}$ and $F$ be a distribution function on $\mathbb{R}$ ($F$ is monotonically nondecreasing and right-continuous, but $F(x) \longrightarrow_{x\to-\infty} 0$ and $F(x) \longrightarrow_{x\to\infty} 1$ are* **not** *necessarily fulfilled). Suppose that*

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) \qquad \text{for all continuity points } x \text{ of } F$$

*and let $\big(\mu_n\big)_{n\in\mathbb{N}}$ and $\mu$ be the corresponding measures on $\big(\mathbb{R},\mathcal{B}\big)$. Then*

*(i)* $\mu(\mathbb{R}) \leq 1$ *and* $\mu_n \overset{v}{\longrightarrow} \mu$.

*(ii)* *If* $\mu(\mathbb{R}) = 1$, *then* $\mu_n \Longrightarrow \mu$.

*(iii)* *If additionally* $\forall \epsilon > 0\ \exists K_\epsilon < \infty$ *such that*

$$F_n(K_\epsilon) - F_n(-K_\epsilon) = \mu_n\big((-K_\epsilon, K_\epsilon]\big) \geq 1 - \epsilon \qquad \forall n \in \mathbb{N}$$

*(i.e., the sequence $\big(\mu_n\big)_{n\in\mathbb{N}}$ is* **tight***), then*

$$\mu(\mathbb{R}) = 1 \qquad \text{and} \qquad \mu_n \Longrightarrow \mu.$$

*Proof.* Since $F$ is a monotonically nondecreasing function it follows that it has at most a countable number of discontinuity points. We denote the set of these points by $D_F$.

(i) Let $(x_m)_{m\in\mathbb{N}}$ be a sequence of monotonically increasing nonnegative numbers such that $x_m, x_{-m} \notin D_F$ and $x_m \longrightarrow_{m\to\infty} \infty$. Since the measure $\mu$ is continuous from below we obtain that

$$\begin{aligned}
\mu(\mathbb{R}) &= \lim_{m\to\infty} \mu\big((-x_m, x_m]\big) \\
&= \lim_{m\to\infty} F(x_m) - F(-x_m) \\
&= \lim_{m\to\infty} \lim_{n\to\infty} \underbrace{F_n(x_m) - F_n(-x_m)}_{\leq 1} \leq 1.
\end{aligned}$$

As for vague convergence, we have to show that

$$\int_{\mathbb{R}} f(x)\, d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\, d\mu(x), \tag{2.1.6}$$

for all **continuous** and **compactly supported functions** $f\colon \mathbb{R} \to \mathbb{R}$. Let $f$ be such a function, i.e., $f(x) = 0$ for all $x \notin (-K, K]$, for some $K < \infty$ such that $K, -K \notin D_F$. Let $\epsilon > 0$ be arbitrary. Since $f$ is continuous there exist points $x_0, \ldots, x_M \notin D_F$ such that $-K = x_0 < \ldots < x_M = K$ and

$$\big|f(x) - f(x_k)\big| \leq \epsilon \qquad \forall x \in (x_{k-1}, x_k].$$

With this choice of a partition of $[-K, K]$, the integrals in (2.1.6) are well approximated by the corresponding right Riemann sums:

$$\begin{aligned}
&\Big| \int_{\mathbb{R}} f(x)\, d\mu(x) \quad - \quad \sum_{k=1}^{M} f(x_k)\, \mu\big((x_{k-1}, x_k]\big) \Big| \\
&= \Big| \int_{(-K,K]} f(x)\, d\mu(x) \quad - \quad \sum_{k=1}^{M} f(x_k) \int_{(x_{k-1},x_k]} d\mu(x) \Big| \\
&\le \sum_{k=1}^{M} \int_{(x_{k-1},x_k]} \underbrace{\big| f(x) - f(x_k) \big|}_{\le \epsilon}\, d\mu(x) \\
&\le \epsilon \sum_{k=1}^{M} \mu\big((x_{k-1}, x_k]\big) \le \epsilon
\end{aligned} \tag{2.1.7}$$

and, analogously,

$$\begin{aligned}
&\Big| \int_{\mathbb{R}} f(x)\, d\mu_n(x) \quad - \quad \sum_{k=1}^{M} f(x_k)\, \mu_n\big((x_{k-1}, x_k]\big) \Big| \\
&\le \sum_{k=1}^{M} \int_{(x_{k-1},x_k]} \underbrace{\big| f(x) - f(x_k) \big|}_{\le \epsilon}\, d\mu_n(x) \\
&\le \epsilon \sum_{k=1}^{M} \mu_n\big((x_{k-1}, x_k]\big) = \epsilon \big(F_n(K) - F_n(-K)\big) \le \epsilon \qquad \forall n \in \mathbb{N}.
\end{aligned} \tag{2.1.8}$$

Finally, from $F_n(x_k) \underset{n\to\infty}{\longrightarrow} F(x_k)$ for $k = 0, \ldots, M$ we obtain

$$\mu_n\big((x_{k-1}, x_k]\big) = F_n(x_k) - F_n(x_{k-1}) \underset{n\to\infty}{\longrightarrow} F(x_k) - F(x_{k-1}) = \mu\big((x_{k-1}, x_k]\big).$$

Therefore,

$$\begin{aligned}
&\Big| \sum_{k=1}^{M} f(x_k)\mu_n\big((x_{k-1}, x_k]\big) - \sum_{k=1}^{M} f(x_k)\mu\big((x_{k-1}, x_k]\big) \Big| \\
&\le \sup\big\{|f(x)| \colon x \in [-K, K]\big\} \sum_{k=1}^{M} \big|\mu_n\big((x_{k-1}, x_k]\big) - \mu\big((x_{k-1}, x_k]\big|\\
&= \sup\big\{|f(x)| \colon x \in [-K, K]\big\} \sum_{k=1}^{M} \big|\big(F_n(x_k) - F_n(x_{k-1})\big) - \big(F(x_k) - F(x_{k-1})\big)\big| \\
&\le \epsilon \qquad \forall n \ge N,
\end{aligned} \tag{2.1.9}$$

if $N$ is sufficiently large. It follows from (2.1.7) to (2.1.9)

$$\Big| \int_{\mathbb{R}} f(x)\, d\mu_n(x) \quad - \quad \int_{\mathbb{R}} f(x)\, d\mu(x) \Big| \le 3\epsilon \qquad \forall n \ge N,$$

which implies that (2.1.6) holds true, i.e., $\mu_n \overset{v}{\longrightarrow} \mu$.

(ii) If additionally $\mu(\mathbb{R}) = 1$, then we can even show weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to $\mu$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be an any **bounded** and **continuous function** and let $\epsilon > 0$ be arbitrary. Furthermore, let $K < \infty$ be such that $-K, K \notin D_F$ and $\mu\big((-K, K]\big) \geq 1 - \epsilon$. Since

$$\begin{aligned} \mu_n\big((-\infty, -K] \cup (K, \infty)\big) &= F_n(-K) + \big(1 - F_n(K)\big) \\ &\underset{n\to\infty}{\longrightarrow} F(-K) + \big(1 - F(K)\big) = 1 - \mu\big((-K, K]\big) \leq \epsilon \end{aligned}$$

we obtain that

$$\mu_n\big((-\infty, -K] \cup (K, \infty)\big) \leq 2\epsilon \qquad \forall n \geq N_1,$$

if $N_1$ is sufficiently large. This implies

$$\begin{aligned} &\int_{(-\infty,-K]\cup(K,\infty)} f(x)\, d\mu_n(x) + \int_{(-\infty,-K]\cup(K,\infty)} f(x)\, d\mu(x) \\ &\quad\leq \sup\big\{|f(x)|:\ x \in \mathbb{R}\big\} \Big(\mu_n\big((-\infty, -K] \cup (K, \infty)\big) + \mu\big((-\infty, -K] \cup (K, \infty)\big)\Big) \\ &\quad\leq 3\,\epsilon\, \sup\big\{|f(x)|:\ x \in \mathbb{R}\big\} \qquad \forall n \geq N_1. \end{aligned} \tag{2.1.10}$$

Moreover, we can show as above that

$$\Big| \int_{(-K,K]} f(x)\, d\mu_n(x) - \int_{(-K,K]} f(x)\, d\mu(x) \Big| \leq \epsilon \qquad \forall n \geq N_2, \tag{2.1.11}$$

if $N_2$ is sufficiently large. We conclude from (2.1.10) and (2.1.11) that

$$\int_{\mathbb{R}} f(x)\, d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\, d\mu(x),$$

i.e. $\mu_n \Longrightarrow \mu$ holds true.

(iii) Let $\epsilon > 0$ be arbitrary. We choose $K < \infty$ such that $-K, K \notin D_F$ and $\mu_n\big((-K, K]\big) \geq 1 - \epsilon \ \forall n \in \mathbb{N}$. Since

$$\begin{aligned} \mu_n\big((-K, K]\big) &= F_n(K) - F_n(-K) \\ &\underset{n\to\infty}{\longrightarrow} F(K) - F(-K) = \mu\big((-K, K]\big) \end{aligned}$$

we obtain that

$$\mu\big((-K, K]\big) \geq 1 - \epsilon.$$

This implies that $\mu(\mathbb{R}) = 1$ and we obtain from (ii) that $\mu_n \Longrightarrow \mu$.

□

The following theorem states the remarkable fact that the autocovariances of any weakly stationary process can be expressed by integrals with a finite measure $\mu$ as the integrator. This does not require absolute summability of the autocovariances.

**Theorem 2.1.4.** *(Herglotz' theorem)*
*A function* $\gamma\colon\ \mathbb{Z} \longrightarrow \mathbb{R}$ *is the autocovariance function of a real-valued stationary process*
*if and only if*
*there exists a finite measure* $\mu$ *on* $\big((-\pi,\pi], \mathcal{B}|_{(-\pi,\pi]}\big)$ *such that* $\mu(B) = \mu(-B)$ *for all* $B \in \mathcal{B}|_{(-\pi,\pi]}$, $B \subseteq (-\pi,\pi)$ *and*

$$\gamma(k) \;=\; \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu(\lambda) \qquad \forall k \in \mathbb{Z}.$$

*Proof.* First of all, recall that it follows from Theorem **??** that $\gamma\colon\ \mathbb{Z} \longrightarrow \mathbb{R}$ is the autocovariance function of a weakly stationary real-valued process
if and only if
$\gamma$ is an even, real-valued and non-negative definite function, i.e.

$$\sum_{j,k=1}^{n} a_j \gamma(t_j - t_k) a_k \;\geq\; 0 \qquad \forall t_1,\ldots,t_n \in \mathbb{Z},\ \forall a_1,\ldots,a_n \in \mathbb{R},\ \forall n \in \mathbb{N}. \tag{2.1.12}$$

($\Longleftarrow$)

Suppose that $\mu$ is a finite measure on $\mathcal{B}\big|_{(-\pi,\pi]}$ such that $\mu(B) = \mu(-B)$ for all $B \in \mathcal{B}\big|_{(-\pi,\pi]}$, $B \subseteq (-\pi,\pi)$ and

$$\gamma(k) \;=\; \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu(\lambda) \qquad \forall k \in \mathbb{Z}.$$

It follows from the symmetry of $\mu$ that $\int_{(-\pi,\pi)} \sin(k\lambda)\, d\mu(\lambda) = 0$ holds for all $k \in \mathbb{Z}$, which implies that

$$\gamma(k) \;=\; \int_{(-\pi,\pi)} \cos(k\lambda)\, d\mu(\lambda) \;+\; i \underbrace{\int_{(-\pi,\pi)} \sin(k\lambda)\, d\mu(\lambda)}_{=\,0} \;+\; \underbrace{e^{ik\pi}}_{\in\,\{1,-1\}}\ \mu\big(\{\pi\}\big).$$

Hence, $\gamma$ is a real-valued function. Furthermore, is also follows from the symmetry of $\mu$ that $\int_{(-\pi,\pi)} \cos(k\lambda)\, d\mu(\lambda) = \int_{(-\pi,\pi)} \cos(-k\lambda)\, d\mu(\lambda)$, which implies that

$$\begin{aligned} \gamma(k) \;&=\; \int_{(-\pi,\pi)} \cos(k\lambda)\, d\mu(\lambda) \;+\; \underbrace{e^{ik\pi}}_{=\,e^{-ik\pi}}\ \mu\big(\{\pi\}\big)\mu\big(\{\pi\}\big) \\ &=\; \int_{(-\pi,\pi)} \cos(-k\lambda)\, d\mu(\lambda) \;+\; e^{-ik\pi}\mu\big(\{\pi\}\big) \;=\; \gamma(-k) \qquad \forall k \in \mathbb{Z}, \end{aligned}$$

i.e., $\gamma$ is an even function. Finally, we have, for arbitrary $n \in \mathbb{N}$, $t_1,\ldots,t_n \in \mathbb{Z}$ and $a_1,\ldots,a_n \in \mathbb{R}$,

$$\begin{aligned} \sum_{j,k=1}^{n} a_j \gamma(t_j - t_k) a_k \;&=\; \int_{(-\pi,\pi]} \sum_{j,k=1}^{n} a_j a_k e^{i(t_j - t_k)\lambda}\, d\mu(\lambda) \\ &=\; \int_{(-\pi,\pi]} \Big| \sum_{j=1}^{n} a_j e^{it_j\lambda} \Big|^2 d\mu(\lambda) \;\geq\; 0. \end{aligned}$$

Hence, $\gamma$ is the autocovariance function of some real-valued stationary process.

($\Longrightarrow$)

Suppose now that $\gamma$ is an even and real-valued function satisfying (2.1.12). We have to show that there exists a finite measure $\mu$ on $\mathcal{B}\big|_{(-\pi,\pi]}$ such that $\mu(B) = \mu(-B)$ for all $B \in \mathcal{B}\big|_{(-\pi,\pi]}$, $B \subseteq (-\pi, \pi)$ and

$$\gamma(k) = \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu(\lambda) \qquad \forall k \in \mathbb{Z}.$$

If $\gamma(0) = 0$, then all other autocovariances vanish as well and the zero measure ($\mu(B) = 0 \;\; \forall B \in \mathcal{B}\big|_{(-\pi,\pi]}$) solves the system of equations.

Let $\gamma(0) > 0$. Since the existence of a spectral density is not guaranteed we start with some sort of "regularized" spectral density. Let

$$\begin{aligned} f_n(\lambda) \;&:=\; \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \left(1 - \frac{|l|}{n}\right)_+ e^{-il\lambda}\gamma(k) \;=\; \frac{1}{2\pi} \sum_{l=-(n-1)}^{n-1} \left(1 - \frac{|l|}{n}\right) \cos(l\lambda)\gamma(k) \\ &=\; \frac{1}{2\pi n} \sum_{r,s=1}^{n} \underbrace{\cos\big((r-s)\lambda\big)}_{=\cos(r\lambda)\cos(s\lambda)+\sin(r\lambda)\sin(s\lambda)} \gamma(r-s) \\ &=\; \frac{1}{2\pi n} \sum_{r,s=1}^{n} \cos(r\lambda)\gamma(r-s)\cos(s\lambda) + \frac{1}{2\pi n} \sum_{r,s=1}^{n} \sin(r\lambda)\gamma(r-s)\sin(s\lambda). \end{aligned}$$

$f_n$ is obviously a real-valued function which is by (2.1.12) also non-negative definite. Let $\mu_n$ be a measure on $\mathcal{B}$ such that $\mu_n((a,b]) = \int_a^b f_n(\lambda)\mathbb{1}_{[-\pi,\pi]}(\lambda)\, d\lambda$ for all $a, b \in \mathbb{R}$, $a \leq b$. Then

$$\begin{aligned} \int_{\mathbb{R}} e^{ik\lambda}\, d\mu_n(\lambda) \;&=\; \int_{-\pi}^{\pi} e^{ik\lambda} f_n(\lambda)\, d\lambda \\ &=\; \int_{-\pi}^{\pi} e^{ik\lambda}\, \frac{1}{2\pi} \sum_{l=-(n-1)}^{n-1} \left(1 - \frac{|l|}{n}\right)\gamma(l)e^{-il\lambda}\, d\lambda \\ &=\; \sum_{l=-(n-1)}^{n-1} \left(1 - \frac{|l|}{n}\right)\gamma(l)\, \underbrace{\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(k-l)\lambda}\, d\lambda}_{=\delta_{k,l}} \\ &=\; \left(1 - \frac{|k|}{n}\right)_+ \gamma(k) \end{aligned}$$

holds for all $k \in \mathbb{Z}$. We have in particular that $\mu_n(\mathbb{R}) = \mu_n([-\pi,\pi]) = \gamma(0)$ for all $n \in \mathbb{N}$. Therefore, $\big(\mu_n/\gamma(0)\big)_{n\in\mathbb{N}}$ is a **tight sequence of probability measures** on $(\mathbb{R}, \mathcal{B})$ and it follows from (iii) of Lemma 2.1.3 that there exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of $\mathbb{N}$ such that $(\mu_{n_j}/\gamma(0))_{j\in\mathbb{N}}$ **converges weakly** to a probability measure $\widetilde{\mu}$ on $(\mathbb{R}, \mathcal{B})$, where $\widetilde{\mu}\big([-\pi,\pi]^c\big) = \liminf_{j\to\infty} \mu_{n_j}\big([-\pi,\pi]^c\big)/\gamma(0) = 0$. Therefore,

$$\frac{1}{\gamma(0)} \int_{\mathbb{R}} e^{ik\lambda}\, d\mu_{n_j}(\lambda) \underset{j\to\infty}{\longrightarrow} \int_{\mathbb{R}} e^{ik\lambda}\, d\widetilde{\mu}(\lambda) = \int_{[-\pi,\pi]} e^{ik\lambda}\, d\widetilde{\mu}(\lambda) \qquad \forall k \in \mathbb{Z},$$

It is easy to see that the measure $\widetilde{\mu}$ is symmetric about zero. Let $\widetilde{\mu}^-$ be such that $\widetilde{\mu}^-(B) = \widetilde{\mu}(-B)$ for all $B \in \mathcal{B}$. Since $\mu_{n_j}/\gamma(0) \Longrightarrow \widetilde{\mu}$ we obtain that

$$\mu_{n_j}\big((a,b]\big)/\gamma(0) \underset{j\to\infty}{\longrightarrow} \widetilde{\mu}\big((a,b]\big)$$

holds for all $a, b$ such that $a < b$ and $\widetilde{\mu}(\{a,b\}) = 0$. For the same reason we obtain that

$$\mu_{n_j}\big([-b,-a)\big)/\gamma(0) \underset{j\to\infty}{\longrightarrow} \widetilde{\mu}\big([-b,-a)\big) \;=\; \widetilde{\mu}^-\big((a,b]\big)$$

holds for all $a, b$ such that $a < b$ and $\widetilde{\mu}(\{-a,-b\}) = 0$. Since $\mu_{n_j}\big((a,b]\big) = \mu_{n_j}\big([-b,-a)\big)$ for all $a < b$ and since the probability measure $\widetilde{\mu}$ has a positive mass in at most a countable number of points we see that

$$\widetilde{\mu}\big((a,b]\big) \;=\; \widetilde{\mu}^-\big((a,b]\big)$$

holds true for all $a < b$, with the possible exception of a countable number of points. In other words, the distribution functions of these two measures coincide at almost all points which means that $\widetilde{\mu} = \widetilde{\mu}^-$.

Let $\mu_\infty(B) = \gamma(0)\widetilde{\mu}(B)\ \forall B \in \mathcal{B}$. It could still be the case that the measure $\mu_\infty$ has some mass in the point $-\pi$. In order to end up with an appropriate measure on $\mathcal{B}\big|_{(-\pi,\pi]}$ we shift this mass to the point $\pi$, i.e. we define

$$\mu(B) \;=\; \mu_\infty(B) \,+\, \mathbb{1}_B(\pi)\mu_\infty(\{-\pi\}) \qquad \forall B \in \mathcal{B}\big|_{(-\pi,\pi]}.$$

Then $\mu(B) = \mu_\infty(B) = \mu_\infty(-B) = \mu(-B)$ for all $B \in \mathcal{B}\big|_{(-\pi,\pi]}$, $B \subseteq (-\pi,\pi)$ and

$$\int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu(\lambda) \;=\; \int_{[-\pi,\pi]} e^{ik\lambda}\, d\mu_\infty(\lambda) \;=\; \lim_{j\to\infty}\int_{[-\pi,\pi]} e^{ik\lambda}\, d\mu_{n_j}(\lambda) \;=\; \gamma(k),$$

which completes the proof. □

In the following we show that both the spectral distribution function $F$ and the spectral measure $\mu$ are uniquely defined by the autocovariances of the corresponding stationary process.

**Lemma 2.1.5.** *Let $\gamma$ be the autocovariance function of a real-valued and weakly stationary process and let $\mu_1$ and $\mu_2$ be measures on $\big((-\pi,\pi],\mathcal{B}\big|_{(-\pi,\pi]}\big)$ such that*

$$\gamma(k) \;=\; \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu_j(\lambda) \qquad \forall k \in \mathbb{Z}, j = 1,2.$$

*Then*

$$\mu_1 \;=\; \mu_2.$$

*Proof.* Since

$$\mu_j\big((-\pi,\pi]\big) \;=\; \int_{(-\pi,\pi]} e^{i0\lambda}\, d\mu_j(\lambda) \;=\; \gamma(0) \;<\; \infty$$

we see that $\mu_1$ and $\mu_2$ are finite measures. According to the uniqueness theorem of measure theory it suffices to show that

$$\mu_1\big((a,b]\big) \;=\; \mu_2\big((a,b]\big) \qquad \forall a,b \;\text{ such that }\; -\pi \le a < b \le \pi.$$

Let $a, b$ be arbitrary such that $-\pi \le a < b \le \pi$. We show that

$$\int_{(-\pi,\pi]} \mathbb{1}_{(a,b]}(x)\, d\mu_1(x) \;=\; \int_{(-\pi,\pi]} \mathbb{1}_{(a,b]}(x)\, d\mu_2(x).$$

To this end we approximate the indicator function $\mathbb{1}_{(a,b]}$ by a sequence of trigonometric functions. First of all, there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of continuous functions $f_n\colon [-\pi,\pi] \to [0,1]$ such that $f_n(-\pi) = f_n(\pi)$ and $f_n(\lambda) \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{(a,b]}(\lambda) \quad \forall \lambda \in (-\pi,\pi]$. Then, by dominated convergence,

$$\int_{(-\pi,\pi]} \mathbb{1}_{(a,b]}(x)\, d\mu_j(x) \;=\; \int_{(-\pi,\pi]} \lim_{n\to\infty} f_n(x)\, d\mu_j(x) \;=\; \lim_{n\to\infty} \int_{(-\pi,\pi]} f_n(x)\, d\mu_j(x), \qquad j = 1,2.$$

Therefore, it suffices to show that

$$\int_{(-\pi,\pi]} f(x)\, d\mu_1(x) \;=\; \int_{(-\pi,\pi]} f(x)\, d\mu_2(x)$$

holds for all continuous functions $f\colon [-\pi,\pi] \to [0,1]$ such that $f(-\pi) = f(\pi)$. Let $f$ be such a function. By Fejér's theorem there exists a sequence of trigonometric functions $\big(T_n f\big)_{n\in\mathbb{N}}$ (i.e., $T_n f(\lambda) = a_0 + \sum_{k=1}^n a_k \cos(k\lambda) + b_k \sin(k\lambda)$, for some $a_0,\dots,a_n, b_1,\dots,b_n \in \mathbb{R}$) such that

$$\delta_n \;:=\; \sup_{\lambda\in[-\pi,\pi]} \big\{|T_n f(\lambda) - f(\lambda)|\big\} \underset{n\to\infty}{\longrightarrow} 0.$$

From

$$\int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu_1(\lambda) \;=\; \int_{(-\pi,\pi]} e^{ik\lambda}\, d\mu_2(\lambda) \;=\; \gamma(k) \;\in\; \mathbb{R} \qquad \forall k \in \mathbb{Z}$$

we obtain

$$\int_{(-\pi,\pi]} \cos(k\lambda)\, d\mu_1(\lambda) \;=\; \int_{(-\pi,\pi]} \cos(k\lambda)\, d\mu_2(\lambda) \qquad \forall k \in \mathbb{N}_0$$

and

$$\int_{(-\pi,\pi]} \sin(k\lambda)\, d\mu_1(\lambda) \;=\; \int_{(-\pi,\pi]} \sin(k\lambda)\, d\mu_2(\lambda) \;=\; 0 \qquad \forall k \in \mathbb{N}.$$

Therefore,

$$\int_{(-\pi,\pi]} T_n f(\lambda)\, d\mu_1(\lambda) \;=\; \int_{(-\pi,\pi]} T_n f(\lambda)\, d\mu_2(\lambda) \qquad \forall n \in \mathbb{N}.$$

Finally, since

$$\int_{(-\pi,\pi]} T_n f(\lambda)\, d\mu_j(\lambda) \;-\; \int_{(-\pi,\pi]} f(\lambda)\, d\mu_j(\lambda) \;\le\; \delta_n\, \mu_j\big((-\pi,\pi]\big) \underset{n\to\infty}{\longrightarrow} 0, \qquad j = 1,2,$$

we obtain the assertion. □

## Exercise

**Ex. 2.1.2** Suppose that $Y$ and $Z$ are uncorrelated random variables with $EY = EZ = 0$ and $EY^2 = EZ^2 = 1$. For $t \in \mathbb{Z}$, let $X_t = Y\cos(\theta t) + Z\sin(\theta t)$, where $\theta \in [0,\pi]$.

(i) Compute the autocovariance function $\gamma_X$ of the process $(X_t)_{t\in\mathbb{Z}}$.
*Hint: Use the trigonometric identity* $\cos(a)\cos(b) + \sin(a)\sin(b) = \cos(a-b)$ $\forall a, b \in \mathbb{R}$.

(ii) Determine the spectral measure of $(X_t)_{t\in\mathbb{Z}}$. Does there exist a spectral density?

## 2.2 Estimation in the spectral domain

Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a weakly stationary and real-valued process with autocovariance function $\gamma$, where $\sum_{k=0}^{\infty} |\gamma(k)| < \infty$. Then the process $\mathbf{X}$ has a spectral density $f_X$, which is defined as

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k)\cos(k\lambda) \qquad \forall \lambda \in [-\pi, \pi].$$

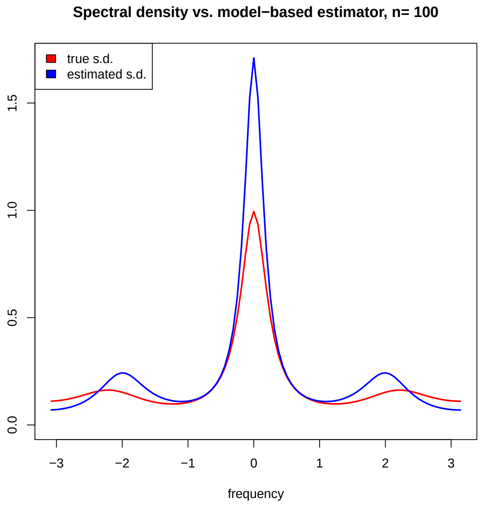
Suppose that realizations of the random variables $X_1, \ldots, X_n$ are available. If we are willing to assume that some model described by a finite-dimensional parameter is adequate, then we can first estimate this parameter and obtain an estimator of the spectral density by plugging this estimator into the formula for the spectral density of the corresponding class of processes. For example, suppose that $(X_t)_{t\in\mathbb{Z}}$ is an autoregressive process of order $p$, where $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p \neq 0 \;\; \forall z \in \mathbb{C}, \, |z| \le 1$ and that the sequence of innovations is given by $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$. Then the spectral density is defined as

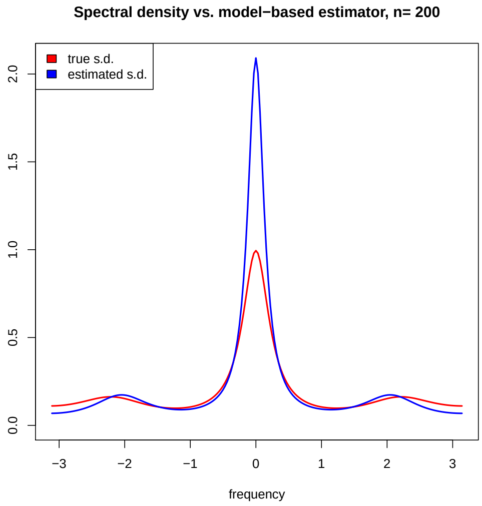
$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi \left|1 - \alpha_1 e^{-i\lambda} - \cdots - \alpha_p e^{-ip\lambda}\right|^2} \qquad \forall \lambda \in [-\pi, \pi].$$

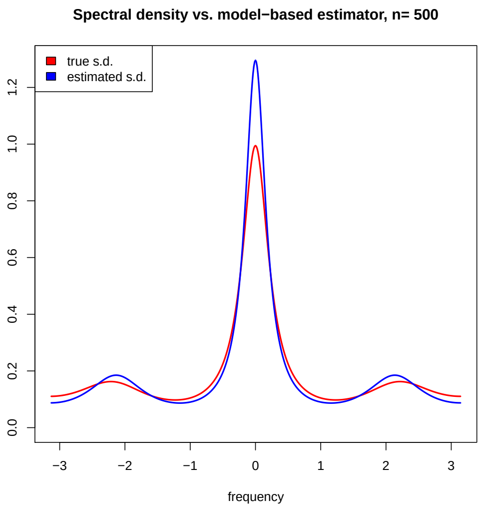
We know that the Yule-Walker estimators $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ and $\widehat{\sigma}_\varepsilon^2$ of the respective parameters $\alpha_1, \ldots, \alpha_p$ and $\sigma_\varepsilon^2$ are consistent, i.e., they converge in probability to their theoretical counterparts as the sample size $n$ tends to infinity. Since $f_X(\lambda)$ is a continuous function of these parameters, a consistent estimator of the spectral density is given by

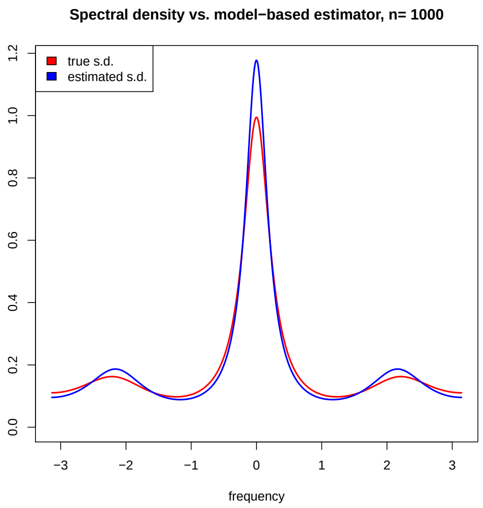
$$\widehat{f}_X(\lambda) = \frac{\widehat{\sigma}_\varepsilon^2}{2\pi \left|1 - \widehat{\alpha}_1 e^{-i\lambda} - \cdots - \widehat{\alpha}_p e^{-ip\lambda}\right|^2} \qquad \forall \lambda \in [-\pi, \pi].$$

The pictures below show the true spectral density $f_X$ of an AR($p$) process with parameters $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$ and independent innovations $\varepsilon_t \sim N(0,1)$ (red lines), together with one realization of the model-based estimator $\widehat{f}_X$ obtained from samples of size $n$ =100, 200, 500 and 1,000, respectively, (blue lines).






If a parametric model with a known relation between the parameters and the spectral density is not available or if consistent estimators of the model parameters are difficult to obtain, then we could use so-called nonparametric approaches to estimate the spectral density or the spectral distribution function. We consider such methods in the following.

Note that $\gamma(k)$ can be estimated by

$$\widehat{\gamma}_n(k) \;=\; \begin{cases} n^{-1}\sum_{t=1}^{n-|k|}\big(X_{t+|k|}-\bar{X}_n\big)\big(X_t-\bar{X}_n\big) & \text{if } |k|\le n-1, \\ 0 & \text{if } |k|\ge n, \end{cases}$$

where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$ is the sample mean. Under the additional assumption of $EX_t = 0$, we could estimate $\gamma(k)$ also by

$$\widetilde{\gamma}_n(k) \;=\; \begin{cases} n^{-1}\sum_{t=1}^{n-|k|} X_{t+|k|}X_t & \text{if } |k|\le n-1, \\ 0 & \text{if } |k|\ge n. \end{cases}$$

Plugging these quantities into the above formula for $f_X(\lambda)$ we obtain the following estimators of the spectral density:

$$I_n(\lambda) \;=\; \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\widetilde{\gamma}_n(k)\cos(k\lambda),$$

$$I_{n,\bar{X}_n}(\lambda) \;=\; \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\widehat{\gamma}_n(k)\cos(k\lambda).$$

$I_n$ is called the **periodogram** and $I_{n,\bar{X}_n}$ the **centered periodogram** of the data set $X_1,\ldots,X_n$. At first glance, the centered periodogram $I_{n,\bar{X}_n}$ seems to be better motivated than the periodogram $I_n$ since it is based on more reliable estimators of the autocovariances which are also appropriate if $EX_t \neq 0$. Nevertheless, we will see below that these estimators coincide at particular frequencies $\lambda$. For a sample of size $n$, we define the so-called **Fourier frequencies** by

$$\lambda_k \;=\; \frac{2\pi k}{n}, \qquad \text{for } k\in\mathbb{Z} \text{ such that } \lambda_k\in(-\pi,\pi].$$

Note that $\lambda_k \in (-\pi, \pi]$ holds if and only if $k \in \big\{ -\lfloor \frac{n-1}{2} \rfloor, -\lfloor \frac{n-1}{2} \rfloor + 1, \ldots, \lfloor \frac{n}{2} \rfloor \big\}$, where $\lfloor a \rfloor$ denotes the largest integer not greater than $a$. For technical reasons, we also define the quantity

$$I_{n,\mu}(\lambda) \;=\; \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \Big( \frac{1}{n} \sum_{t=1}^{n-|k|} \big(X_{t+|k|} \;-\; \mu\big)\big(X_t \;-\; \mu\big) \Big) \cos(k\lambda).$$

The following lemma provides a few algebraic properties of the periodogram.

**Lemma 2.2.1.** *(i)* $\quad I_n(\lambda) \;=\; \frac{1}{2\pi n} \Big| \sum_{t=1}^n X_t \, e^{-it\lambda} \Big|^2 \qquad \forall \lambda \in [-\pi, \pi].$
*(Analogous formulas hold true for $I_{n,\mu}$ and $I_{n,\bar{X}_n}$ with $X_t - \mu$ and $X_t - \bar{X}_n$ instead of $X_t$, respectively.)*
*(ii) For each Fourier frequency $\lambda_k \neq 0$,*

$$I_n(\lambda_k) \;=\; I_{n,\mu}(\lambda_k) \;=\; I_{n,\bar{X}_n}(\lambda_k).$$

*Proof.* (i) It holds that

$$\begin{aligned} \frac{1}{2\pi n} \Big| \sum_{t=1}^n X_t \, e^{-it\lambda} \Big|^2 \;&=\; \frac{1}{2\pi n} \sum_{s,t=1}^n X_s X_t \, e^{-i(s-t)\lambda} \\ &=\; \frac{1}{2\pi} \frac{1}{n} \sum_{s,t=1}^n X_s X_t \cos\big((s-t)\lambda\big) \\ &=\; \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \widetilde{\gamma}_n(k) \cos(k\lambda). \end{aligned}$$

(ii) Since $c + c^2 + \cdots + c^n = c(1 + c + \cdots + c^{n-1}) = c \, \frac{1-c^n}{1-c}$ holds for all $c \in \mathbb{C}$ with $c \neq 1$, we obtain, for a Fourier frequency $\lambda_k = 2\pi k/n \neq 0$,

$$\sum_{t=1}^n e^{-it\lambda_k} \;=\; \sum_{t=1}^n \big(e^{-i\lambda_k}\big)^t \;=\; e^{-i\lambda_k} \frac{1 \;-\; e^{-i\,2\pi k}}{1 \;-\; e^{-i\lambda_k}} \;=\; 0.$$

This implies that

$$\sum_{t=1}^n X_t \, e^{-it\lambda_k} \;=\; \sum_{t=1}^n \big(X_t - \mu\big) e^{-it\lambda_k} \;=\; \sum_{t=1}^n \big(X_t - \bar{X}_n\big) e^{-it\lambda_k},$$

which proves (ii). $\square$

**Theorem 2.2.2.** *Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a stationary and real-valued process with autocovariance function $\gamma$, $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$.*

*(i)* *If $EX_t = 0$, then*

$$\sup_{\lambda\in[-\pi,\pi]} \big|EI_n(\lambda) \,-\, f_X(\lambda)\big| \underset{n\to\infty}{\longrightarrow} 0.$$

*(ii)* *If $EX_t = \mu$, then*

$$EI_n(0) \,-\, \frac{n\,\mu^2}{2\pi} \underset{n\to\infty}{\longrightarrow} f_X(0)$$

*and*

$$\max\Big\{\big|EI_n(\lambda_k) \,-\, f_X(\lambda_k)\big|: \quad \lambda_k = 2\pi k/n \in (-\pi,\pi], \lambda_k \neq 0\Big\} \underset{n\to\infty}{\longrightarrow} 0.$$

*Proof.* (i) Since $EI_n(\lambda) = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}(1-|k|/n)\gamma(k)\cos(k\lambda)$ we obtain by dominated convergence

$$\big|EI_n(\lambda) \,-\, f_X(\lambda)\big| \,\leq\, \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \big||k|/n\wedge 1\big|\,|\gamma(k)| \underset{n\to\infty}{\longrightarrow} 0.$$

(ii) It holds that

$$\begin{aligned} EI_n(0) &= \frac{1}{2\pi n} E\Big|\sum_{t=1}^{n} X_t\Big|^2 \\ &= \frac{1}{2\pi n}\Big\{E\Big[\Big(\sum_{t=1}^{n}(X_t-\mu)\Big)^2\Big] \quad + \quad n^2\mu^2\Big\} \\ &= \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\big(1-|k|/n\big)\gamma(k) \quad + \quad \frac{n\mu^2}{2\pi}, \end{aligned}$$

which implies, again by dominated convergence,

$$EI_n(0) \,-\, \frac{n\,\mu^2}{2\pi} \,=\, \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\big(1-|k|/n\big)\gamma(k) \underset{n\to\infty}{\longrightarrow} \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k) \,=\, f_X(0).$$

For each Fourier frequency $\lambda_k \neq 0$, we have that $I_n(\lambda_k) = I_{n,\mu}(\lambda_k)$. Therefore, we have the equality

$$EI_n(\lambda_k) \,=\, EI_{n,\mu}(\lambda_k) \,=\, \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\big(1-|k|/n\big)\gamma(k)\cos(k\lambda).$$

Hence, we obtain the second statement of part (ii) in complete analogy to (i). □

Although the periodogram is asymptotically unbiased for $f_X(\lambda)$ at the Fourier frequencies $\lambda_k \neq 0$, the following lemma shows the disappointing fact that its variance does not vanish as $n \to \infty$, even in case of a white noise.

**Lemma 2.2.3.** *Let* $(X_t)_{t\in\mathbb{Z}} \sim IID(0,\sigma^2)$, $EX_t^4 =: \eta < \infty$. *Then, for Fourier frequencies* $\lambda_k = 2\pi k/n \quad (k = -\left[\frac{n-1}{2}\right], \dots, \left[\frac{n}{2}\right])$,

*(i)* *If* $|\lambda_j| \neq |\lambda_k|$, *then*

$$\operatorname{cov}\big(I_n(\lambda_j), I_n(\lambda_k)\big) \;=\; \frac{\eta - 3\sigma^2}{4\pi^2 n},$$

*i.e., the values of the periodogram are asymptotically uncorrelated.*

*(ii)* *It holds that*

$$\operatorname{var}\big(I_n(\lambda_k)\big) \;=\; \begin{cases} \frac{\eta-3\sigma^4}{4\pi^2 n} + \frac{\sigma^4}{2\pi^2}, & \text{if } \lambda_k \in \{0,\pi\}, \\ \frac{\eta-3\sigma^4}{4\pi^2 n} + \frac{\sigma^4}{4\pi^2}, & \text{if } \lambda_k \notin \{0,\pi\}. \end{cases}$$

*Proof.* We start off with a few preparatory calculations. It holds that

$$EI_n(\lambda) \;=\; \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \big(1-|k|/n\big) \underbrace{\gamma(k)}_{=\sigma^2\delta_{0,k}} \cos(k\lambda) \;=\; \frac{\sigma^2}{2\pi} \qquad \forall \lambda \in [-\pi,\pi].$$

Furthermore, we have that

$$\begin{aligned} E\big[I_n(\lambda_j)\, I_n(\lambda_k)\big] \;&=\; \frac{1}{4\pi^2 n^2} E\Big[\Big|\sum_{t=1}^n X_t e^{-it\lambda_j}\Big|^2 \Big|\sum_{t=1}^n X_t e^{-it\lambda_k}\Big|^2\Big] \\ &=\; \frac{1}{4\pi^2 n^2} \sum_{s,t=1}^n \sum_{u,v=1}^n E\big[X_s X_t X_u X_v\big] e^{-i(s-t)\lambda_j} e^{-i(u-v)\lambda_k}. \end{aligned}$$

Note that $E\big[X_s X_t X_u X_v\big] \neq 0$ is only possible if $s=t=u=v$ or if each index appears twice. Therefore,

$$\begin{aligned} E\big[I_n(\lambda_j)\, I_n(\lambda_k)\big] \;&=\; \frac{1}{4\pi^2 n^2} \sum_{t=1}^n E\big[X_t^4\big] \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{(s,t,u,v):\; s=t\neq u=v} \underbrace{E\big[X_s^2\big]}_{=\sigma^2} \underbrace{E\big[X_u^2\big]}_{=\sigma^2} \underbrace{e^{-i(s-t)\lambda_j}}_{=1} \underbrace{e^{-i(u-v)\lambda_k}}_{=1} \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{(s,t,u,v):\; s=u\neq t=v} \underbrace{E\big[X_s^2\big]}_{=\sigma^2} \underbrace{E\big[X_t^2\big]}_{=\sigma^2} e^{-i(s-t)\lambda_j} e^{-i(u-v)\lambda_k} \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{(s,t,u,v):\; s=v\neq t=u} \underbrace{E\big[X_s^2\big]}_{=\sigma^2} \underbrace{E\big[X_t^2\big]}_{=\sigma^2} e^{-i(s-t)\lambda_j} e^{-i(u-v)\lambda_k} \\ &=\; \frac{1}{4\pi^2 n^2} \sum_{t=1}^n \big(\eta - 3\sigma^2\big) \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{s,u=1}^n \sigma^4 \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{s,t=1}^n \sigma^4\, e^{-i(s-t)(\lambda_j+\lambda_k)} \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{s,t=1}^n \sigma^4\, e^{-i(s-t)(\lambda_j-\lambda_k)} \\ &=:\; T_1 + \cdots + T_4, \end{aligned}$$

say. Now we are in a position to prove (i) and (ii).
(i) We have that

$$T_2 \;=\; \frac{\sigma^4}{4\pi^2} \;=\; EI_n(\lambda_j)\, EI_n(\lambda_k).$$

Since $|\lambda_j| \neq |\lambda_k|$ we have $\lambda_j + \lambda_k, \lambda_j - \lambda_k \notin \{0, 2\pi\}$, which leads to $\sum_{t=1}^n e^{-i(s-t)(\lambda_j+\lambda_k)} = \sum_{t=1}^n e^{-i(s-t)(\lambda_j-\lambda_k)} = 0$ and, therefore,

$$T_3 \;=\; T_4 \;=\; 0.$$

Hence,

$$\operatorname{cov}\big(I_n(\lambda_j), I_n(\lambda_k)\big) \;=\; T_1 \;=\; \frac{\eta \,-\, 3\sigma^4}{4\pi^2 n}.$$

(ii) We have, as above,

$$T_1 \;=\; \frac{\eta \,-\, 3\sigma^4}{4\pi^2 n} \qquad \text{and} \qquad T_2 \;=\; \big(EI_n(\lambda_k)\big)^2.$$

If $\lambda_k \in \{0, \pi\}$, then $2\lambda_k \in \{0, 2\pi\}$, which implies

$$T_3 \;=\; \frac{\sigma^4}{4\pi^2}.$$

On the other hand, if $\lambda_k \notin \{0, \pi\}$, then $2\lambda_k \notin \{0, 2\pi\}$, which leads to

$$T_3 \;=\; 0.$$

In both cases,

$$T_4 \;=\; \frac{\sigma^4}{4\pi^2}.$$

Collecting all terms we see that

$$\operatorname{var}\big(I_n(\lambda_k)\big) \;=\; T_1 \,+\, T_3 \,+\, T_4,$$

which completes the proof of the second statement. □

**Theorem 2.2.4.** *Suppose that $(X_t)_{t\in\mathbb{Z}} \sim IID(0, \sigma^2)$. Then, for Fourier frequencies $\lambda_k = 2\pi k/n$ $(k = -\lfloor \frac{n-1}{2} \rfloor, \ldots, \lfloor \frac{n}{2} \rfloor)$,*

*(i)* $I_n(\lambda_k) \stackrel{d}{\longrightarrow} \frac{\sigma^2}{2\pi} Z_1^2$, *if* $\lambda_k \in \{0, \pi\}$,

*(ii)* $I_n(\lambda_k) \stackrel{d}{\longrightarrow} \frac{\sigma^2}{2\pi}\big(Z_1^2 + Z_2^2\big)/2$, *if* $\lambda_k \notin \{0, \pi\}$,

*where $Z_1, Z_2 \sim N(0,1)$ are independent.*

*Proof.* First of all, we rewrite the periodogram in a suitable form:

$$\begin{aligned} I_n(\lambda_k) \;&=\; \frac{1}{2\pi n}\Big|\sum_{t=1}^n X_t\, e^{-it\lambda_k}\Big|^2 \\ &=\; \frac{1}{2\pi n}\Big|\sum_{t=1}^n X_t\,\big\{\cos(t\lambda_k) \,-\, i\,\sin(t\lambda_k)\big\}\Big|^2 \\ &=\; \frac{1}{2\pi}\Big\{\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^n X_t\,\cos(t\lambda_k)\Big)^2 \,+\, \Big(\frac{1}{\sqrt{n}}\sum_{t=1}^n X_t\,\sin(t\lambda_k)\Big)^2\Big\}. \end{aligned}$$

(i) Let $\lambda_k \in \{0, \pi\}$. Then

$$\sin\big(t\lambda_k\big) \;=\; 0 \quad \text{and} \quad \cos\big(t\lambda_k\big) \;=\; \begin{cases} 1, & \text{if } \lambda_k = 0, \\ (-1)^k, & \text{if } \lambda_k = \pi \end{cases}$$

holds for all $t = 1, \ldots, n$. Since, by the Lindeberg-Feller central limit theorem (Theorem 1.5.10),

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \cos\big(t\lambda_k\big)\, X_t \;\xrightarrow{d}\; \sigma\, Z_1 \;\sim\; N\big(0, \sigma^2\big).$$

we obtain by the Continuous Mapping Theorem that

$$I_n(\lambda_k) \;=\; \frac{1}{2\pi}\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \cos\big(t\lambda_k\big)\, X_t\Big)^2 \;\xrightarrow{d}\; \frac{1}{2\pi}(\sigma Z_1)^2 \;=\; \frac{\sigma^2}{2\pi} Z_1^2.$$

(ii) Let $\lambda_k$ be a Fourier frequency, $\lambda_k \notin \{0, \pi\}$. In this case, both the cosine and sine terms matter and we show that

$$Z^{(n)} \;:=\; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} X_t\, \cos(t\lambda_k) \\ X_t\, \sin(t\lambda_k) \end{pmatrix} \;\xrightarrow{d}\; \frac{\sigma}{\sqrt{2}} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \;\sim\; N\Bigg(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2/2 & 0 \\ 0 & \sigma^2/2 \end{pmatrix}\Bigg). \tag{2.2.1}$$

From (2.2.1) we obtain, again by the Continuous Mapping Theorem, that

$$\begin{aligned} I_n(\lambda_k) \;&=\; \frac{1}{2\pi}\Big\{\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t\, \cos(t\lambda_k)\Big)^2 \;+\; \Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t\, \sin(t\lambda_k)\Big)^2\Big\} \\ &\xrightarrow{d}\; \frac{1}{2\pi}\Big\{\Big(\frac{\sigma}{\sqrt{2}} Z_1\Big)^2 \;+\; \Big(\frac{\sigma}{\sqrt{2}} Z_2\Big)^2\Big\} \;=\; \frac{\sigma^2}{2\pi}\big(Z_1^2 + Z_2^2\big)/2. \end{aligned}$$

It remains to prove (2.2.1). According to the Cramér-Wold device, (2.2.1) is equivalent to

$$c^T Z^{(n)} = \sum_{t=1}^{n} \underbrace{\frac{1}{\sqrt{n}}\big(c_1 \cos(t\lambda_k) \;+\; c_2 \sin(t\lambda_k)\big) X_t}_{=:\, Z_{n,t}} \;\xrightarrow{d}\; \frac{\sigma}{\sqrt{2}}\, c^T \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\Big(0, \frac{\sigma^2}{2}\big(c_1^2 + c_2^2\big)\Big)$$

$$\text{for all } c = (c_1, c_2)^T \in \mathbb{R}^2. \tag{2.2.2}$$

To show this, we use once more the Lindeberg-Feller central limit theorem (Theorem 1.5.10). We verify that the conditions of this theorem are satisfied. For each $n \in \mathbb{N}$, the random variables $Z_{n,1}, \ldots, Z_{n,n}$ are stochastically independent and we have

$$E Z_{n,t} \;=\; 0$$

and

$$\begin{aligned} \sum_{t=1}^{n} E\big[Z_{n,t}^2\big] \;&=\; \frac{1}{n} \sum_{t=1}^{n} \Big\{c_1^2 \cos^2(t\lambda_k) \;+\; c_2^2 \sin^2(t\lambda_k) \;+\; 2c_1 c_2 \cos(t\lambda_k) \sin(t\lambda_k)\Big\} E\big[X_t^2\big] \\ &=\; \frac{\sigma^2(c_1^2 \;+\; c_2^2)}{2}. \end{aligned} \tag{2.2.3}$$

Note that the second equality in (2.2.3) follows from the trigonometric identities

$$\sum_{t=1}^{n} \cos^2(t\lambda_k) \;=\; \sum_{t=1}^{n} \sin^2(t\lambda_k) \;=\; \frac{n}{2} \tag{2.2.4a}$$

and

$$\sum_{t=1}^{n} \cos(t\lambda_k)\sin(t\lambda_k) \;=\; 0, \tag{2.2.4b}$$

which hold true for all Fourier frequencies $\lambda_k \notin \{0, \pi\}$. Finally, we obtain by Lebesgue's theorem on dominated convergence that, for arbitrary $\epsilon > 0$,

$$\begin{aligned} L_n(\epsilon) \;&=\; \sum_{t=1}^{n} E\Big[ \underbrace{Z_{n,t}^2}_{\le (|c_1|+|c_2|)^2 X_t^2/n} \mathbb{1}\big(|Z_{n,t}| > \epsilon\big)\Big] \\ &\le\; \frac{(|c_1| + |c_2|)^2}{n} \sum_{t=1}^{n} E\Big[X_t^2\, \mathbb{1}\big((|c_1| + |c_2|)|X_t| \;>\; \sqrt{n}\,\varepsilon\big)\Big] \\ &=\; (|c_1| + |c_2|)^2\, E\Big[X_1^2\, \mathbb{1}\big((|c_1| + |c_2|)|X_1| \;>\; \sqrt{n}\,\varepsilon\big)\Big] \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Hence, all conditions of the Lindeberg-Feller central limit theorem are satisfied. Since the variance of the limit variable is given by the right-hand side of (2.2.3) we obtain that (2.2.2) holds true.

□

## Exercise

**Ex. 2.2.1** Let $\lambda_k = 2\pi k/n$ be a Fourier frequency, $\lambda_k \notin \{0, \pi\}$. Show that

$$\sum_{t=1}^{n} \cos^2(t\lambda_k) \;=\; \sum_{t=1}^{n} \sin^2(t\lambda_k) \;=\; \frac{n}{2}$$

and

$$\sum_{t=1}^{n} \cos(t\lambda_k)\sin(t\lambda_k) \;=\; 0.$$

*Hint: Prove first that* $\cos^2(x) - \sin^2(x) = \sin(2x)$ *and* $\cos(x)\sin(x) = \sin(2x)$ $\forall x \in \mathbb{R}$.

Now we generalize the result of Theorem 2.2.4 to the case of linear processes. Recall that, for $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \text{IID}(0,\sigma^2)$ and $X_t = \sum_{t=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$, where $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$, the process $(X_t)_{t\in\mathbb{Z}}$ has a spectral density $f_X$,

$$f_X(\lambda) \;=\; \left|\beta(e^{-i\lambda})\right|^2 f_\varepsilon(\lambda) \qquad \forall \lambda \in [-\pi,\pi],$$

where $\beta(e^{-i\lambda}) = \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda}$ and $f_\varepsilon(\lambda) = \sigma^2/(2\pi)$. The following theorem shows that there exists a similar relation between the respective periodograms based on $X_1,\ldots,X_n$ and $\varepsilon_1,\ldots,\varepsilon_n$.

**Theorem 2.2.5.** *Suppose that* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim IID(0,\sigma^2)$, $X_t = \sum_{t=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$, *where* $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$. *Let* $I_n^X$ *and* $I_n^\varepsilon$ *be the periodograms based on* $X_1,\ldots,X_n$ *and* $\varepsilon_1,\ldots,\varepsilon_n$, *respectively. Then*

$$I_n^X(\lambda) \;=\; \left|\beta(e^{-i\lambda})\right|^2 I_n^\varepsilon(\lambda) \;+\; R_n(\lambda),$$

*where*

$$\sup_{\lambda\in[-\pi,\pi]} E\left|R_n(\lambda)\right| \underset{n\to\infty}{\longrightarrow} 0.$$

*Proof.* Let

$$J_n^X(\lambda) \;:=\; n^{-1/2} \sum_{t=1}^n X_t\, e^{-it\lambda} \qquad \text{and} \qquad J_n^\varepsilon(\lambda) \;:=\; n^{-1/2} \sum_{t=1}^n \varepsilon_t\, e^{-it\lambda}$$

denote the so-called finite Fourier transforms of $X_1,\ldots,X_n$ and $\varepsilon_1,\ldots,\varepsilon_n$, respectively. We establish the connection between $J_n^X(\lambda)$ and $J_n^\varepsilon(\lambda)$:

$$\begin{aligned}
J_n^X(\lambda) &= n^{-1/2} \sum_{t=1}^n \Big( \sum_{k=-\infty}^{\infty} \beta_k \varepsilon_{t-k} \Big) e^{-it\lambda} \\
&= n^{-1/2} \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda} \sum_{t=1}^n \varepsilon_{t-k} e^{-i(t-k)\lambda} \\
&= n^{-1/2} \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda} \sum_{t=1-k}^{n-k} \varepsilon_t e^{-it\lambda} \\
&= \beta\big(e^{-i\lambda}\big) J_n^\varepsilon(\lambda) \quad + \quad n^{-1/2} \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda} \Big\{ \sum_{t=1-k}^{n-k} \varepsilon_t e^{-it\lambda} - \sum_{t=1}^n \varepsilon_t e^{-it\lambda} \Big\},
\end{aligned}$$

i.e.,

$$J_n^X(\lambda) \;=\; \beta(e^{-i\lambda})\, J_n^\varepsilon(\lambda) \;+\; Z_n,$$

where $Z_n = n^{-1/2} \sum_{k=-\infty}^{\infty} \beta_k e^{-ik\lambda} U_{n,k}$ and $U_{n,k} = \sum_{t=1-k}^{n-k} \varepsilon_t e^{-it\lambda} - \sum_{t=1}^n \varepsilon_t e^{-it\lambda}$. Since $I_n^X(\lambda) = |J_n^X(\lambda)|^2/(2\pi)$ and $I_n^\varepsilon(\lambda) = |J_n^\varepsilon(\lambda)|^2/(2\pi)$ we obtain

$$\begin{aligned}
I_n^X(\lambda) \;-\; \left|\beta(e^{-i\lambda})\right|^2 I_n^\varepsilon(\lambda) &= \frac{1}{2\pi}\left|Z_n\right|^2 + \frac{1}{2\pi}\beta(e^{-i\lambda}) J_n^\varepsilon(\lambda)\, \overline{Z_n} + \frac{1}{2\pi}\overline{\beta(e^{-i\lambda}) J_n^\varepsilon(\lambda)}\, Z_n \\
&=: R_{n,1} \;+\; R_{n,2} \;+\; R_{n,3},
\end{aligned}$$

say. Note that if $|k| < n$, then $U_{n,k}$ is a sum of $2|k|$ independent random variables, whereas if $|k| \geq n$, $U_{n,k}$ is a sum of $2n$ independent random variables. It follows that

$$E\big[|U_{n,k}|^2\big] \;=\; 2\sigma^2 \min\{|k|, n\}.$$

We obtain by the Minkowski inequality (named after the German mathematician Hermann Minkowski) that

$$\begin{aligned}
\sqrt{E\big[|Z_n|^2\big]} \quad &= \quad \frac{1}{\sqrt{n}} \Big\| \sum_{k=-\infty}^{\infty} \beta_k\, e^{-ik\lambda}\, U_{n,k} \Big\| \\
&\leq \quad \frac{1}{\sqrt{n}} \Big\| \sum_{k=-\infty}^{\infty} \big|\beta_k\, e^{-ik\lambda}\, U_{n,k}\big| \Big\| \\
&= \quad \lim_{m\to\infty} \frac{1}{\sqrt{n}} \Big\| \sum_{k=-m}^{m} \big|\beta_k\, e^{-ik\lambda}\, U_{n,k}\big| \Big\| \\
&\underbrace{\leq}_{Minkowski} \quad \lim_{m\to\infty} \frac{1}{\sqrt{n}} \sum_{k=-m}^{m} \Big\|\beta_k\, e^{-ik\lambda}\, U_{n,k}\Big\| \\
&= \quad O\Big( \sum_{k=-\infty}^{\infty} |\beta_k| \sqrt{|k|/n \wedge 1} \Big) \underset{n\to\infty}{\longrightarrow} 0, \qquad (2.2.5)
\end{aligned}$$

where the latter step follows by dominated convergence. It follows directly from (2.2.5) that

$$\sup_{\lambda\in[-\pi,\pi]} \big\{ E\big|R_{n,1}\big| \big\} \underset{n\to\infty}{\longrightarrow} 0.$$

Since $|\beta(e^{-i\lambda})| \leq \sum_{k=-\infty}^{\infty} |\beta_k| < \infty$ and

$$E\Big[\big|J_n^{\varepsilon}(\lambda)\big|^2\Big] \;=\; 2\pi\, E\big[I_n^{\varepsilon}(\lambda)\big] \;=\; \sigma^2$$

we obtain, by Cauchy-Schwarz and again (2.2.5), that

$$\sup_{\lambda\in[-\pi,\pi]} \big\{ E|R_{n,2}| \,+\, E|R_{n,3}| \big\} \underset{n\to\infty}{\longrightarrow} 0,$$

which completes the proof. □

Theorems 2.2.4 and 2.2.5 allow us to derive the asymptotic behavior of the periodogram for general linear processes.

**Corollary 2.2.6.** *Suppose that* $(\varepsilon_t)_{t\in\mathbb{Z}} \sim IID(0, \sigma^2)$, $X_t = \sum_{t=-\infty}^{\infty} \beta_k \varepsilon_{t-k}$, *where* $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$. *Let* $I_n^X$ *be the periodogram based on* $X_1, \ldots, X_n$. *Then, for Fourier frequencies* $\lambda_k = 2\pi k/n$ $(k = -\lfloor \frac{n-1}{2} \rfloor, \ldots, \lfloor \frac{n}{2} \rfloor)$,

*(i)* $\quad I_n^X(\lambda_k) \overset{d}{\longrightarrow} f_X(\lambda_k) Z_1^2, \qquad$ *if* $\lambda_k \in \{0, \pi\}$,

*(ii)* $\quad I_n^X(\lambda_k) \overset{d}{\longrightarrow} f_X(\lambda_k)\big(Z_1^2 + Z_2^2\big)/2, \qquad$ *if* $\lambda_k \notin \{0, \pi\}$,

*where* $Z_1, Z_2 \sim N(0, 1)$ *are independent.*

*Proof.* Let $I_n^\varepsilon$ be the periodogram based on $\varepsilon_1, \ldots, \varepsilon_n$. It follows from Theorem 2.2.5 that

$$I_n^X(\lambda_k) \;=\; \left|\beta(e^{-i\lambda_k})\right|^2 I_n^\varepsilon(\lambda_k) \;+\; R_n(\lambda_k),$$

where $R_n(\lambda_k) \xrightarrow{P} 0$.

Since, by Theorem 2.2.4,

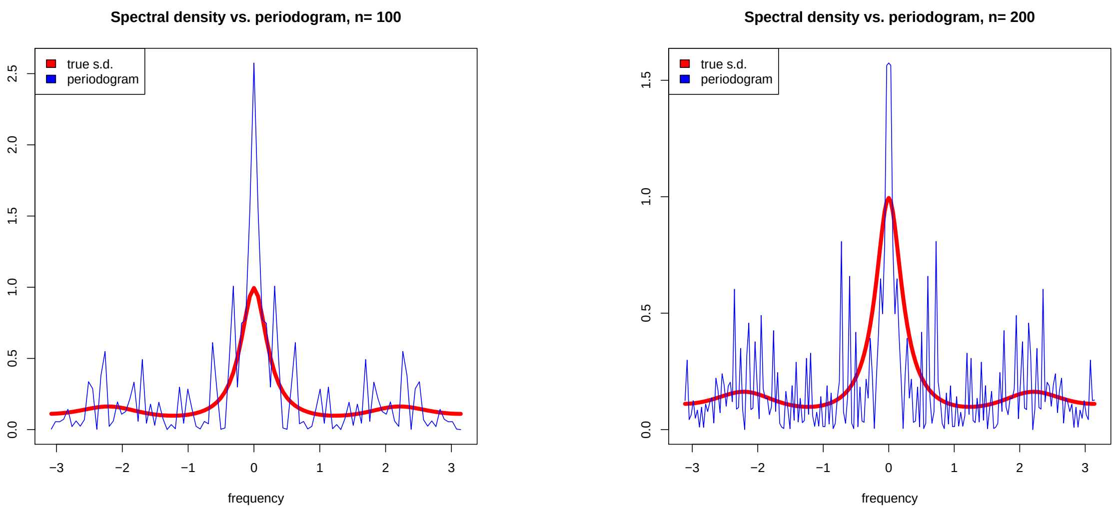
$$I_n^\varepsilon(\lambda_k) \;\xrightarrow{d}\; \begin{cases} \frac{\sigma^2}{2\pi} Z_1^2, & \text{if } \lambda_k \in \{0,\pi\}, \\[2ex] \frac{\sigma^2}{2\pi}\big(Z_1^2 + Z_2^2\big)/2, & \text{if } \lambda_k \notin \{0,\pi\} \end{cases}$$

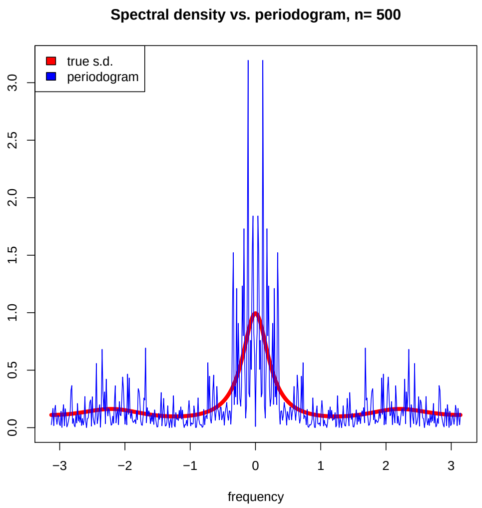
and $f_X(\lambda_k) = f_\varepsilon(\lambda_k)\left|\beta(e^{-i\lambda_k})\right|^2 = \frac{\sigma^2}{2\pi}\left|\beta(e^{-i\lambda_k})\right|^2$ we obtain that

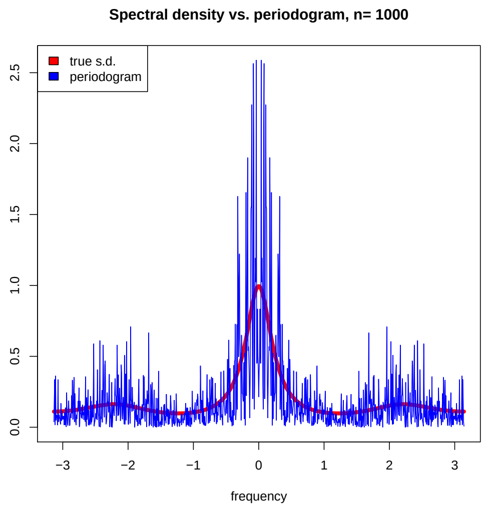
$$\begin{aligned} I_n^X(\lambda_k) \;&\xrightarrow{d}\; \begin{cases} \frac{\sigma^2}{2\pi}\left|\beta(e^{-i\lambda_k})\right|^2 Z_1^2, & \text{if } \lambda_k \in \{0,\pi\}, \\[2ex] \frac{\sigma^2}{2\pi}\left|\beta(e^{-i\lambda_k})\right|^2 \big(Z_1^2 + Z_2^2\big)/2, & \text{if } \lambda_k \notin \{0,\pi\} \end{cases} \\ &= \begin{cases} f_X(\lambda_k) Z_1^2, & \text{if } \lambda_k \in \{0,\pi\}, \\[2ex] f_X(\lambda_k)\big(Z_1^2 + Z_2^2\big)/2, & \text{if } \lambda_k \notin \{0,\pi\} \end{cases} \end{aligned}$$

□

Corollary 2.2.6 underlines once more that the periodogram is not a consistent estimator of the spectral density. This will be corroborated by the following pictures which show the true spectral density $f_X$ of an AR($p$) process with parameters $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$ and independent innovations $\varepsilon_t \sim N(0,1)$ (red lines), together with one realization of the periodogram $I_n$ obtained from samples of size $n$ =100, 200, 500 and 1,000, respectively, (blue lines). As can be seen, the periodogram fluctuates around the true spectral density, however, its variability stays high even for moderately large sample sizes $n$. This is disappointing because we should hope that observing the time series $(X_t)_{t\in\mathbb{Z}}$ long enough will enable us to estimate its spectral density with arbitrary precision.






## Exercise

**Ex. 2.2.2** Suppose that the conditions of Corollary 2.2.6 are fulfilled.

(i) Show that, for $\lambda \in \{0, \pi\}$ and $f_X(\lambda) \neq 0$,

$$P\Big(f_X(\lambda) \in \big[I_n^X(\lambda)/u^2_{1-\alpha_1}, I_n^X(\lambda)/u^2_{1-\alpha_2}\big]\Big) \underset{n\to\infty}{\longrightarrow} 2\big(\alpha_2 - \alpha_1\big),$$

where $u_{1-\alpha} = \Phi^{-1}(1-\alpha)$ and $1/2 < \alpha_1 < \alpha_2 < 1$.

(ii) Show that, for a Fourier frequency $\lambda \notin \{0, \pi\}$ such that $f_X(\lambda) \neq 0$,

$$P\Big(f_X(\lambda) \in \big[I_n^X(\lambda)/(-\ln(\alpha_1)), I_n^X(\lambda)/(-\ln(\alpha_2))\big]\Big) \underset{n\to\infty}{\longrightarrow} \alpha_2 - \alpha_1,$$

where $0 < \alpha_1 < \alpha_2 < 1$.
*Hint: Use the fact that* $\chi^2_2 = Exp(1/2)$.

In the following we consider consistent estimators of the spectral density and the spectral distribution function. We begin with the simpler case of estimating the spectral distribution function $F_X$ which is given by

$$F_X(\lambda) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\lambda} \{\gamma(0) + \sum_{k:\; k\neq 0} \gamma(k)\cos(k\omega)\}\, d\omega \;=\; \frac{1}{2\pi}\Big\{\gamma(0)(\lambda+\pi) + 2\sum_{k=1}^{\infty}\gamma(k)\frac{\sin(k\lambda)}{k}\Big\}.$$

Recall that the periodogram $I_n(\lambda)$ and the centered periodogram $I_{n,\bar{X}_n}(\lambda)$ are estimators of $f_X(\lambda)$ based on $\widetilde{\gamma}_n(k) = n^{-1}\sum_{t=1}^{n-|k|} X_{t+|k|}X_t$ and $\widehat{\gamma}_n(k) = n^{-1}\sum_{t=1}^{n-|k|}\big(X_{t+|k|} - \bar{X}_n\big)\big(X_t - \bar{X}_n\big)$, respectively. ($\widetilde{\gamma}_n(k) = \widehat{\gamma}_n(k) = 0$ if $|k| \geq n$.) This leads to the following estimators of the spectral distribution function.

$$\widehat{F}_n(\lambda) \;=\; \int_{-\pi}^{\lambda} I_n(\omega)\, d\omega \;=\; \frac{1}{2\pi}\Big\{\widetilde{\gamma}_n(0)(\pi+\lambda) \;+\; 2\sum_{k=1}^{n-1}\widetilde{\gamma}_n(k)\frac{\sin(k\lambda)}{k}\Big\}$$

and

$$\widehat{F}_{n,\bar{X}_n}(\lambda) \;=\; \int_{-\pi}^{\lambda} I_{n,\bar{X}_n}(\omega)\, d\omega \;=\; \frac{1}{2\pi}\Big\{\widehat{\gamma}_n(0)(\pi+\lambda) \;+\; 2\sum_{k=1}^{n-1}\widehat{\gamma}_n(k)\frac{\sin(k\lambda)}{k}\Big\}.$$

The following lemma shows that both $\widehat{F}_n(\lambda)$ and $\widehat{F}_{n,\bar{X}_n}(\lambda)$ are asymptotically unbiased estimators of $F_X(\lambda)$.

**Lemma 2.2.7.** *Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a weakly stationary and real-valued process with autocovariance function $\gamma$, $\sum_{k=-\infty}^{\infty}\big|\gamma(k)\big| < \infty$.*

*(i) If $EX_t = 0$, then*

$$\sup_{\lambda\in[-\pi,\pi]}\big|E\widehat{F}_n(\lambda) \;-\; F_X(\lambda)\big| \;=\; O\big(n^{-1}\big).$$

*(ii) For arbitrary $EX_t$,*

$$\sup_{\lambda\in[-\pi,\pi]}\big|E\widehat{F}_{n,\bar{X}_n}(\lambda) \;-\; F_X(\lambda)\big| \;=\; O\big(n^{-1}\ln(n)\big).$$

*Proof.* (i) Suppose that $EX_t = 0$. Then $E\widetilde{\gamma}_n(k) - \gamma(k) = \big((1-|k|/n)_+ - 1\big)\gamma(k)$, which implies that

$$\begin{aligned}\big|E\widehat{F}_n(\lambda) \;-\; F_X(\lambda)\big| \;&=\; \Big|\frac{1}{\pi}\sum_{k=1}^{\infty}\Big(\big(1-|k|/n\big)_+ \;-\; 1\Big)\gamma(k)\frac{\sin(k\lambda)}{k}\Big| \\ &\leq\; \frac{1}{\pi}\sum_{k=1}^{\infty}\underbrace{\Big(1 \;-\; \big(1-|k|/n\big)_+\Big)}_{\leq k/n}\frac{1}{k}\big|\gamma(k)\big| \;=\; O\big(n^{-1}\big).\end{aligned}$$

(ii) In this case, the calculations in the proof of Lemma 1.4.8 show that

$$E\widehat{\gamma}_n(k) \;=\; \big(1 \;-\; |k|/n\big)_+\gamma(k) \;+\; O\big(n^{-1}\big),$$

which implies that

$$\begin{aligned}\big|E\widehat{F}_{n,\bar{X}_n}(\lambda) \;-\; F_X(\lambda)\big| \;&\leq\; \frac{1}{n}\sum_{k=1}^{n}\Big(1 \;-\; \big(1 \;-\; |k|/n\big)_+\Big)\frac{1}{k}\big|\gamma(k)\big| \;+\; O\Big(\frac{1}{n}\,\big(1 \;+\; \sum_{k=1}^{n}\frac{1}{k}\big)\Big) \\ &=\; O\big(n^{-1}\ln(n)\big).\end{aligned}$$

□

**Lemma 2.2.8.** *Let* $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ *be a* **strictly** *stationary process with autocovariance function* $\gamma$ *such that* $EX_t = \mu$, $\sum_{k=-\infty}^{\infty} \big|\gamma(k)\big| < \infty$, $EX_t^4 < \infty$, *and* $\sum_{t,u,v=-\infty}^{\infty} \big|\operatorname{cum}\big(X_0-\mu, X_t-\mu, X_u-\mu, X_v-\mu\big)\big| < \infty$.

*(i)* *If* $\mu = 0$, *then*

$$\operatorname{var}\big(\widehat{F}_n(\lambda)\big) \;=\; O\big(n^{-1}\big).$$

*(ii)* *For arbitrary* $\mu$,

$$\operatorname{var}\big(\widehat{F}_{n,\bar{X}_n}(\lambda)\big) \;=\; \operatorname{var}\big(\widehat{F}_{n,\mu}(\lambda)\big) \;+\; O\big(n^{-2/3}\ln(n)\big),$$

*where* $\widehat{F}_{n,\mu}(\lambda) = \int_{-\pi}^{\lambda} I_{n,\mu}(\omega)\,d\omega$).

*Proof.* (i) We represent $\widehat{F}_n(\lambda)$ as a quadratic form, $\widehat{F}_n(\lambda) = \sum_{s,t=1}^n A_{st}X_sX_t$, where

$$A_{st} \;=\; \begin{cases} \frac{1}{2\pi n}(\pi+\lambda) & \text{if } s=t, \\ \frac{1}{2\pi n}\frac{\sin((s-t)\lambda)}{s-t} & \text{if } s\neq t. \end{cases}$$

It follows from Lemma 1.4.9 that

$$\begin{aligned} \operatorname{var}\big(\widehat{F}_n(\lambda)\big) \;&=\; \sum_{s,t,u,v=1}^n \underbrace{A_{st}}_{=O(n^{-1})}\;\underbrace{A_{uv}}_{=O(n^{-1})}\;\operatorname{cum}\big(X_s,X_t,X_u,X_v\big) \;+\; 2\operatorname{tr}\big(A\Sigma A\Sigma\big) \\ &=:\; T_{n,1} \;+\; T_{n,2}, \end{aligned}$$

where $A = \big((A_{st})\big)_{s,t=1,\dots,n}$, $\Sigma = \big((\gamma(s-t))\big)_{s,t=1,\dots,n}$. It is obvious, that

$$T_{n,1} \;=\; O\big(n^{-1}\big).$$

The term $|T_{n,2}|$ can be estimated by

$$\begin{aligned} \big|T_{n,2}\big| \;&\leq\; 2\sum_{s,t,u,v=1}^n \big|A_{st}\Sigma_{tu}A_{uv}\Sigma_{vs}\big| \\ &\leq\; \sum_{s,t,u,v=1}^n \big(A_{st}^2 \,+\, A_{uv}^2\big)\,\big|\Sigma_{tu}\big|\,\big|\Sigma_{vs}\big| \\ &\leq\; \max_{1\leq s,t\leq n}\Big\{\sum_{u=1}^n \underbrace{\big|\Sigma_{tu}\big|}_{=|\gamma(t-u)|}\;\sum_{v=1}^n \underbrace{\big|\Sigma_{vs}\big|}_{=|\gamma(v-s)|}\Big\}\sum_{s,t=1}^n A_{st}^2 \\ &\quad+\; \max_{1\leq u,v\leq n}\Big\{\sum_{s=1}^n \underbrace{\big|\Sigma_{vs}\big|}_{=|\gamma(v-s)|}\;\sum_{t=1}^n \underbrace{\big|\Sigma_{tu}\big|}_{=|\gamma(t-u)|}\Big\}\sum_{u,v=1}^n A_{uv}^2 \\ &\leq\; 2\Big(\sum_{k=-\infty}^{\infty}|\gamma(k)|\Big)^2\,\sum_{s,t=1}^n A_{st}^2 \end{aligned} \tag{2.2.6}$$

and it follows from $\max_{1\leq s\leq n}\sum_{t=1}^n A_{st}^2 \;\leq\; \frac{1}{n^2} \,+\, \frac{1}{4n^2\pi^2}\sum_{k=1}^{n-1}\frac{1}{k^2} \;=\; O\big(n^{-2}\big)$ that

$$T_{n,2} \;=\; O\big(n^{-1}\big).$$

(ii) We have

$$\begin{aligned}
\widehat{\gamma}_n(k) &= \frac{1}{n} \sum_{t=1}^{n-|k|} \big(X_{t+|k|} - \mu \,+\, \mu - \bar{X}_n\big)\big(X_t - \mu \,+\, \mu - \bar{X}_n\big) \\
&= \underbrace{\frac{1}{n} \sum_{t=1}^{n-|k|} \big(X_{t+|k|} - \mu\big)\big(X_t - \mu\big)}_{=:\widehat{\gamma}_{n,\mu}(k)} + \sum_{s,t=1}^{n} M_{st}^{(k,n)} \big(X_s - \mu\big)\big(X_t - \mu\big),
\end{aligned}$$

where $\max_{s,t}\big\{\big|M_{st}^{(k,n)}\big|\big\} = O\big(n^{-2}\big)$. This implies

$$\begin{aligned}
\widehat{F}_{n,\bar{X}_n}(\lambda) &= \frac{1}{2\pi}\Big\{\widehat{\gamma}_n(0)(\pi+\lambda) + 2\sum_{k=1}^{n-1} \widehat{\gamma}_n(k)\frac{\sin(k\lambda)}{k}\Big\} \\
&= \underbrace{\frac{1}{2\pi}\Big\{\widehat{\gamma}_{n,\mu}(0)(\pi+\lambda) + 2\sum_{k=1}^{n-1} \widehat{\gamma}_{n,\mu}(k)\frac{\sin(k\lambda)}{k}\Big\}}_{=\widehat{F}_{n,\mu}(\lambda)} \quad + \quad \sum_{s,t=1}^{n} R_{st}^{(n)}(X_s - \mu)(X_t - \mu),
\end{aligned}$$

where it follows from $R_{st}^{(k,n)} = M_{st}^{(0,n)}(\pi+\lambda) + 2\sum_{k=1}^{n-1} M_{st}^{(k,n)} \sin(k\lambda)/k$ that $\max_{s,t}\big\{\big|R_{st}^{(n)}\big|\big\} = O\big(n^{-2}\ln(n)\big)$. It follows from Lemma 1.4.9 that

$$\begin{aligned}
&\operatorname{var}\Big(\sum_{s,t=1}^{n} R_{st}^{(n)}(X_s - \mu)(X_t - \mu)\Big) \\
&= \sum_{s,t,u,v=1}^{n} R_{st}^{(n)} R_{uv}^{(n)} \operatorname{cum}\big(X_s - \mu, X_t - \mu, X_u - \mu, X_v - \mu\big) + 2\sum_{s=1}^{n}\sum_{t,u,v=1}^{n} R_{st}^{(n)}\Sigma_{tu} R_{uv}^{(n)}\Sigma_{vs} \\
&= \ldots = O\big(n^{-2}(\ln(n))^2\big).
\end{aligned}$$

Since, according to (i), $\operatorname{var}\big(\widehat{F}_{n,\mu}(\lambda)\big) = O\big(n^{-1}\big)$ we obtain by the Cauchy-Schwarz inequality that

$$\operatorname{var}\big(\widehat{F}_{n,\bar{X}_n}(\lambda)\big) = \operatorname{var}\big(\widehat{F}_{n,\mu}(\lambda)\big) + O\big(n^{-3/2}\ln(n)\big).$$

□

We conclude from Lemmas 2.2.7 and 2.2.8 that the squared error risk of the estimators of the spectral distribution function tends to zero as $n \to \infty$. This is in sharp contrast to the behavior of $I_n(\lambda)$ and $I_{n,\bar{X}_n}(\lambda)$. The essential difference between estimators of the spectral distribution function and the periodogram can be found when we consider their respective representations as quadratic forms, $\sum_{s,t} A_{st}X_sX_t$. In case of the estimators of the spectral distribution function we have that $A_{st} = O\big(n^{-1}(1 \wedge 1/|s-t|)\big)$. In contrast, in case of $I_n(\lambda)$ and $I_{n,\bar{X}_n}(\lambda)$, we have that $A_{st} = O\big(n^{-1}\big)$ which provides some hint why the variance of the spectral density estimators does not vanish as $n \to \infty$.

In the following we introduce two methods to obtain consistent (sequences of) estimators of the spectral density. We assume that the process $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary with $EX_t = 0$ and that its autocovariance function $\gamma$ is absolutely summable. As we did for the estimator of the spectral distribution function, we also assume that $E[X_t^4] < \infty$ and that $\sum_{t,u,v=-\infty}^{\infty} \operatorname{cum}\big(X_0, X_t, X_u, X_v\big) < \infty$. The first method is motivated by the representation

$$f_X(\lambda) \;=\; \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k)\cos(k\lambda).$$

Since the autocovariance function is absolutely summable, we obtain that

$$\Big|\frac{1}{2\pi} \sum_{k:\ |k|>m} \gamma(k)\cos(k\lambda)\Big| \;\leq\; \frac{1}{2\pi} \sum_{k:\ |k|>m} \big|\gamma(k)\big| \underset{m\to\infty}{\longrightarrow} 0,$$

which suggests the so-called **lag window estimator** (since a "window" ensures that only a limited number of lags are included). This estimator is defined as

$$\widehat{f}_n(\lambda) \;:=\; \frac{1}{2\pi} \sum_{k:\ |k|\leq m_n} \widetilde{\gamma}_n(k)\cos(k\lambda) \qquad \forall \lambda \in [-\pi,\pi].$$

Of course, if $m_n \geq n-1$, then $\widehat{f}_n$ is equal to the periodogram $I_n$. To obtain consistency, the sequence $(m_n)_{n\in\mathbb{N}}$ has to fulfill two requirements: It is easy to see that $m_n \underset{n\to\infty}{\longrightarrow} \infty$ leads to a vanishing bias and, as we see below, $m_n/n \underset{n\to\infty}{\longrightarrow} 0$ yields a vanishing variance. Here are details of the corresponding calculations:

Since

$$E\widetilde{\gamma}_n(k) \;=\; \begin{cases} (1-|k|/n)\gamma(k), & \text{if } |k| \leq n-1, \\ 0, & \text{if } |k| > n-1 \end{cases}$$

we obtain that

$$E\widehat{f}_n(\lambda) \;=\; \frac{1}{2\pi} \sum_{k:\ |k|\leq m_n} \big(1-|k|/n\big)_+ \gamma(k)\cos(k\lambda).$$

Hence, we obtain by Lebesgue's theorem on dominated convergence that

$$f_X(\lambda) \;-\; E\widehat{f}_n(\lambda) \;=\; \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Big(1 \;-\; \underbrace{(1-|k|/n)_+ \mathbb{1}(|k|\leq m_n)}_{\underset{n\to\infty}{\longrightarrow}\ 1}\Big)\gamma(k)\cos(k\lambda) \underset{n\to\infty}{\longrightarrow} 0 \tag{2.2.7}$$

holds true if $m_n \underset{n\to\infty}{\longrightarrow} \infty$. To estimate the variance of $\widehat{f}_n(\lambda)$, we rewrite this estimator as a quadratic form,

$$\widehat{f}_n(\lambda) \;=\; \sum_{s,t=1}^{n} A_{st} X_s X_t,$$

where

$$A_{st} \;=\; \begin{cases} \frac{1}{2\pi n}\cos(\lambda(s-t)), & \text{if } |s-t| \leq m_n, \\ 0, & \text{if } |s-t| > m_n. \end{cases}$$

It follows from Lemma 1.4.9 that

$$\begin{aligned} \operatorname{var}\big(\widehat{f}_n(\lambda)\big) \;&=\; \sum_{s,t,u,v=1}^{n} A_{st}\, A_{uv} \operatorname{cum}(X_s, X_t, X_u, X_v) \;+\; 2\operatorname{tr}\big(A\Sigma A\Sigma\big) \\ &=:\; T_{n,1} \;+\; T_{n,2}, \end{aligned}$$

where $A = \big((A_{st})\big)_{s,t=1,\dots,n}$ and $\Sigma = \big((\gamma(s-t))\big)_{s,t=1,\dots,n}$ denotes the covariance matrix of the vector $(X_1,\dots,X_n)^T$. To estimate $|T_{n,1}|$ it suffices to use the fact that $|A_{st}| \le 1/(2\pi n)$. Therefore,

$$\big|T_{n,1}\big| \;\le\; \frac{1}{4\pi^2 n^2} \sum_{s=1}^{n} \underbrace{\sum_{t,u,v=1}^{n} \big|\operatorname{cum}(X_s,X_t,X_u,X_v)\big|}_{\le \sum_{t,u,v=-\infty}^{\infty} |\operatorname{cum}(X_0,X_t,X_u,X_v)| < \infty} \;=\; O\big(n^{-1}\big).$$

The term $|T_{n,2}|$ can be estimated as follows:

$$\begin{aligned}
\big|T_{n,2}\big| \;&\le\; 2 \sum_{s,t,u,v=1}^{n} \big|A_{st}\Sigma_{tu}A_{uv}\Sigma_{vs}\big| \\
&\le\; \sum_{s,t,u,v=1}^{n} \big(A_{st}^2 \,+\, A_{uv}^2\big)\, \big|\Sigma_{tu}\big|\, \big|\Sigma_{vs}\big| \\
&\le\; \max_{1\le s,t\le n} \Big\{ \sum_{u=1}^{n} \underbrace{\big|\Sigma_{tu}\big|}_{=|\gamma(t-u)|} \sum_{v=1}^{n} \underbrace{\big|\Sigma_{vs}\big|}_{=|\gamma(v-s)|} \Big\} \sum_{s,t=1}^{n} A_{st}^2 \\
&\quad +\; \max_{1\le u,v\le n} \Big\{ \sum_{s=1}^{n} \underbrace{\big|\Sigma_{vs}\big|}_{=|\gamma(v-s)|} \sum_{t=1}^{n} \underbrace{\big|\Sigma_{tu}\big|}_{=|\gamma(t-u)|} \Big\} \sum_{u,v=1}^{n} A_{uv}^2 \\
&\le\; 2 \Big( \sum_{k=-\infty}^{\infty} |\gamma(k)| \Big)^2 \sum_{s,t=1}^{n} A_{st}^2 .
\end{aligned}$$

Since $|A_{st}| \le 1/(2\pi n)$ and only the main diagonal and $2m_n$ minor diagonals of $A$ can contain nonzero entries we see that

$$\sum_{s,t=1}^{n} A_{st}^2 \;\le\; \frac{1}{(2\pi n)^2}\, n\, (2m_n + 1),$$

which implies

$$\big|T_{n,2}\big| \;=\; O\big(m_n/n\big).$$

Hence,

$$\operatorname{var}\big(\widehat{f}_n(\lambda)\big) \;=\; O\big(n^{-1}\big) \,+\, O\big(m_n/n\big) \;=\; O\big(m_n/n\big). \tag{2.2.8}$$

(2.2.7) and (2.2.8) imply that

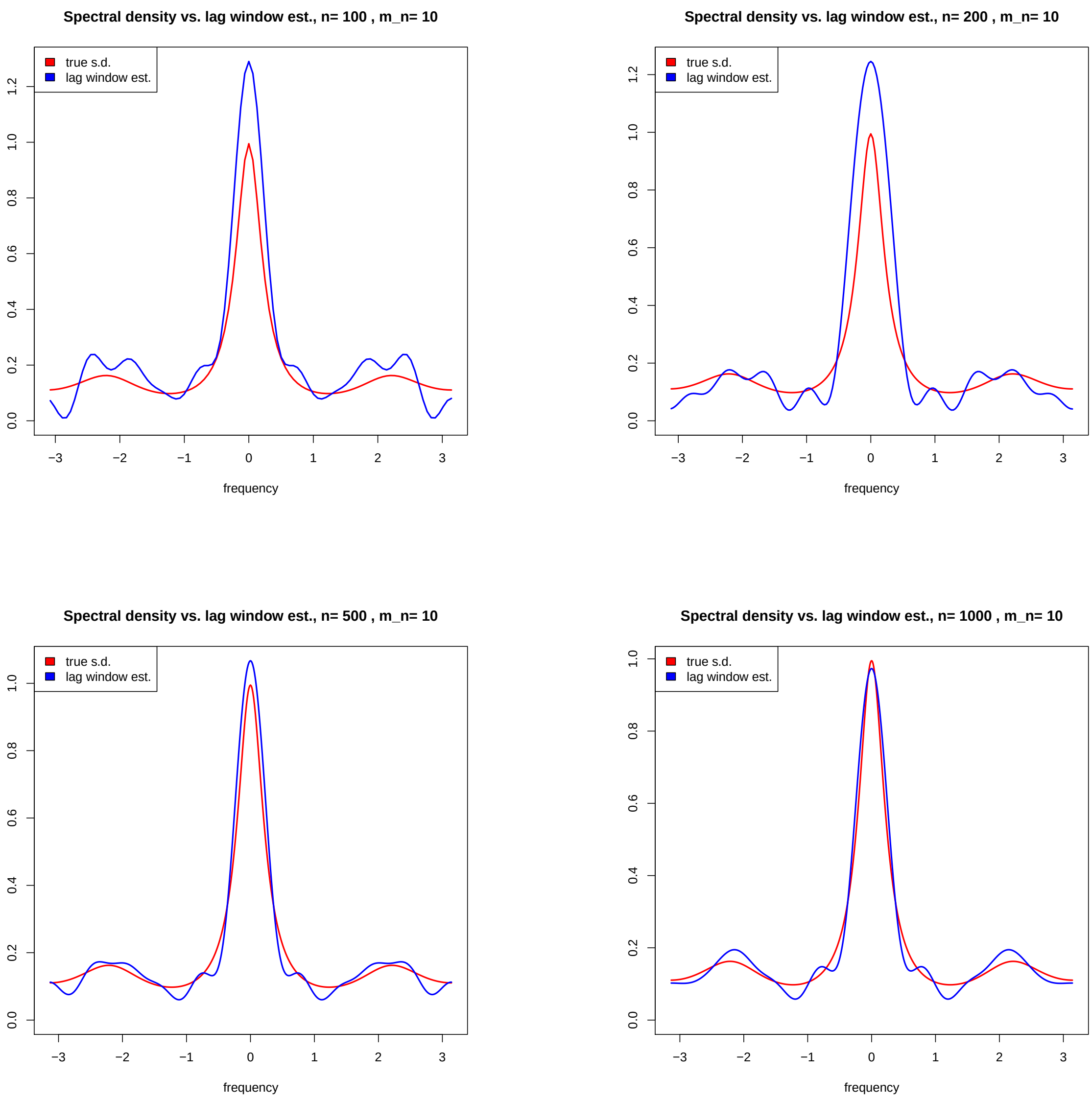
$$E\Big[\big(\widehat{f}_n(\lambda) \,-\, f_X(\lambda)\big)^2\Big] \;=\; \operatorname{var}\big(\widehat{f}_n(\lambda)\big) \,+\, \big(E\widehat{f}_n(\lambda) \,-\, f_X(\lambda)\big)^2 \underset{n\to\infty}{\longrightarrow} 0,$$

if the sequence $(m_n)_{n\in\mathbb{N}}$ of truncation parameters is chosen such that $m_n \underset{n\to\infty}{\longrightarrow} \infty$ and $m_n/n \underset{n\to\infty}{\longrightarrow} 0$. In this case, we obtain by the Markov inequality that

$$\widehat{f}_n(\lambda) \;\stackrel{P}{\longrightarrow}\; f_X(\lambda),$$

i.e., the sequence $\big(\widehat{f}_n(\lambda)\big)_{n\in\mathbb{N}}$ is **consistent** for $f_X(\lambda)$.

The following pictures show the true spectral density $f_X$ of a stationary AR($p$) process with parameters $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$ and independent innovations $\varepsilon_t \sim N(0,1)$ (red lines), together with one realization of the lag window estimator $\widehat{f}_n$ obtained from samples of size $n$ =100, 200, 500 and 1,000, respectively, (blue lines). Although the parameter $m_n$ was chosen independently of $n$ to be equal to 10, we see that the lag window estimator approximates $f_X$ quite well for samples of size 500 and 1,000. This is in sharp contrast to the periodogram, which fluctuates around the true spectral density regardless of the size of the sample.



A second method to obtain a consistent estimator of the spectral density is motivated by the fact that the values of the periodogram at different Fourier frequencies are asymptotically uncorrelated. (We have shown such a result for a white noise but this also holds true e.g. for linear processes.) On the other hand, the periodogram is an asymptotically unbiased estimator of the spectral density and the spectral density itself is a continuous function. Therefore, it seems promising to "smooth" the periodogram which is achieved in a particularly simple way by

$$\widehat{\widehat{f}}_n(\lambda) \;:=\; \frac{1}{2h_n} \int_{\lambda-h_n}^{\lambda+h_n} I_n(\omega)\, d\omega.$$

The parameter $h_n$ is called **"bandwidth"** and it controls the degree of smoothing. We will see that this estimator is asymptotically unbiased if $h_n \underset{n\to\infty}{\longrightarrow} 0$ and that its variance tends to zero if $nh_n \underset{n\to\infty}{\longrightarrow} \infty$.

Recall from (i) of Theorem 2.2.2 that

$$\sup_{\omega\in[-\pi,\pi]} \big|EI_n(\omega) \;-\; f_X(\omega)\big| \underset{n\to\infty}{\longrightarrow} 0.$$

Since $f_X$ is continuous we obtain that

$$\begin{aligned}
&\big|E\widehat{\widehat{f}}_n(\lambda) \;-\; f_X(\lambda)\big| \\
&= \;\frac{1}{2h_n}\Big| \int_{\lambda-h_n}^{\lambda+h_n} EI_n(\omega) \;-\; f_X(\lambda)\, d\omega \Big| \\
&\leq \;\frac{1}{2h_n} \int_{\lambda-h_n}^{\lambda+h_n} \big|f_X(\omega) \;-\; f_X(\lambda)\big|\, d\omega \;+\; \sup_{\omega\in[-\pi,\pi]} \big|EI_n(\omega) \;-\; f_X(\omega)\big| \underset{n\to\infty}{\longrightarrow} 0. \qquad (2.2.9)
\end{aligned}$$

To estimate the variance of $\widehat{\widehat{f}}_n(\lambda)$, we rewrite it as a quadratic form,

$$\widehat{\widehat{f}}_n(\lambda) \;=\; \sum_{s,t=1}^{n} A'_{st} X_s X_t,$$

where

$$A'_{st} \;=\; \frac{1}{2\pi n}\,\frac{1}{2h_n} \int_{\lambda-h_n}^{\lambda+h_n} \cos\big(\omega(s-t)\big)\, d\omega \;=\; \begin{cases} \frac{1}{2\pi n}, & \text{if } s=t, \\ \frac{1}{2\pi n}\;\frac{\sin((s-t)(\lambda+h_n))-\sin((s-t)(\lambda-h_n))}{2h_n\,(s-t)}, & \text{if } s\neq t. \end{cases}$$

It follows again from Lemma 1.4.9 that

$$\begin{aligned}
\operatorname{var}\Big(\widehat{\widehat{f}}_n(\lambda)\Big) \;&=\; \sum_{s,t,u,v=1}^{n} A'_{st}\, A'_{uv}\, \operatorname{cum}(X_s, X_t, X_u, X_v) \;+\; 2\operatorname{tr}\big(A'\Sigma A'\Sigma\big) \\
&=: \; T'_{n,1} \;+\; T'_{n,2},
\end{aligned}$$

where $A' = \big((A'_{st})\big)_{s,t=1,\ldots,n}$ and $\Sigma = \big((\gamma(s-t))\big)_{s,t=1,\ldots,n}$ denotes the covariance matrix of the vector $(X_1,\ldots,X_n)^T$. Since $|A'_{s,t}| \leq 1/(2\pi n)$ we obtain

$$\big|T'_{n,1}\big| \;\leq\; \frac{1}{4\pi^2 n^2} \sum_{s=1}^{n} \underbrace{\sum_{t,u,v=1}^{n} \big|\operatorname{cum}(X_s, X_t, X_u, X_v)\big|}_{\leq \sum_{t,u,v=-\infty}^{\infty} |\operatorname{cum}(X_0,X_t,X_u,X_v)| < \infty} \;=\; O\big(n^{-1}\big).$$

In order to estimate $|T'_{n,2}|$, we have to take a closer look at the $A'_{st}$. For $s \neq t$, we have that $\left|\frac{\sin((s-t)(\lambda+h_n))-\sin((s-t)(\lambda-h_n))}{2h_n\,(s-t)}\right| \leq \min\left\{1, 1/(h_n|s-t|)\right\}$, which implies that

$$\sum_{s,t=1}^{n} {A'_{s,t}}^2 \leq \frac{1}{4\pi^2 n}\Big(1 + \sum_{k:\ k\neq 0} \min\{1, 1/(h_n^2 k^2)\}\Big) = O\big(\frac{1}{nh_n}\big).$$

Therefore, we obtain in analogy to the computations above that

$$\left|T'_{n,2}\right| \leq 2\left(\sum_{k=-\infty}^{\infty} |\gamma(k)|\right)^2 \sum_{s,t=1}^{n} {A'_{st}}^2 = O\Big(\frac{1}{nh_n}\Big),$$

which implies that

$$\operatorname{var}\left(\widehat{\widehat{f}}_n(\lambda)\right) = O\big(n^{-1}\big) + O\Big(\frac{1}{nh_n}\Big) = O\Big(\frac{1}{nh_n}\Big). \qquad (2.2.10)$$

(2.2.9) and (2.2.10) imply that

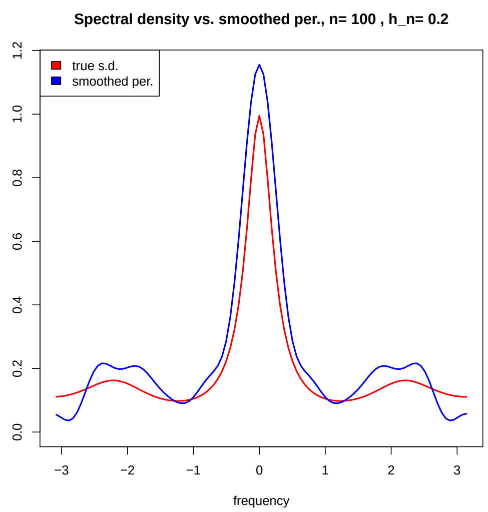
$$E\Big[\Big(\widehat{\widehat{f}}_n(\lambda) - f_X(\lambda)\Big)^2\Big] = \operatorname{var}\left(\widehat{f}_n(\lambda)\right) + \left(E\widehat{f}_n(\lambda) - f_X(\lambda)\right)^2 \underset{n\to\infty}{\longrightarrow} 0,$$

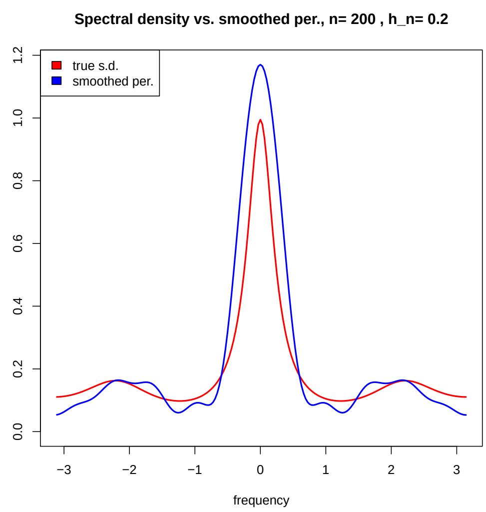
if the sequence $(h_n)_{n\in\mathbb{N}}$ of bandwidths is chosen such that $h_n \underset{n\to\infty}{\longrightarrow} 0$ and $nh_n \underset{n\to\infty}{\longrightarrow} \infty$. Then we obtain by the Markov inequality that

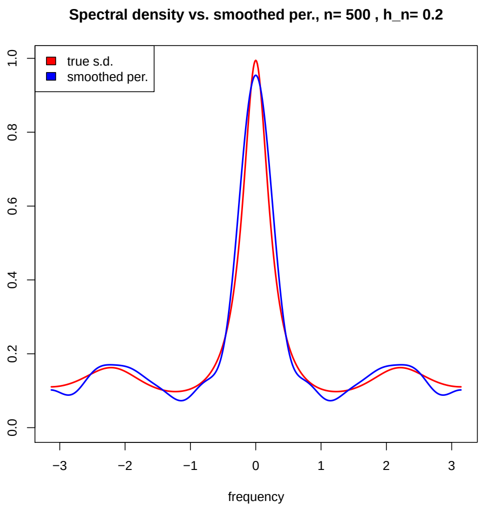
$$\widehat{\widehat{f}}_n(\lambda) \stackrel{P}{\longrightarrow} f_X(\lambda),$$

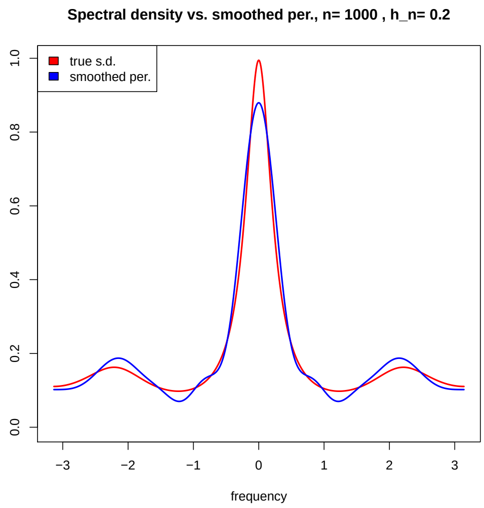
i.e., the sequence $\big(\widehat{\widehat{f}}_n(\lambda)\big)_{n\in\mathbb{N}}$ is **consistent**.

The following pictures show the true spectral density $f_X$ of a stationary AR($p$) process with parameters $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$ and independent innovations $\varepsilon_t \sim N(0,1)$ (red lines), together with one realization of the smoothed periodogram $\widehat{\widehat{f}}_n$ obtained from samples of size $n$ =100, 200, 500 and 1,000, respectively, (blue lines). The bandwidth $h_n$ was chosen independently of $n$ to be equal to 0.2. As in case of the lag window estimator, we obtain quite a good approximation of the spectral density $f_X$ for samples of size 500 and 1,000.






Although the lag window estimator $\widehat{f}_n(\lambda)$ and the smoothed periodogram $\widehat{\widehat{f}}_n(\lambda)$ are motivated by different considerations, they have an important feature in common. If we represent these estimators as quadratic forms, $\widehat{f}_n(\lambda) = \sum_{s,t=1}^n A_{st} X_s X_t$ and $\widehat{\widehat{f}}_n(\lambda) = \sum_{s,t=1}^n A'_{st} X_s X_t$, respectively, then we see that the coefficients $A_{st}$ and $A'_{st}$ are considerably smaller than their counterparts for the periodogram $(I_n(\lambda) = \sum_{s,t=1}^n \frac{\cos((s-t)\lambda)}{2\pi n} X_s X_t)$, as $|s-t|$ gets large. This is in fact the key for obtaining a vanishing variance by these two estimators as the sample size $n$ tend to infinity. On the other hand, the downweighting of $\widetilde{\gamma}_n(k)$ for large values of $|k|$ does not matter much since it follows from absolute summability of $\gamma$ that the contribution of the corresponding autocovariances to the spectral density decreases as $|k|$ gets large. Therefore, the estimators $\widehat{f}_n(\lambda)$ and $\widehat{\widehat{f}}_n(\lambda)$ are still asymptotically unbiased if the tuning parameters $m_n$ and $h_n$ are chosen within the range described above.

# 3 Solutions to the exercises

**Ex. 1.1.1** Show that $\mathcal{C}$ is an algebra but not a $\sigma$-algebra on $\mathbb{R}^\infty$.

**Solution**
We verify that $\mathcal{C}$ satisfies the axioms of an algebra:

a) $\mathbb{R}^\infty \in \mathcal{C}$ since $\underbrace{\mathbb{R}}_{\in \mathcal{B}} \times \mathbb{R}^\infty \in \mathcal{C}_1$.

b) Let $A \in \mathcal{C}$. Then $A = B \times \mathbb{R}^\infty$ for some $B \in \mathcal{B}^d$ and some $d \in \mathbb{N}$. Since $A^c = B^c \times \mathbb{R}^\infty$ and $B^c \in \mathcal{B}^d$ we obtain that $A^c \in \mathcal{C}$.

c) Let $A_1, A_2 \in \mathcal{C}$. Then $A_1 = B_1 \times \mathbb{R}^\infty$ and $A_2 = B_2 \times \mathbb{R}^\infty$, for some $B_1 \in \mathcal{B}^{d_1}$, $B_2 \in \mathcal{B}^{d_2}$, and some $d_1, d_2 \in \mathbb{N}$.

If $d_1 = d_2$, then

$$A_1 \cup A_2 \;=\; \underbrace{\big(B_1 \cup B_2\big)}_{\in \mathcal{B}^{d_1}} \times \mathbb{R}^\infty \;\in\; \mathcal{C}_{d_1}.$$

If $d_1 > d_2$, then $A_2 = \underbrace{B_2 \times \mathbb{R}^{d_2 - d_1}}_{\in \mathcal{B}^{d_1}} \times \mathbb{R}^\infty \in \mathcal{C}_{d_1}$ and

$$A_1 \cup A_2 \;=\; \underbrace{\big(B_1 \cup (B_2 \times \mathbb{R}^{d_1 - d_2})\big)}_{\in \mathcal{B}^{d_1}} \times \mathbb{R}^\infty \;\in\; \mathcal{C}_{d_1}.$$

If $d_1 < d_2$, then $A_1 \cup A_2 \in \mathcal{C}_{d_2}$ follows analogously.

It follows from a) to c) that $\mathcal{C}$ is an algebra on $\mathbb{R}^\infty$.

To disprove that $\mathcal{C}$ is not a $\sigma$-algebra, consider the sets $A_n := [0,1]^n \times \mathbb{R}^\infty$. Then $A_n \in \mathcal{C}$ for all $n \in \mathbb{N}$, however,

$$\bigcap_{n=1}^{\infty} A_n \;=\; [0,1] \times [0,1] \times \cdots \;=\; \big\{(x_1, x_2, \ldots)\colon\; x_t \in [0,1] \text{ for all } t\big\} \;\notin\; \mathcal{C}.$$

**Ex. 1.1.2** Show that, for $\mu \in \mathbb{R}$,

$$\Big\{x \in \mathbb{R}^\infty\colon\ \frac{1}{n}\sum_{t=1}^{n} x_t \underset{n\to\infty}{\longrightarrow} \mu\Big\} \in \sigma\big(\mathcal{C}\big).$$

**Solution**
Note that a sequence $(y_n)_{n\in\mathbb{N}}$ of real numbers converges to some $\mu \in \mathbb{R}$ if there exists for each $\epsilon > 0$ some sufficiently large $N = N(\epsilon) \in \mathbb{N}$ such that $|y_n - \mu| \le \epsilon$ holds for all $n \ge N$.
Therefore,

$$\Big\{x \in \mathbb{R}^\infty\colon\ \frac{1}{n}\sum_{t=1}^{n} x_t \underset{n\to\infty}{\longrightarrow} \mu\Big\} = \bigcap_{K=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\underbrace{\Big\{x \in \mathbb{R}^\infty\colon\ \Big|\frac{1}{n}\sum_{t=1}^{n} x_t - \mu\Big| \le \frac{1}{K}\Big\}}_{\in\,\mathcal{C}_n} \in \mathcal{C}.$$

**Ex. 1.1.3** Let $(X_t)_{t\in[0,\infty)}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ such that

- $X_0 = 0$ with probability 1,
- for $0 < t_1 < t_2 < \ldots < t_k$, $k \in \mathbb{N}$, the increments $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are stochastically independent.
- for $s < t$, $X_t - X_s \sim \mathcal{N}\big(0, t-s\big)$.

Find the finite-dimensional distributions $P^{X_{t_1},\ldots,X_{t_k}}$.

**Solution**

Let $0 \le t_1 < t_2 < \ldots < t_k$. Since the increments are independent and normally distributed we obtain that

$$\begin{pmatrix} X_{t_1} \\ X_{t_2} - X_{t_1} \\ \vdots \\ X_{t_k} - X_{t_{k-1}} \end{pmatrix} \sim \mathcal{N}_k\Big(0_k,\ \mathrm{Diag}\big(t_1, t_2 - t_1, \ldots, t_k - t_{k-1}\big)\Big),$$

where $0_k = \big(\underbrace{0, \ldots, 0}_{k\,times}\big)^T$. It follows that

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} X_{t_1} \\ X_{t_2} - X_{t_1} \\ \vdots \\ X_{t_k} - X_{t_{k-1}} \end{pmatrix} \sim \mathcal{N}_k\big(0_k,\ M\big),$$

where

$$M = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 - t_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & t_k - t_{k-1} \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \cdots & 1 & 1 \end{pmatrix}^T.$$

The elements of the matrix $M$ can also be computed as follows. Since, for $i < j$,

$$M_{i,j} = \mathrm{cov}\left(X_{t_i}, X_{t_j}\right) = \underbrace{\mathrm{cov}\left(X_{t_i}, X_{t_i}\right)}_{=t_i} + \underbrace{\mathrm{cov}\left(X_{t_i}, X_{t_j} - X_{t_i}\right)}_{=0}$$

we obtain, for $1 \le i, j \le k$,

$$M = \min\big\{t_i, t_j\big\}.$$

Finally, if $t_1, \ldots, t_k$ are arbitrary distinct and non-negative numbers,

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_k} \end{pmatrix} \sim \mathcal{N}_k\left(0_k, \begin{pmatrix} t_1 \wedge t_1 & \cdots & t_1 \wedge t_k \\ \cdots & \ddots & \cdots \\ t_k \wedge t_1 & \cdots & t_k \wedge t_k \end{pmatrix}\right).$$

**Ex. 1.2.1** Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, $E\varepsilon_0 = 0$, $E\varepsilon_0^2 =: \sigma_\varepsilon^2 < \infty$, and

$$X_t := \varepsilon_t + \beta\,\varepsilon_{t-1}.$$

Is the process $(X_t)_{t\in\mathbb{Z}}$ (weakly) stationary?

**Solution**

We have that

$$EX_t = E\big[\varepsilon_t + \beta\,\varepsilon_{t-1}\big] = E\varepsilon_t + \beta\,E\varepsilon_{t-1} = 0$$

and

$$\begin{aligned}\operatorname{cov}\big(X_{t+k}, X_t\big) &= \operatorname{cov}\big(\varepsilon_{t+k} + \beta\,\varepsilon_{t+k-1},\ \varepsilon_t + \beta\,\varepsilon_{t-1}\big)\\ &= \begin{cases}(1+\beta^2)\,\sigma_\varepsilon^2, & \text{if } k = 0,\\ \beta\,\sigma_\varepsilon^2, & \text{if } k = \pm 1,\\ 0, & \text{if } |k| > 1\end{cases}\end{aligned}$$

holds for all $t \in \mathbb{Z}$. Therefore, the process $(X_t)_{t\in\mathbb{Z}}$ is (weakly) stationary?

**Ex. 1.2.2** Let $(\beta_k)_{k\in\mathbb{Z}}$ be a sequence of real numbers with $\sum_{k=-\infty}^{\infty}\beta_k^2 < \infty$. The function $\gamma\colon\ \mathbb{Z}\to\mathbb{R}$ is defined by $\gamma(k)=\sum_{j=-\infty}^{\infty}\beta_{j+k}\beta_j$.

Is $\gamma$ an autocovariance function?

**Solution**

It follows from Theorem 1.2.5 that it suffices to show that

- $\gamma$ is real-valued ($\neq \pm\infty$) and $\gamma(k)=\gamma(-k)\ \forall k\in\mathbb{Z}$,
- $\sum_{i,j=1}^{n} a_i\gamma(t_i-t_j)a_j \geq 0 \quad \forall a_1,\dots,a_n\in\mathbb{R},\ \forall t_1,\dots,t_n\in\mathbb{Z},\ \forall n\in\mathbb{N}$.

First of all, we obtain that

$$\sum_{j=-\infty}^{\infty}\left|\beta_{j+k}\right|\left|\beta_j\right| \leq \sqrt{\sum_{j=-\infty}^{\infty}\beta_{j+k}^2}\sqrt{\sum_{j=-\infty}^{\infty}\beta_j^2} < \infty.$$

Hence, the infinite series $\sum_{j=-\infty}^{\infty}\beta_{j+k}\beta_j$ is absolutely convergent and its sum therefore is finite. Moreover, it is obvious that

$$\gamma(k) = \sum_{j=-\infty}^{\infty}\beta_{j+k}\,\beta_j = \sum_{j=-\infty}^{\infty}\beta_j\,\beta_{j-k} = \gamma(-k) \qquad \forall k\in\mathbb{Z},$$

i.e., $\gamma$ is indeed a real-valued and even function.

To prove that $\gamma$ is non-negative definite, suppose that $n\in\mathbb{N}$, $a_1,\dots,a_n\in\mathbb{R}$, and $t_1,\dots,t_n\in\mathbb{Z}$ are arbitrary. Then

$$\sum_{i,j=1}^{n} a_i\,\gamma(t_i-t_j)\,a_j = \sum_{i,j=1}^{n} a_i\Big(\sum_{k\in\mathbb{Z}}\beta_{k+t_i-t_j}\,\beta_k\Big)\,a_j.$$

Since the infinite series $\sum_{k\in\mathbb{Z}}\beta_{k+t_i-t_j}\,\beta_k$ is absolutely summable we conclude that above triple sum is also absolutely convergent and any change of the order of summation does not change the value of the sum. Therefore, we obtain

$$\begin{aligned}\sum_{i,j=1}^{n} a_i\,\gamma(t_i-t_j)\,a_j &= \sum_{i,j=1}^{n} a_i\Big(\sum_{k\in\mathbb{Z}}\beta_{k+t_i}\,\beta_{k+t_j}\Big)\,a_j \\ &= \sum_{k\in\mathbb{Z}}\Big(\sum_{i=1}^{n} a_i\beta_{k+t_i}\Big)\Big(\sum_{j=1}^{n} a_j\beta_{k+t_j}\Big) \\ &= \sum_{k\in\mathbb{Z}}\Big(\sum_{i=1}^{n} a_i\beta_{k+t_i}\Big)^2 \geq 0.\end{aligned}$$

To summarize, $\gamma$ is an even, real-valued and non-negative definite function. It follows from Theorem 1.2.5 that there exists a stochastic process with $\gamma$ as its autocovariance function.

**Ex. 1.3.1** Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, P)$ and $(\beta_k)_{k\in\mathbb{Z}}$ be a sequence of real numbers. Assume that $E\varepsilon_t = 0$, $\sigma_\varepsilon^2 := E\varepsilon_t^2 < \infty$, and $\sum_{k=-\infty}^{\infty} \beta_k^2 < \infty$.

(i) Show that $\big(X_{t,m}\big)_{m\in\mathbb{N}}$ defined by

$$X_{t,m} := \sum_{k=-m}^{m} \beta_k \varepsilon_{t-k}$$

is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$.

(ii) Let $X_t$ be the $L^2$-limit of $\big(X_{t,m}\big)_{m\in\mathbb{N}}$.
Compute $\operatorname{cov}(X_{t+k}, X_t)$.

(iii) Let $\pi\colon \mathbb{Z} \to \mathbb{Z}$ be a bijective function. Then, for each $t \in \mathbb{Z}$, $\big(\beta_{\pi(k)}\varepsilon_{t-\pi(k)}\big)_{k\in\mathbb{Z}}$ is a rearrangement of the sequence $\big(\beta_k\varepsilon_{t-k}\big)_{t\in\mathbb{Z}}$.

a) Show that $\big(\widetilde{X}_{t,m}\big)_{m\in\mathbb{N}}$ defined by

$$\widetilde{X}_{t,m} := \sum_{k=-m}^{m} \beta_{\pi(k)} \varepsilon_{t-\pi(k)}$$

is also a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$.

b) Show that

$$\big\|X_{t,m} - \widetilde{X}_{t,m}\big\| \underset{m\to\infty}{\longrightarrow} 0$$

and conclude that

$$P\big(X_t = \widetilde{X}_t\big) = 1,$$

where $\widetilde{X}_t$ denotes the $L^2$-limit of $\big(\widetilde{X}_{t,m}\big)_{m\in\mathbb{N}}$.

## Solution

(i) Let, w.l.o.g., $m < n$. Then

$$\begin{aligned}
\big\|X_{t,n} - X_{t,m}\big\|^2 &= \big\langle X_{t,n} - X_{t,m},\, X_{t,n} - X_{t,m}\big\rangle \\
&= \Big\langle \sum_{j:\; m<|j|\le n} \beta_j\, \varepsilon_{t-j},\, \sum_{k:\; m<|k|\le n} \beta_k\, \varepsilon_{t-k}\Big\rangle \\
&= \sum_{j,k:\; m<|j|,|k|\le n} \beta_j\beta_k \underbrace{\langle \varepsilon_{t-j},\, \varepsilon_{t-k}\rangle}_{=\,\delta_{j,k}\,\sigma_\varepsilon^2} \\
&= \sum_{k:\; m<|k|\le n} \sigma_\varepsilon^2\, \beta_k^2.
\end{aligned}$$

Since the sequence $(\beta_k)_{k\in\mathbb{Z}}$ is absolutely summable we see that the right-hand side of the above display can be made arbitrarily small if $m$ and $n$ are sufficiently large. Hence, $(X_{t,m})_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$.

(ii) Since $EX_{t,m} = 0$ for all $t, m$ and $E|X_{t,m} - X_t| \leq \|X_{t,m} - X_t\| \to_{m\to\infty} 0$ we conclude that $EX_t = 0$ for all $t \in \mathbb{Z}$. Now we obtain by continuity of the inner product (see (ii) of Lemma 1.3.4) that

$$\begin{aligned}
\operatorname{cov}\big(X_{t+k}, X_t\big) &= \big\langle X_{t+k},\, X_t \big\rangle \\
&= \lim_{m\to\infty} \big\langle X_{t+k,m},\, X_{t,m} \big\rangle \\
&= \lim_{m\to\infty} \Big\langle \sum_{j=-m}^{m} \beta_j \varepsilon_{t+k-j},\, \sum_{l=-m}^{m} \beta_l \varepsilon_{t-l} \Big\rangle \\
&= \lim_{m\to\infty} \sum_{j,l=-m}^{m} \beta_j \beta_l \underbrace{\big\langle \varepsilon_{t+k-j},\, \varepsilon_{t-l} \big\rangle}_{=\,\delta_{j-k,l}\, \sigma_\varepsilon^2} \\
&= \sigma_\varepsilon^2 \sum_{j\in\mathbb{Z}} \beta_j \beta_{j-k}.
\end{aligned}$$

(iii) a) Let $\epsilon > 0$. Since $\sum_{k\in\mathbb{Z}} \beta_{\pi(k)}^2 = \sum_{k\in\mathbb{Z}} \beta_k^2 < \infty$ we obtain for $m < n$, as in the proof of part (i),

$$\begin{aligned}
\big\|\widetilde{X}_{t,m} - \widetilde{X}_{t,n}\big\|^2 &= \Big\| \sum_{k:\ m<|k|\leq n} \beta_{\pi(k)} \varepsilon_{t-\pi(k)} \Big\|^2 \\
&= \sigma_\varepsilon^2 \sum_{k:\ m<|k|\leq n} \beta_{\pi(k)}^2 \;\leq\; \epsilon \qquad \text{if } m \geq N(\epsilon),
\end{aligned}$$

for some sufficiently large $N(\epsilon)$.

b) Let $\epsilon > 0$. It follows from $\sum_{k\in\mathbb{Z}} \beta_k^2 < \infty$ that there exists some $K_\epsilon \in \mathbb{N}$ such that $\sum_{k:\ |k|>K_\epsilon} \beta_k^2 < \epsilon$. If $m$ is such that $\{-K_\epsilon, \ldots, K_\epsilon\} \subseteq \{\pi(-m), \ldots, \pi(m)\}$, then

$$\begin{aligned}
\big\|X_{t,m} - \widetilde{X}_{t,m}\big\|^2 &= \Big\| \sum_{k:\ |k|\leq m} \beta_k \varepsilon_{t-k} - \beta_{\pi(k)} \varepsilon_{t-\pi(k)} \Big\|^2 \\
&= \sigma_\varepsilon^2 \sum_{k\in\{-m,\ldots,m\}\setminus\{\pi(-m),\ldots,\pi(m)\}} \beta_k^2 \\
&\quad + \sigma_\varepsilon^2 \sum_{k\in\{\pi(-m),\ldots,\pi(m)\}\setminus\{-m,\ldots,m\}} \beta_{\pi(k)}^2 \\
&\leq 2\,\sigma_\varepsilon^2\,\epsilon.
\end{aligned}$$

We obtain by the triangle inequality that

$$\big\|X_t - \widetilde{X}_t\big\| \leq \big\|X_t - X_{t,m}\big\| + \big\|X_{t,m} - \widetilde{X}_{t,m}\big\| + \big\|\widetilde{X}_{t,m} - \widetilde{X}_t\big\| \underset{m\to\infty}{\longrightarrow} 0,$$

which implies that $\big\|X_t - \widetilde{X}_t\big\| = 0$, and so

$$P\big(X_t = \widetilde{X}_t\big) = 1.$$

**Ex. 1.4.1** Suppose that $Y$ and $Z$ are uncorrelated random variables with $EY = EZ = 0$ and $EY^2 = EZ^2 = 1$. For $t \in \mathbb{N}$, let $X_t = Y\cos(\theta t) + Z\sin(\theta t)$, where $\theta \in \mathbb{R}$.

Show that $\widehat{X}_3 = 2\cos(\theta)X_2 - X_1$ is the best linear predictor of $X_3$ given $X_1$, $X_2$.

*Hint:* $E[X_s X_t] = \cos(\theta(s-t))$ *and* $\cos(2\theta) = (\cos(\theta))^2 - (\sin(\theta))^2$.

**Solution**

According to (ii) of Theorem 1.3.6, it suffices to show that

$$\langle X_3 - \widehat{X}_3, X_k\rangle = 0 \qquad \text{for } k = 1, 2.$$

We can easily verify this:

$$\begin{aligned}\langle X_3 - 2\cos(\theta)X_2 + X_1, X_1\rangle &= \cos(2\theta) - 2\big(\cos(\theta)\big)^2 + 1 \\ &= -\big(\sin(\theta)\big)^2 - \big(\cos(\theta)\big)^2 + 1 = 0\end{aligned}$$

and

$$\langle X_3 - 2\cos(\theta)X_2 + X_1, X_2\rangle = \cos(\theta) - 2\cos(\theta) + \cos(-\theta) = 0.$$

**Ex. 1.4.2** Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma_\varepsilon^2)$ and $X_t = \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}$, for some $\alpha \in \mathbb{R}$, $|\alpha| < 1$.
Show that $\widehat{X}_{n+1} := \alpha X_n$ is the best linear predictor of $X_{n+1}$ given $X_1, \ldots, X_n$.
*Hint: Argue that $X_{n+1} - \alpha X_n = \varepsilon_{n+1}$ and use (ii) of Lemma 1.3.4.*

**Solution**
Since $\sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}$ converges absolutely we obtain that

$$\begin{aligned} X_{n+1} - \alpha X_n &= \sum_{k=0}^{\infty} \alpha^k \varepsilon_{n+1-k} - \alpha \cdot \sum_{k=0}^{\infty} \alpha^k \varepsilon_{n-k} \\ &= \varepsilon_{n+1} + \underbrace{\left( \sum_{k=1}^{\infty} \alpha^k \varepsilon_{n+1-k} - \sum_{k=0}^{\infty} \alpha^{k+1} \varepsilon_{n-k} \right)}_{=0}. \end{aligned}$$

According to Theorem 1.3.6, we have to show that

$$\langle X_{n+1} - \alpha X_n,\, X_k \rangle = 0 \qquad \forall k = 1, \ldots, n.$$

In order to use linearity of the inner product, we first replace $X_k$ by its truncated counterpart $X_{k,m} = \sum_{l=0}^{m} \alpha^l \varepsilon_{k-l}$. We obtain

$$\begin{aligned} \langle X_{n+1} - \alpha X_n,\, X_{k,m} \rangle &= \langle \varepsilon_{n+1},\, \sum_{l=0}^{m} \alpha^l \varepsilon_{k-l} \rangle \\ &= \sum_{l=0}^{m} \alpha^l \underbrace{\langle \varepsilon_{n+1},\, \varepsilon_{k-l} \rangle}_{=0} = 0. \end{aligned}$$

Finally, we obtain by continuity of the inner product ((ii) of Lemma 1.3.4) that

$$\langle X_{n+1} - \alpha X_n,\, X_k \rangle = \lim_{m\to\infty} \langle X_{n+1} - \alpha X_n,\, X_{k,m} \rangle = 0 \qquad \forall k = 1, \ldots, n.$$