Aus dem Institut für Mathematik
der Universität zu Lübeck
Direktor: Prof. Dr. Jürgen Prestin

Estimation of Ordinal Pattern Probabilities in Stochastic Processes

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck
- Aus der Technisch-Naturwissenschaftlichen Fakultät -

vorgelegt von
Mathieu Sinn
aus Lüneburg
Lübeck 2009

1. Berichterstatter: PD Dr. Karsten Keller

2. Berichterstatter: Prof. Dr. Christoph Bandt

Prüfungsausschussvorsitzender: Prof. Dr. Alfred Mertins

Tag der mündlichen Prüfung: 21. Dezember 2009

Zusammenfassung

Diese Arbeit untersucht die Verteilung ordinaler Muster in stochastischen Prozessen und die Schätzung der Auftretenswahrscheinlichkeit ordinaler Muster. Ordinale Muster beschreiben die Ordnungsrelationen zwischen einer festen Anzahl von Werten einer Zeitreihe. Sind die Werte der Zeitreihe paarweise verschieden, so können ordinale Muster durch Permutationen dargestellt werden. Die Verteilung ordinaler Muster in einer Zeitreihe (bzw. in Teilen der Zeitreihe) dient dazu, Charakteristiken der zu Grunde liegenden Dynamik zu berechnen, oder zwischen der Dynamik in unterschiedlichen Teilen der Zeitreihe zu unterscheiden. Ein Beispiel solch einer Charakteristik ist die Permutationsentropie, die durch die Shannon-Entropie der Verteilung ordinaler Muster gegeben ist und als Maß für die Komplexität von Zeitreihen angesehen werden kann. Eine Anwendung der Permutationsentropie ist die Analyse epileptischer Aktivität in EEG Zeitreihen.

Der Kontext der Untersuchungen dieser Arbeit ist eine parametrische Familie stochastischer Prozesse mit stationären, nicht-degenerierten, zentrierten Gauss'schen Zuwächsen. Diese Prozessklasse beinhaltet equidistante Diskretisierungen Fraktaler Brown'scher Bewegung sowie integrierte ARFIMA(0,d,0) und AR(1) Prozesse. In Kapitel 3 zeigen wir, dass die Verteilung ordinaler Muster in solchen Prozessen stationär ist, und dass jedes ordinale Muster eine strikt positive Auftretenswahrscheinlichkeit hat. Ist eine endliche Anzahl von Beobachtungen ordinaler Muster gegeben, so ist die relative Häufigkeit eines Musters ein unverzerrter Schätzer für die entsprechende Auftretenswahrscheinlichkeit. Da die Verteilung stationärer und zentrierter Gauss'scher Prozesse invariant ist bezüglich einer Umkehrung der Raum- bzw. Zeitachse, haben bestimmte ordinale Muster dieselbe Auftretenswahrscheinlichkeit. Indem man die relativen Häufigkeiten dieser Muster mittelt, erhält man Schätzer mit niedrigerer Varianz.

Eine hinreichende Bedingung für schwache Konsistenz der Schätzer ist, dass die Autokovarianzen des Zuwachsprozesses für wachsende Zeitabstände gegen null gehen. Wie wir zeigen, ist diese Bedingung auch hinreichend für starke Konsistenz. Hinreichend für asymptotische Normalität ist, dass die Autokovarianzen des Zuwachsprozesses schneller abklingen als $k \mapsto \frac{1}{\sqrt{k}}$. Diese Aussage gilt allgemeiner auch für bestimmte differenzierbare Funktionen der Auftretenswahrscheinlichkeiten, sowie im mehrdimensionalen Fall, wenn die Auftretenswahrscheinlichkeiten verschiedener Muster gleichzeitig geschätzt werden.

In Kapitel 4 untersuchen wir die Kovarianzen von Nulldurchläufen in nicht-degenerierten, stationären, zentrierten Gauss'schen Prozessen. Kern der Untersuchungen ist die Analyse vierdimensionaler Gauss'scher Orthant-Wahrscheinlichkeiten sowie ihrer Ableitungen bezüglich bestimmter Korrelationskoeffizienten. Wie wir zeigen, lassen sich die Kovarianzen von Nulldurchläufen als Summen eindimensionaler Integrale darstellen, die numerisch mit beliebiger Genauigkeit ausgewertet werden können. Wir bestimmen außerdem das asymptotische Verhalten der Kovarianzen und geben untere und obere Schranken sowie Approximationen an. Auf Grundlage dieser Ergebnisse leiten wir Eigenschaften der Varianz empirischer Nulldurchlaufsraten her.

In Kapitel 5 betrachten wir als Spezialfall ordinale Muster, die die Ordnungsrelationen zwischen genau drei aufeinanderfolgenden Werten einer Zeitreihe beschreiben. Wie wir zeigen, läßt sich in diesem Fall jeder "vernünftige" Schätzer von Auftretenswahrscheinlichkeiten ordinaler Muster als affine Funktion der empirischen Nulldurchlaufsrate im Zuwachsprozess darstellen. Auf Grundlage der Ergebnisse aus Kapitel 4 berechnen wir die Varianz der Schätzer in equidistanten Diskretisierungen Fraktaler Brown'scher Bewegung, sowie in integrierten ARFIMA(0,d,0) und AR(1) Prozessen.

Stehen die Parameter der stochastischen Prozesse in einer monotonen Beziehung zur Wahrscheinlichkeit eines Nullduchlaufes, so erhält man Schätzer für diese Parameter, indem man die empirische Nulldurchlaufsrate in die inverse monotone Beziehung einsetzt. Mittels der Ergebnisse aus Kapitel 3 leiten wir Eigenschaften dieser Schätzer her. Unter zusätzlichen Anforderungen an die Autokovarianzen des Zuwachsprozesses bestimmen wir Konfidenzintervalle für die Prozessparameter. Wir illustrieren die Methoden für die Schätzung des Hurst Parameters in Fraktaler Brown'scher Bewegung, des fraktalen Differenzierungsparameters in ARFIMA(0,d,0) Prozessen, sowie des autoregressiven Koeffizienten in AR(1) Prozessen.

In einer Simulationsstudie untersuchen wir die Güte der Schätzer sowie die Überdeckung der Prozessparameter durch die Konfidenzintervalle. Im Fall des Hurst Parameters vergleichen wir die Güte mit der eines alternativen Schätzers. Weiterhin betrachten wir die Verteilung der empirischen Nulldurchlaufsrate in den Zuwächsen Fraktaler Brown'scher Bewegung. Wie sich herausstellt, ist die Verteilung für große Werte des Hurst Parameters äußerst unregelmäßig.

In Kapitel 6 betrachten wir ordinale Muster, die die Ordnungsrelationen zwischen Werten zu beliebigen Zeitpunkten beschreiben (an Stelle von unmittelbar aufeinanderfolgenden Werten). Wir zeigen, wie ordinale Muster auf großen Zeitskalen verwendet werden können, um den Index asymptotisch selbstähnlicher Prozesse zu schätzen. Eine Anwendung ist die Schätzung des Hurst Parameters in Fraktaler Brown'scher Bewegung, die von schwach korreliertem "Rauschen" überlagert ist. Wir illustrieren die Anwendung dieser Methode für zwei empirische Zeitreihen, nämlich, Pegelstände des Nils sowie Präzisionsmessungen des amerikanischen NBS Institutes.

Abstract

This thesis studies the distribution of ordinal patterns in stochastic process and the estimation of occurrence probabilities of ordinal patterns. Ordinal patterns represent the order relations among a fixed number of values in a time series. Under the assumption that the values of the time series are pairwise different, it is natural to identify ordinal patterns with permutations. The distribution of ordinal patterns in a time series (or parts of it) can be used to compute characteristics of the underlying dynamics, or to discriminate between the dynamics in different parts of the time series. For instance, permutation entropy (which is the Shannon entropy of ordinal pattern distribution) has been proposed as a measure for the complexity of time series. Permutation entropy measurements have been applied, for example, to the analysis of epileptic acivity in EEG data.

The framework of our analysis is a parametric family of stochastic processes with stationary, non-degenerate and zero-mean Gaussian increments. This class of processes includes, e.g., equidistant discretizations of Fractional Brownian Motion, and processes where the increments are ARFIMA(0,d,0) or AR(1) processes. In Chapter 3, we show that the distribution of ordinal patterns in such processes is stationary, and each pattern has a strictly positive probability of occurrence. Given a finite number of observations, the relative frequency of an ordinal pattern is an unbiased estimator of the corresponding occurrence probability. By the fact that the distribution of stationary zero-mean Gaussian processes is invariant with respect to reversions of the time and space orientation, certain ordinal patterns have the same probability. We show that averaging the relative frequencies of these patterns yields unbiased estimators with smaller variance.

A sufficient condition for the estimators of ordinal pattern probabilities to be consistent is that the autocovariances of the increment process tend to zero. We show that this condition is also sufficient for strong consistency. A sufficient condition for asymptotic normality of the estimators is that the autocovariances of the increment process decay faster than $k \mapsto \frac{1}{\sqrt{k}}$. More generally, this statement is true for certain differentiable functions of ordinal pattern probabilities and also in the multidimensional case when the probabilities of several patterns are jointly estimated.

In Chapter 4, we study covariances of zero crossings in non-degenerate and stationary zero-mean Gaussian processes. The results are obtained by analyzing four-dimensional

normal orthant probabilities and their derivatives with respect to correlation coefficients. We propose a representation of the zero crossing covariances by one-dimensional integrals which can be numerically evaluated using standard quadrature rules. Furthermore, we derive asymptotics of the covariances and establish approximations and bounds. Based on these results, we derive properties of the variance of empirical zero crossing rates.

In Chapter 5, we focus on ordinal patterns describing the order relations among three subsequent values of a time series. In this case, any "reasonable" estimator of ordinal pattern probabilities can be expressed as an affine function of the empirical zero crossing rate in the increment process. Using the results of Chapter 4, we evaluate the variance of the estimators in equidistant discretizations of Fractional Brownian Motion and in processes where the increments are ARFIMA(0,d,0) and AR(1), respectively.

When the parameters of the family of stochastic processes are real numbers and monotonically related to the probability of a zero crossing, an estimator of the parameters is obtained by plugging the empirical zero crossing rate into the inverse of the monotonic relation. Using the results of Chapter 3, we establish properties of this estimator. Under additional conditions on the autocovariances of the increment process, we also derive confidence intervals. We show how the results apply to the estimation of the Hurst parameter in Fractional Brownian Motion, of the fractional differencing parameter in ARFIMA(0,d,0) processes and of the autoregressive coefficient in AR(1) processes.

In a simulation study, we evaluate the performance of the estimators and the coverage of the parameters by the confidence intervals. For the Hurst parameter, we compare the performance to that of an alternative estimator. We also consider the distribution of empirical zero crossing rates in the increment process of FBM. It turns out that the distribution is very irregular for large values of the Hurst parameter.

Chapter 6 generalizes the results of the previous chapters to ordinal patterns describing the order relations among values at arbitrary times instead of immediately subsequent values. We demonstrate how ordinal patterns on increasing time scales can be used for estimating the index of asymptotically self-similar processes. One application is the estimation of the Hurst parameter in equidistant discretizations of Fractional Brownian Motion superimposed with short range dependent "noise". We illustrate our method for two practical time series, namely, River Nile data and NBS precision measurements.

Acknowledgements

I would like to express my profound gratitude to my supervisor, PD Dr. Karsten Keller. Without his continuous guidance, support and encouragement, this thesis would have never been accomplished. Ever since I started working with him, I have been very much impressed by his gentle and modest personality. His sincerity and enthusiasm towards mathematical science will always serve as a lightning example to me.

I would also like to thank Prof. Dr. Lutz Mattner for his help and useful suggestions throughout this research work. It was his excellent lectures in probability theory which sparked my interest in mathematical statistics at the first place, and many fruitful discussions with him greatly enhanced my knowledge in this subject. From his instructions and criticism, this thesis has benefited significantly.

My gratitude also extends to my secondary supervisor, Prof. Dr. Rolf Verleger, and to my mentor, Prof. Dr. Bernd Fischer. I would like to thank Prof. Dr. Heinz Lauffer from the Department of Pediatric Medicine at the University of Greifswald and Dr. Björn Rasch from the Department of Neuroendocrinology at the University of Lübeck for providing and explaining the data behind Figures 1.2 and 1.3. I am grateful to all the staff at the Institute for Mathematics and at the Graduate School for Computing in Medicine and Life Sciences at the University of Lübeck for all the help and support, particularly, to Jens Keiner and Dr. Peter Dencker for proofreading parts of the thesis and for many useful suggestions.

Finally, special thanks to my family and friends. Particularly, I want to thank my fiancée Bei Chen, who also helped me proofreading, and my parents Volker and Marguerite Sinn, to whom I dedicate this work.

Contents

Chapter 1

Introduction

Modern science progresses through the collection of data and the development of new methods for their analysis. Often data are obtained by sequential measurements of a quantity in time, so-called *time series*. Examples are manifold: daily temperatures, river gauges, currency exchange rates or the electric activity of the human brain. With the advances in computer technology which make it possible to store gigabytes of data, one of the main challenges of time series analysis these days is the computational complexity due to the large amount of data.

Major approaches to time series analysis are the theory of *dynamical systems* and the theory of *stochastic processes*. A dynamical system is a deterministic model for the generation of time series. Thus, if the dynamics and all initial states are known, any future value can be predicted. A stochastic process includes a random component which makes it impossible to exactly forecast future values.

Since the 1980's, *non-linear dynamical systems* have been receiving particular attention. One reason is that linear dynamical systems are unable to capture important characteristics of many time series observed in nature, such as aperiodic behaviour and high sensitivity to initial states. Another reason is that, nowadays, computers allow to study properties of non-linear dynamics, for instance, by means of simulations (see Galka [42]). Although non-linear dynamical systems are deterministic, even an approximate prediction of future values can be impossible when the initial states are not known exactly. Therefore, an essential concept in the study of non-linear dynamics is *complexity* which, roughly speaking, quantifies the unpredictability of a dynamical system. For instance, the complexity of linear systems is low, while "chaotic" non-linear systems have high complexity.

The complexity of dynamical systems can be measured by *entropies*, *dimensions* and *Lyapunov exponents* (see Walters [99], Grassberger and Procaccia [44], Galka [42]). While these quantities are well-motivated from the theoretical viewpoint, it is often difficult to

estimate them in practice. In particular, the estimates depend on the specific choice of a sequence of families of sets over the state space, so there is no standardized way for computing them. Moreover, the computation of estimates is time-consuming, and when the complexity of the system is high, the estimates themselves are very sensitive to observational noise.

Permutation entropy. An interesting alternative for measuring the complexity of a one-dimensional system is the permutation entropy introduced by Bandt and Pompe [10]. The basic idea is to partition the state space of a system in such a way that two points lie in the same part of the state space if and only if their first d iterates are correspondingly ordered. The *permutation entropy of order* d is given by the Shannon entropy of this partition. The *permutation entropy rate* is defined as the upper limit of the permutation entropy of order d divided by d as d tends to infinity.

Surprisingly at first glance, there are fundamental relations between the permutation entropy rate and complexity measures which take the whole metrical or topological information of time series into account. For piecewise monotone interval maps, Bandt et al. [9] shows coincidence with the Kolmogorov-Sinai entropy, and a similar result holds for the topological entropy (see Bandt et al. [9], Misiurewicz [75]). Amigó et al. [3, 4, 5] generalize the result for the Kolmogorov-Sinai entropy to the multi-dimensional case, however, based on a concept of permutation entropy which uses non-standardized partitions of the state space. For ergodic systems, Keller and Sinn [61, 62] show that the permutation entropy rate is always an upper bound for Kolmogorov-Sinai entropy.

Partitioning the state space of a system (or the space of a fixed number of consecutive states of the system) and considering only the information in which piece of the partition the actual state is offers a coarse-grained view of the system. This is helpful, e.g., to detect periodicities in the dynamics. Because the pieces of partitions are often associated with "symbols", the general methodology is known as *symbolic dynamics* (see Kitchens [64]). In order to analyze the symbol sequences obtained by the orbit of a system, nominal statistics and techniques from information theory can be applied, such as measuring the complexity of an information source by the Shannon entropy.

From the practical viewpoint, one major advantage of using permutation entropy for measuring the complexity of a system is that the computation of estimates can be realized by simple and fast algorithms (see Bandt and Pompe [10], Keller and Sinn [57]). Furthermore, the estimates are robust with respect to observational and dynamical noise (see Bandt and Pompe [10], Veisi et al. [98]). For these reasons, permutation entropy has been applied to real data in recent years, with the main focus on the analysis of epileptic activity in EEG recordings.

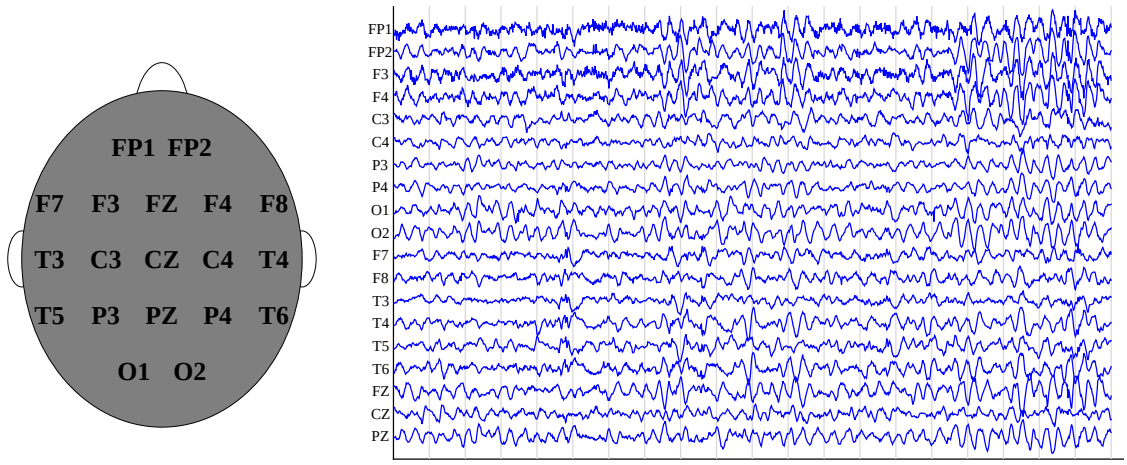

Figure 1.1: The 10-20 system for electrode placement and 20 seconds of an EEG recording.

Application to EEG data analysis. An electroencephalogram (EEG) measures the electric activity of the brain (of humans or animals) by electrodes on the scalp. Figure 1.1 displays the 10-20 system, which is a standard scheme for the placement of 19 electrodes on the human scalp. Furthermore, Figure 1.1 shows a 20 seconds parts from an EEG which was recorded according to the 10-20 system. Note that each line represents one channel, that is, the signal from one electrode. EEG recordings are usually very long (some minutes up to several hours, digitalized with sampling rates of 250 Hertz and higher). Moreover, the brain is a very complex non-linear system (some researchers even believe it is chaotic). Therefore, measuring the complexity of brain dynamics is a particularly challenging task.

Epileptic activity is related to a loss of complexity in the brain dynamics, caused by abnormal synchronizations of large clusters of neuronal cells. Hence, detecting and quantifying epileptic activity by estimating the complexity of the brain dynamics is obvious. See Galka [42] for an overview of applications of entropies, dimensions and Lyapunov exponents. In recent years, many authors have reported epileptic activity to be associated with a decrease of permutation entropy. In particular, Faul et al. [39] discusses permutation entropy for analyzing the EEG of newborn children, Veisi et al. [98] studies the detection of epilepsy in noisy signals and Keller and Lauffer [55] investigates the effect of Vagus stimulation.

In a methodological study, Staniek and Lehnertz [93] discusses the impact of the *order* and *delay* of *ordinal patterns* (see below) on permutation entropy measurements. Li et al. [70] examines the predictability of epileptic seizures for rats, and Bretschneider et al. [18] uses permutation entropy to measure the coherence of in vivo/in vitro field potential activities and of the EEG. See also Cao et al. [24] and Keller et al. [58].

Besides epileptic seizures, other causes of changes in the complexity of the brain dynamics

are, e.g., aging and sleep. For instance, one finds typical differences between the EEGs of a child and an adult, or the EEGs of a person with eyes open and eyes closed. In a study on the detection of brain states preceding epileptic seizures, Bruzzo et al. [21] reports reduced vigilance to result in lower permutation entropy. Olofsen et al. [77] observes a similar effect for anaesthetic drugs and discusses the applicability of permutation entropy for measuring anaesthetic drug effects (see also Li et al. [69]).

In addition to EEG data, permutation entropy has also been used for the analysis of electrocardiogram recordings (Cammarota and Rogora [22, 23], Frank et al. [41]) and econometric data (Matilla-Garcia [74]). Another application is the evaluation of pseudo-random number generators (Larrondo et al. [65]).

1.1 Ordinal time series analysis

In the light of the relation between permutation entropy and the complexity of dynamical systems, together with the results obtained by permutation entropy in practice, the basic approach of considering only the order relations between the values of a time series instead of the values themselves has been further investigated. A general theoretical and methodological framework is established by *ordinal time series analysis* (see Bandt [11], Keller and Sinn [57]).

The central concept of ordinal time series analysis are *ordinal patterns* (or *order patterns*, according to the terminology of Bandt and Shiha [12]). Ordinal patterns represent the order relations among a fixed number of values in a time series. While the *order* of an ordinal pattern specifies the number of values taken into account, the *delay* defines the distances between them. When the values are pairwise different, there exists a unique rank order indicating which is the largest value, which is the second largest value, and so on. If there are equal values, a unique rank order can be established, for instance, by defining earlier values to be larger than subsequent ones. Ordinal patterns are identified with the *permutations* representing the rank orders. Actually, the one-to-one relation between ordinal patterns and permutations gives permutation entropy its name.

For the statistical analysis of time series, the *distribution of ordinal patterns* is of particular interest. Bandt and Shiha [12] considers the difference between certain ordinal pattern probabilities as a measure of the symmetry of time series and shows the application to the detection of trends. Keller and Wittfeld [56] quantifies local differences in the EEG by means of correspondence analysis. Groth [45] uses recurrence plots of ordinal patterns for analyzing speech signals. Keller et al. [57, 58] propose statistics of ordinal pattern distributions which are related to the probability of changes between "upwards" and "downwards" and to the mean length of monotone parts. Note that also permutation entropy is a statistic of ordinal pattern distributions (namely, the Shannon entropy of the distributions).

General results on distributions of ordinal patterns are established by Shiha [91], Bandt and Shiha [12] and Keller et al. [60]. As a major finding, interdependencies among successive ordinal patterns yield fundamental inequalities among ordinal pattern probabilities. Furthermore, the measure of ordinal pattern distributions is concentrated either on a very "thin" or a very "thick" subset of the space of ordinal pattern sequences (see Keller et al. [59, 60], Amigó et al. [6]).

At first glance, considering only the ordinal structure of a time series may seem to be a waste of information. However, compared to methods which take the whole metrical information into account, ordinal time series analysis has some major advantages:

- The ordinal methods are computationally simple and fast (in particular, the computation of ordinal pattern distributions only takes linear time with respect to the time series length). This allows to apply ordinal methods to one and the same time series for various parameters. For example, by considering ordinal patterns with varying delays, time series can be "simultaneously" analyzed on different time scales. Also an online analysis of time series is possible.

- Ordinal methods are robust with respect to transformations of time series which do not affect any order relation among the values, such as offsets and (non-linear) scalings. Thus, ordinal methods are particularly well-suited for the analysis of physical time series where the exact calibration of the measurement device is unknown. Moreover, ordinal methods are relatively robust with respect to artefacts which affect only few of the order relations among values (e.g., slow drifts, abrupt level-changes).

Ordinal patterns in stochastic processes. Originally, the concepts of permutation entropy and ordinal time series analysis have been introduced in the context of dynamical systems. In recent years, also the distribution of ordinal patterns in stochastic processes has been considered. Bandt and Shiha [12] gives explicit formulas for ordinal pattern probabilities in Gaussian processes. These results are used by Rosso et al. [87] for a numerical analysis of permutation entropy in equidistant discretizations of Fractional Brownian Motion. Keller et al. [58] consider *ordinal processes*, with ordinal patterns obtained from real-valued processes as a special case.

From the statistical viewpoint, time series generated by non-linear dynamical systems and by stochastic processes share many similarities. In particular, because of the complexity of the dynamics and the impossibility to exactly measure the initial conditions, time series from non-linear dynamical systems often appear "random" to the observer, despite of their intrinsically deterministic nature. Furthermore, it is often hard to decide by a statistical test whether a time series is generated by a deterministic or a stochastic system. Actually, any dynamical system can formally be represented by a stationary stochastic process and vice versa. In this thesis, we are only interested in statistical properties

of time series. We use the framework of stochastic processes for modelling time series, because it is particularly well suited for describing the problem of estimating ordinal pattern probabilities.

Considering only the order relations among observations is a typical approach in non-parametric statistics. However, while so-called *rank-based statistics* are widely applied in the case of independent observations, e.g., for testing whether there is a difference between two distributions, only few literature is available for the dependent case. Basically, ordinal patterns generate same sigma-fields as *rank vectors*, and if permutations with the same number of inversions are identified, they yield the same information as *Kendall's tau* (see Lehmann [67]). A test for serial dependence in time series based on Kendall's tau is introduced by Ferguson et al. [40]. Garel and Hallin [43] and Hallin and Jurečkova [46] propose rank-based methods for the identification of the order of autoregressive models.

Relation to zero crossings. Most contributions to the investigation of ordinal patterns in stochastic processes originate from the study of *zero crossings*. The event that two successive values of a process have different signs (and thus a straight line connecting these values crosses the level zero) is called a zero crossing. Zero crossings are related to ordinal patterns by the fact that changes between "upwards" and "downwards" (which are fully described by ordinal patterns) are equivalent to zero crossings in the process of first-order differences. In particular, a change from "upwards" to "downwards" is equivalent to a positive difference followed by a negative one, and a change from "downwards" to "upwards" is equivalent to a negative difference followed by a positive one.

In engineering, statistics based on zero crossings are applied to the processing of speech signals. Dating back to the 1940's, telephony engineers found that replacing speech signals with rectangular waves having the same zero crossings retained high intelligibility (see Chang et al. [26]). A mathematical explanation of this phenomenon is given by the formula of Rice [84] which relates the zero crossing rate in a sum of random sinusoids to the dominant frequency in the spectral domain. Since the beginning of digital speech signal processing, zero crossing analysis is used for the detection of pitch frequencies and to distinguish voiced and unvoiced intervals (see, e.g., Ewing and Taylor [38], Rabiner and Schafer [82]).

Kedem [53] proposes estimators for autocorrelations and spectral frequencies based on empirical zero crossing rates of higher order, that is, zero crossings in the processes of higher-order differences. Moreover, Kedem [52] considers zero crossings for the modelling of *binary time series*. Here, the basic idea is that the presence or absence of a phenomenon (represented by "1" and "0", respectively) can often be explained by a real-valued process exceeding or not exceeding a critical level. If the critical level is zero, then the event of observing subsequently "1" and "0" or "0" and "1" corresponds to a zero crossing of the real-valued process. Applications are the modelling of binary data in ecology (see Damsleth and El-Shaarawi [31]) and in hydrology (see Salas et al. [88]).

Note that there is also extensive literature on zero crossings in time-continuous stochastic processes. The questions arising in this context, however, are different from those in the time-discrete case. For example, in the continuous case, the number of zero crossings in a finite time interval is not necessarily finite (see Leadbetter et al. [66], Piterbarg [80]).

Ordinal patterns in equidistant discretizations of Fractional Brownian motion. The class of stochastic processes called *Fractional Brownian Motion* (abbrev. *FBM*) plays an important role both in the theory of stochastic processes and in the statistical modelling of time series. For instance, any *selfsimilar* Gaussian processes with stationary increments belongs to FBM (see Embrechts and Maejima [36]). FBM occurs as the functional limit of partial sums of random variables if the summands have a certain dependence structure (see Taqqu [95]).

In statistics, FBM has been the first model to capture the so-called *Hurst phenomenon*, which is also known as *long range dependence* or *long memory* of time series (see Mandelbrot [71] and Molchan [76] for historical bibliographical remarks). Since long range dependent time series appear in diverse fields such as hydrology, meteorology, econometrics and computer network design, there is a considerable number of monographs devoted to this subject (see Beran [15], Robinson [86], Doukhan et al. [35], Samorodnitsky [89]). Besides long range dependence, FBM is also applied to the modelling of turbulences in physics and in finance (see Mantegna and Stanley [72]).

Bandt and Shiha [12] shows that, as a consequence of the selfsimilarity of FBM, the distribution of ordinal patterns is identical on every time scale. Coeurjolly [28] proposes an estimator of the Hurst parameter based on the empirical zero crossing rate in the increment process, an idea which can generally be applied to the estimation of monotonic functions of the first-order autocorrelation in stationary Gaussian processes. Coeurjolly's estimator is applied to the analysis of hydrological time series (Marković and Koch [73]) and atmospheric turbulence data (Shi et al. [90]).

1.2 Illustration: EEG data analysis

In this section, we illustrate how ordinal methods can be applied to the analysis of EEG data. The reader can skip this part and continue with Section 1.3 where we outline the thesis.

Here, we do not give a mathematical definition of ordinal pattern distributions but an intuitive explanation. Table 1.1 shows the six ordinal patterns which can occur for three sequent values of a time series. We identify the patterns with numbers from 1 to 6. For instance, if the third value is larger than the second one and the second value is larger than the first one, then the ordinal pattern is equal to 1. If the third value is larger than the first one and the first value is larger than the second one, then the ordinal pattern

					
1	2	3	4	5	6

Table 1.1: Ordinal patterns of order $d = 2$.

is equal to 2, and so on. The empirical distribution of ordinal patterns in a time series is obtained by determining the ordinal pattern for each triple of three consecutive values and counting the occurrences of 1 to 6.

Epileptic activity. The upper plot of Figure 1.2 shows a 250 seconds long part of the EEG from an 8-years old boy. The signal has been recorded from the electrode F8 in the 10-20 scheme (see Figure 1.1) and digitalized with a sampling rate of 256 Hertz. Thus, the displayed time series has a length of $250 \cdot 256 = 64\,000$ data points. After 160 seconds, there is an increase of the amplitude which is related to the onset of an epileptic seizure.

The lower plot of Figure 1.2 shows the empirical distribution of ordinal patterns in subsequent non-overlapping parts of 2 seconds (= 512 data points) each. Altogether, we obtain 125 such distributions. The space between horizontal lines represents the relative frequency of ordinal patterns. For instance, the space between the ordinate and the first line from below represents the relative frequency of the ordinal pattern 1, the space between the first and the second line from below represents the relative frequency of the ordinal pattern 2, and so on. Clearly, the relative frequencies of any of the ordinal patterns 1 to 6 add up to 1.

As can be seen from Figure 1.2, the sequence of ordinal pattern distributions clearly reflects the change in the dynamics related to the onset of the epileptic seizure. Typically during epileptic seizures, the EEG exhibits pronounced waveforms with high amplitudes. Therefore, the ordinal patterns 1 and 6, which represent monotone behaviour of the time series (see Table 1.1), are prevailing during this stage.

The background colors of Figure 1.2 visualize the result of a cluster analysis of the ordinal pattern distributions. We have used the total variation distance measure for distributions and the complete linkage algorithm to group the distributions into three clusters (see Härdle and Simar [47] for more information on the clustering method). Roughly speaking, the white cluster corresponds to the time before and after the epileptic seizure, the dark gray cluster to peaks of the seizure, and the light gray cluster to transitions between normal and epileptic activity.

Classification of sleep stages. Figure 1.3 shows another application of ordinal time series analysis. According to the methodology of Rechtschaffen and Kales [83], six differ-

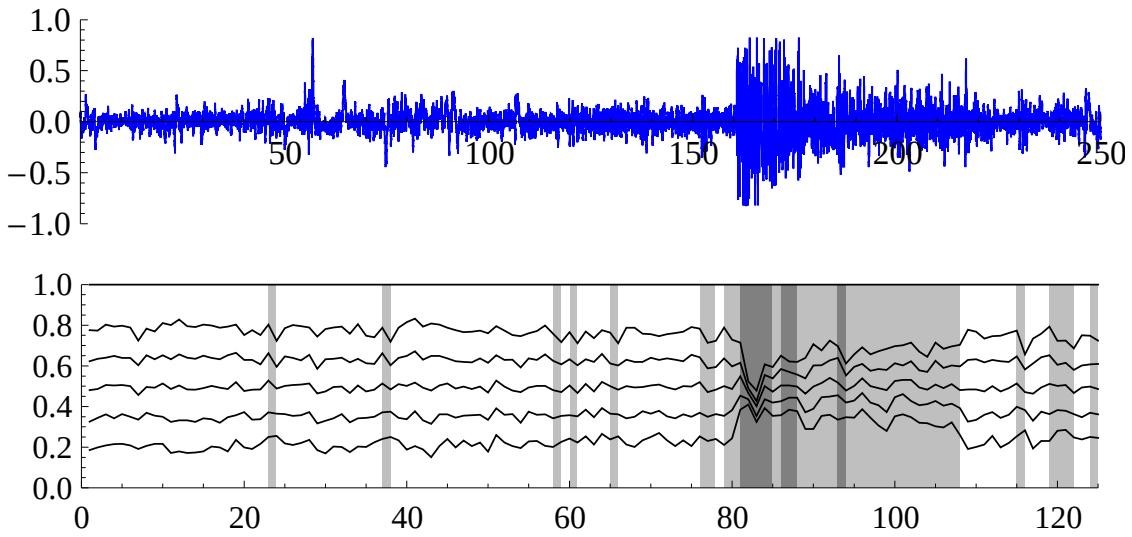

Figure 1.2: EEG time series and ordinal pattern distributions.

ent stages in the sleep of an adult are distinguished: awake (W), sleep stages S1, S2, S3, S4 and rapid eye movement (REM) sleep. The classification is based on the recordings of EEG, electrooculogram (EOG, measuring eye movements) and electromyogram (EMG, measuring muscle activity). The recordings from the whole night are divided into subsequent non-overlapping parts of 30 second, the so-called *epochs*. Separately for each epoch, the corresponding sleep stage is determined. For instance, the transition from W to S1 is associated with a slowdown of the EEG and a decrease of muscle tone. A necessary criterion for S2 is the occurence of so-called sleep spindles and K-complexes in the EEG. REM sleep is characterized by special eye movement patterns and very low mucle tone. S3 and S4 are defined by the amount of delta activity in the EEG, that is, slow waves in a frequency range smaller than 4 Hertz. For S3, the proportion of delta activity is 20 to 50 percent of an epoch, for S4, it is higher than 50 percent.

The upper plot in Figure 1.3 shows a *hypnogram* visualizing the sequence of sleep stages in the first 300 minutes of sleep of an adult. Note that the corresponding 600 epochs have been manually classified by an expert. The hypnogram shows the typical cycles of light and deep sleep: The awake stage (W) at the beginning is followed by light sleep (S1 and S2), deep sleep (S3 and S4), short light sleep (S1 and S2) and REM sleep, before the next cycle begins.

The lower plot in Figure 1.3 displays the empirical ordinal pattern distributions obtained for the 600 epochs. Here, the ordinal patterns are not obtained for directly consecutive values, but for values with a distance of 12 time points in between. We chose this distance because it stresses contrasts between ordinal pattern distributions in different parts of the time series to a maximum extent.

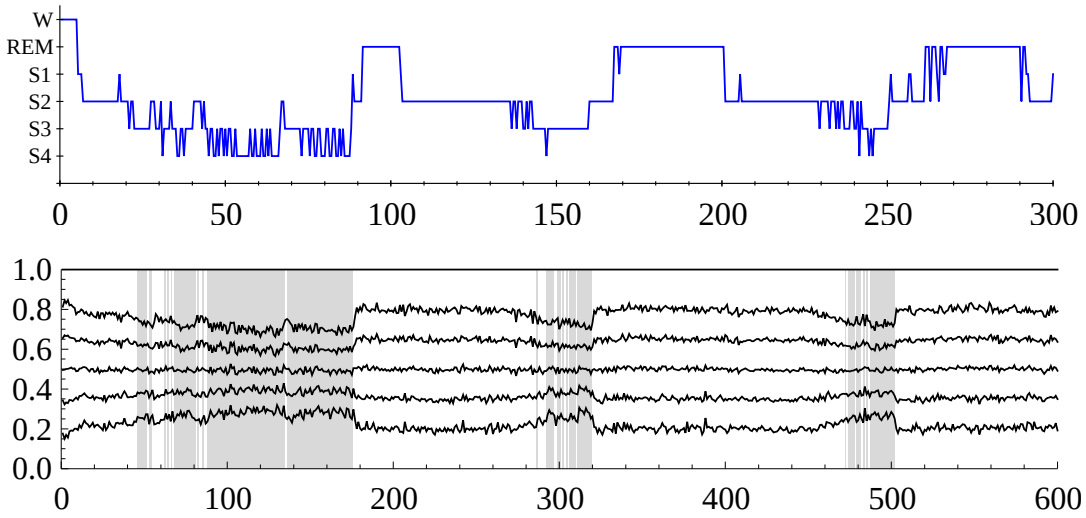

Figure 1.3: Hypnogram and ordinal pattern distributions.

The sleep cycles can be recognized very well. In particular, S3 and S4 coincide with an increase of the frequency of the ordinal patterns 1 and 6, corresponding to a higher proportion of monotone parts in the EEG during these stages. Hardly any difference can be found among the distributions obtained during the stages W, S1, S2 and REM, which is not surprising since the main criterions to distinguish these stages refer to features of EOG and EMG. S3 and S4 result in very similar ordinal pattern distributions, corresponding to the only gradual difference in the proportion of delta activity. Note that more recent methodologies do not distinguish between S3 and S4 (see Iber et al. [51]).

Similar to Figure 1.2, the background color visualizes the result of a cluster analysis. Here, the 600 epochs are gouped into two clusters. It is remarkable how well the white cluster matches the epochs classified as W, S1, S2 and REM, and the gray cluster the epochs classfied as S3 and S4. We find that

- 408 of the 414 epochs ($\approx 98.6\%$) classified as W, S1, S2 and REM are covered by the white cluster.
- 155 of the 186 epochs ($\approx 83.3\%$) classified as S3 and S4 are covered by the gray cluster.

Altogether, 563 of the 600 epochs ($\approx 93.8\%$) are covered by the "right" cluster. Note that we obtain even higher rates if we cluster the distributions of ordinal patterns of higher orders (that is, ordinal patterns describing the order relations among more than three values).

1.3 Outline of the thesis

This thesis studies the estimation of ordinal pattern probabilities. The framework of our analysis are parametric families of stochastic processes with stationary, non-degenerate and zero-mean Gaussian increments. In Chapter 2, we introduce notation and review concepts from the theory of stochastic processes and their distributions. Furthermore, we investigate properties of Fractional Gaussian Noise, ARFIMA(0,d,0) and AR(1) processes.

Chapter 3 is devoted to a general discussion of the estimation of ordinal pattern probabilities. We show that the distribution of ordinal patterns is stationary, and each pattern occurs with a strictly positive probability. Given a finite number of observations, the relative frequency of an ordinal pattern is an unbiased estimator of the corresponding occurrence probability. By the fact that the distribution of stationary zero-mean Gaussian processes is invariant with respect to reversions of the time and space orientation, certain ordinal patterns have the same probability. We show that averaging the relative frequencies of these patterns yields unbiased estimators with smaller variance.

A sufficient condition for the estimators of ordinal pattern probabilities to be consistent is that the autocovariances of the increment process tend to zero. We show that this condition is also sufficient for strong consistency. For asymptotic normality of the estimators it is sufficient that the autocovariances of the increment process decay faster than $k \mapsto \frac{1}{\sqrt{k}}$. More generally, this statement is true for certain differentiable functions of ordinal pattern probabilities and also in the multidimensional case when the probabilities of several patterns are jointly estimated. We illustrate the results for equidistant discretizations of Fractional Brownian Motion and for processes where the increments are ARFIMA(0,d,0) and AR(1), respectively.

In Chapter 4, we study covariances of zero crossings. The results are obtained by analyzing four-dimensional normal orthant probabilities and their derivatives with respect to correlation coefficients. We propose a representation of zero crossing covariances by one-dimensional integrals which can be numerically evaluated using standard quadrature rules. Furthermore, we derive asymptotics of the covariances and establish approximations and bounds. Based on these results, we derive properties of the variance of empirical zero crossing rates.

In Chapter 5, we focus on ordinal patterns of order $d = 2$. We show that any "reasonable" estimator of ordinal pattern probabilities can be expressed as an affine function of the empirical zero crossing rate in the increment process. Using the results of Chapter 4, we evaluate the variance of the estimators in equidistant discretizations of Fractional Brownian Motion and in processes where the increments are ARFIMA(0,d,0) and AR(1), respectively.

When the parameters of the family of stochastic processes are real numbers and monotonically related to the probability of a change, an estimator of the parameters is obtained by

plugging the frequency of changes into the inverse of the monotonic relation. Using the results of Chapter 3, we establish properties of this estimator. Under additional conditions on the autocovariances of the increment process, we also derive confidence intervals. We show how the results apply to the estimation of the Hurst parameter in Fractional Brownian Motion, of the fractional differencing parameter in ARFIMA(0,d,0) processes and of the autoregressive coefficient in AR(1) processes.

In a simulation study, we evaluate the performance of the estimators and the coverage of the parameters by the confidence intervals. For the Hurst parameter, we compare the performance to that of an alternative estimator. We also consider the distribution of empirical zero crossing rates in the increment process of FBM. It turns out that the distribution is very irregular when the Hurst parameter is large and thus the increment process exhibits long range dependence.

Chapter 6 generalizes the results of the previous chapters to ordinal patterns with arbitrary delays. We demonstrate how patterns with increasing delays can be used for estimating the Hurst parameter in asymptotically self-similar processes. One application is the estimation of the Hurst parameter in equidistant discretizations of Fractional Brownian Motion superimposed with short range dependent "noise". We illustrate our method for two practical time series, namely, River Nile data and NBS precision measurements.

Chapter 2

Preliminaries

This chapter introduces notation and concepts from the theory of stochastic processes. Furthermore, we provide some properties of important classes of Gaussian processes.

2.1 Notation

Except for commonly used notation, we will explain the meaning of symbols at their first occurence in the text. For a list of symbols used throughout the thesis, see p. 147.

Sets. We use "$\subset$" to denote inclusion or equality of sets. $\mathbb{N} = \{ 1, 2, \dots \}$ stands for the set of natural numbers and $\mathbb{N}_0$ for $\mathbb{N} \cup \{0\}$. We write $\mathbb{Z} = \{ \dots, -2, -1, 0, 1, 2, \dots \}$ for the set of integers and $\mathbb{R}$ for the set of real numbers.

If $\mathcal{Y}$ is a set and $\mathbf{T} = \mathbb{R}$ or $\mathbf{T} = \mathbb{Z}$, then $\mathcal{Y}^{\mathbf{T}}$ denotes the set of sequences $(y_t)_{t \in \mathbf{T}}$ with $y_t \in \mathcal{Y}$ for $t \in \mathbf{T}$. By $\sharp\, \mathcal{Y}$ we denote the cardinality of $\mathcal{Y}$. If $\mathcal{Y}$ is a topological space, we write $\mathbb{B}(\mathcal{Y})$ for the Borel σ-field of $\mathcal{Y}$, that is, the smallest σ-field containing all open subsets of $\mathcal{Y}$.

Matrices. We use bold uppercase letters to denote matrices and bold lowercase letters to denote vectors. For $n \in \mathbb{N}$, we write $\mathbb{R}^n$ for the set of n-dimensional row vectors (i.e., $\mathbb{R}^n$ stands for $\mathbb{R}^{1 \times n}$). For $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$, let

$$\mathbf{diag}(a_1, a_2, \dots, a_n) \;=\; \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix}$$

be the $n \times n$ matrix with the entries $a_1, a_2, \ldots, a_n$ on the main diagonal and zero entries, otherwise. By $\mathbf{0}$ we denote the vector $(0, 0, \ldots, 0) \in \mathbb{R}^n$, and $\mathbf{I}_n$ stands for the matrix $\mathbf{diag}(1, 1, \ldots, 1) \in \mathbb{R}^{n \times n}$.

Asymptotics. We write $g(k) = O(h(k))$ for mappings $g, h : \mathbb{N} \to \mathbb{R}$ iff

$$\limsup_{k \to \infty} \frac{|g(k)|}{|h(k)|} \quad < \quad \infty \, .$$

We write $g(k) = o(h(k))$ iff

$$\lim_{k \to \infty} \frac{g(k)}{h(k)} \quad = \quad 0 \, .$$

With $\frac{0}{0} := 1$, we write $g(k) \sim h(k)$ and say g *is asymptotically equivalent to* h iff

$$\lim_{k \to \infty} \frac{g(k)}{h(k)} \quad = \quad 1 \, .$$

2.2 Stochastic processes

In Section 2.2.1, we review basic concepts of random variables and their distributions and give corresponding definitions for stochastic processes. We recall the definition of Gaussian processes in Section 2.2.2 and give properties of Fractional Gaussian Noise, ARFIMA(0,d,0) and AR(1) processes in Sections 2.2.3 - 2.2.5.

2.2.1 Basic concepts

Throughout this section, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{Y}, \mathcal{B})$ a measurable space.

Definition 2.1. Let $Y, Y_1, Y_2, \ldots$ and Z be $(\mathcal{Y}, \mathcal{B})$-valued random variables on $(\Omega, \mathcal{A}, \mathbb{P})$.

(*i*) We write $Y \stackrel{\text{dist}}{=} Z$ and say Y *and* Z *have the same distribution* (or Y *and* Z *are identically distributed*) iff $\mathbb{P}(Y \in B) = \mathbb{P}(Z \in B)$ for every $B \in \mathcal{B}$.

(*ii*) Suppose $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ for some $k \in \mathbb{N}$. We say Y *is non-degenerate* iff

$$\mathbb{P}(Y \in B) \quad > \quad 0$$

only if $B \in \mathbb{B}(\mathbb{R}^k)$ has strictly positive Lebesgue measure.

(*iii*) Let $\mathcal{Y}$ be a topological space, $\mathcal{B} = \mathbb{B}(\mathcal{Y})$ and ∂B the boundary of a set $B \in \mathcal{B}$. We write

$$Y_n \quad \xrightarrow{\mathbb{P}} \quad Y$$

and say Y_n *converges to* Y *in distribution* iff

$$\lim_{n\to\infty} \mathbb{P}(Y_n \in B) \quad = \quad \mathbb{P}(Y \in B)$$

for every $B \in \mathcal{B}$ with $\mathbb{P}(Y \in \partial B) = 0$.

Clearly, if $Y = Z$ $\mathbb{P}$-almost surely (that is, $\mathbb{P}(Y = Z) = 1$), then Y and Z are identically distributed. Some authors use a different notion of non-degeneracy. For example, Bauer [14] calls $\mathbf{Y}$ non-degenerate if there does not exist a $\mathbf{y} \in \mathbb{R}^k$ with $\mathbb{P}(Y = \mathbf{y}) = 1$.

Note that there are equivalent ways to define convergence in distribution (see Theorem 29.1 in Billingsley [17]). In the special case where $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ with $k \in \mathbb{N}$, a common definition is given by pointwise convergence of the distribution functions of $Y_1, Y_2, \ldots$ to the distribution function of Y at all continuity points of the distribution function of Y (see Billingsley [17], pp. 327, 329).

Clearly, we may replace Y in (*iii*) with any other random variable having the same distribution as Y. When Y is normally distributed with mean μ and variance σ^2, we occasionally write

$$Y_n \quad \xrightarrow{\mathbb{P}} \quad \mathrm{N}(\mu, \sigma^2)$$

(for the definition of $\mathrm{N}(\mu, \sigma^2)$, see Section 2.2.2).

Stochastic processes. Let $\mathbf{T} = \mathbb{R}$ or $\mathbf{T} = \mathbb{Z}$. Any family $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ of $(\mathcal{Y}, \mathcal{B})$-measurable mappings on $(\Omega, \mathcal{A}, \mathbb{P})$ is called a $(\mathcal{Y}, \mathcal{B})$-*valued stochastic process*. We say $\mathbf{Y}$ *is time-continuous* if $\mathbf{T} = \mathbb{R}$, and $\mathbf{Y}$ *is time-discrete*, otherwise. In the following chapters, when $\mathbf{Y}$ is time-discrete, we write time indices as subscripts, that is, Y_t instead of $Y(t)$.

For $k \in \mathbb{N}$, let $\mathcal{B}^{\otimes k}$ denote the smallest σ-field containing any set $B_1 \times B_2 \times \ldots \times B_k$ with $B_1, B_2, \ldots, B_k \in \mathcal{B}$. Clearly, for all $t_1, t_2, \ldots, t_k \in \mathbf{T}$ with $k \in \mathbb{N}$, the mapping $(Y(t_1), Y(t_2), \ldots, Y(t_k))$ is $(\mathcal{Y}^k, \mathcal{B}^{\otimes k})$-measurable. If $\mathcal{Y}$ is finite and $\mathcal{B}$ is the power set of $\mathcal{Y}$, or $\mathcal{Y} = \mathbb{R}$ and $\mathcal{B} = \mathbb{B}(\mathbb{R})$, then the distributions of all such finite-dimensional random vectors uniquely determine a probability measure on $(\mathcal{Y}^{\mathbf{T}}, \mathbb{B}(\mathcal{Y}^{\mathbf{T}}))$. (More generally, this statement is true when $\mathcal{Y}$ is a separable and completely metrizable topological space and $\mathcal{B}$ is the corresponding Borel σ-field, see Theorem 35.3 in Bauer [14]. Indeed, finite sets equipped with the discrete topology and $\mathbb{R}$ equipped with the Euclidean topology are separable and completely metrizable, see Engelking [37], pp. 25-26 and pp. 268-269). We refer to this probability measure as the *distribution of* $\mathbf{Y}$. Clearly, $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ and

$\mathbf{Z} = (Z(t))_{t\in\mathbf{T}}$ have the same distribution iff all finite-dimensional distributions of $\mathbf{Y}$ and $\mathbf{Z}$ are identical, that is,

$$(Y(t_1),\, Y(t_2),\, \ldots,\, Y(t_k)) \;\stackrel{\text{dist}}{=}\; (Z(t_1),\, Z(t_2),\, \ldots,\, Z(t_k))$$

for all $t_1, t_2, \ldots, t_k \in \mathbf{T}$ with $k \in \mathbb{N}$. In this case, we write $\mathbf{Y} \stackrel{\text{dist}}{=} \mathbf{Z}$.

The following definition introduces basic properties of stochastic processes.

Definition 2.2. Let $\mathbf{T} = \mathbb{R}$ or $\mathbf{T} = \mathbb{Z}$. Assume that $\mathcal{Y}$ is finite and $\mathcal{B}$ is the power set of $\mathcal{Y}$, or $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}, \mathbb{B}(\mathbb{R}))$. Let $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$, $\mathbf{Y}^{(n)} = (Y^{(n)}(t))_{t\in\mathbf{T}}$ for $n = 1, 2, \ldots$ and $\mathbf{Z} = (Z(t))_{t\in\mathbf{T}}$ be $(\mathcal{Y}, \mathcal{B})$-valued stochastic processes on $(\Omega, \mathcal{A}, \mathbb{P})$.

(*i*) $\mathbf{Y}$ *is stationary* iff $(Y(t))_{t\in\mathbf{T}} \stackrel{\text{dist}}{=} (Y(t+\tau))_{t\in\mathbf{T}}$ for all $\tau \in \mathbf{T}$.

(*ii*) Let $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}, \mathbb{B}(\mathbb{R}))$. $\mathbf{Y}$ *is weakly stationary* iff $\mathbb{E}(Y(t)) = \mathbb{E}(Y(0))$ for all $t \in \mathbf{T}$ and $\mathrm{Cov}(Y(s), Y(t)) = \mathrm{Cov}(Y(0), Y(t-s))$ for all $s, t \in \mathbf{T}$.

(*iii*) Let $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}, \mathbb{B}(\mathbb{R}))$. $\mathbf{Y}$ *is zero-mean* iff $\mathbb{E}(Y(t)) = 0$ for all $t \in \mathbf{T}$. $\mathbf{Y}$ *has unit variance* iff $\mathrm{Var}(Y(t)) = 1$ for all $t \in \mathbf{T}$.

(*iv*) Let $(\mathcal{Y}, \mathcal{B}) = (\mathbb{R}, \mathbb{B}(\mathbb{R}))$. $\mathbf{Y}$ *is non-degenerate* iff $(Y(t_1),\, Y(t_2),\, \ldots,\, Y(t_k))$ is non-degenerate for all $t_1 < t_2 < \ldots < t_k \in \mathbf{T}$ with $k \in \mathbb{N}$.

(*v*) We say $\mathbf{Y}^{(n)}$ *converges to* $\mathbf{Y}$ *in distribution* and write

$$\mathbf{Y}^{(n)} \;\stackrel{\mathbb{P}}{\longrightarrow}\; \mathbf{Y}$$

iff all finite-dimensional distributions of $\mathbf{Y}^{(n)}$ converge to those of $\mathbf{Y}$, that is,

$$(Y^{(n)}(t_1),\, Y^{(n)}(t_2),\, \ldots,\, Y^{(n)}(t_k)) \;\stackrel{\mathbb{P}}{\longrightarrow}\; (Y(t_1),\, Y(t_2),\, \ldots,\, Y(t_k))$$

for all $t_1, t_2, \ldots, t_k \in \mathbf{T}$ with $k \in \mathbb{N}$.

Clearly, we may replace $\mathbf{Y}$ in (*v*) with any other stochastic process having the same distribution as $\mathbf{Y}$. For instance, we write

$$\mathbf{Y}^{(n)} \;\stackrel{\mathbb{P}}{\longrightarrow}\; \mathbf{G}_{\boldsymbol{H}}$$

if the finite-dimensional distributions of $\mathbf{Y}^{(n)}$ converge to those of Fractional Gaussian Noise with the Hurst parameter $\boldsymbol{H}$ (see Section 2.2.3).

2.2.2 Gaussian processes

Throughout this section, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

Definition 2.3. Let Y and $(Y_1, Y_2, \ldots, Y_k)$ with $k \in \mathbb{N}$ be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ and $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$, respectively.

(*i*) *Y has a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in [0, \infty)$* (or *Y is distributed according to* $\mathrm{N}(\mu, \sigma^2)$) iff, in the case $\sigma^2 = 0$,

$$\mathbb{P}(Y = \mu) \;=\; 1$$

or, otherwise,

$$\mathbb{P}(Y \in B) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma} \int_B \exp\big(-\frac{(x-\mu)^2}{2\,\sigma^2}\big)\,\mathrm{d}x$$

for every $B \in \mathbb{B}(\mathbb{R})$. The normal distribution with mean 0 and variance 1 is called the *standard normal distribution*.

(*ii*) *$(Y_1, Y_2, \ldots, Y_k)$ has a normal distribution with means $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$* (or *$(Y_1, Y_2, \ldots, Y_k)$ is distributed according to* $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) iff, for every $\mathbf{a} = (a_1, a_2, \ldots, a_k) \in \mathbb{R}^k$,

$$a_1 Y_1 + a_2 Y_2 + \ldots + a_k Y_k \text{ is distributed according to } \mathrm{N}(\mathbf{a}\,\boldsymbol{\mu}^T,\ \mathbf{a}\,\boldsymbol{\Sigma}\,\mathbf{a}^T)\,.$$

The normal distribution with means $\mathbf{0}$ and covariance matrix $\mathbf{I}_k$ is called the *standard normal distribution*.

It is well-known that if Y is distributed according to $\mathrm{N}(\mu, \sigma^2)$, then $\mathbb{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$. Furthermore, if $(Y_1, Y_2, \ldots, Y_k)$ has a normal distribution with means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^k$, then

$$\mathbb{E}(Y_i) = \mu_i \text{ for } i = 1, 2, \ldots, k \;\text{ and }\; \mathrm{Cov}(Y_i, Y_j) = \sigma_{ij} \text{ for } i, j = 1, 2, \ldots, k.$$

In particular, the covariance matrix $\boldsymbol{\Sigma}$ is symmetric and positive definite, that is, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$ and $\mathbf{a}\,\boldsymbol{\Sigma}\,\mathbf{a}^T \geq 0$ for every $\mathbf{a} \in \mathbb{R}^k$. The condition that $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$ is symmetric and positive definite is also sufficient for the existence of a normal distribution with means $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance matrix $\boldsymbol{\Sigma}$. If $(Y_1, Y_2, \ldots, Y_k)$ is distributed according to $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A} \in \mathbb{R}^{l \times k}$, then $\mathbf{A}\,(Y_1, Y_2, \ldots, Y_k)^T$ is distributed according to $\mathrm{N}(\mathbf{A}\,\boldsymbol{\mu}^T,\ \mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}^T)$ (see Billingsley [17], p. 384).

The following theorem gives other well-known properties of normal distributions. We state these properties in a theorem, because we will frequently refer to them throughout the rest of the thesis.

Theorem 2.4.

(*i*) *If* $(Y_1, Y_2, \ldots, Y_k)$ *is distributed according to* $\mathrm{N}(\boldsymbol{\mu},\, \boldsymbol{\Sigma})$, *then* $(Y_1, Y_2, \ldots, Y_k)$ *is non-degenerate if and only if* $\boldsymbol{\Sigma}$ *is strictly positive definite, that is,* $\mathbf{a}\,\boldsymbol{\Sigma}\,\mathbf{a}^T > 0$ *for every* $\mathbf{a} \in \mathbb{R}^k$ *with* $\mathbf{a} \neq (0, 0, \ldots, 0)$. *If* $\boldsymbol{\Sigma}$ *is strictly positive definite, then*

$$\mathbb{P}((Y_1, Y_2, \ldots, Y_k) \in B) \;\; = \;\; \int_B \phi(\boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \mathbf{x})\, d\mathbf{x}$$

for every $B \in \mathbb{B}(\mathbb{R}^k)$, *with* $\phi(\boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \cdot)$ *given by*

$$\phi(\boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \mathbf{x}) \;\; := \;\; \big((2\pi)^k \det(\boldsymbol{\Sigma})\big)^{-\frac{1}{2}} \exp\big(-\frac{1}{2}\,(\mathbf{x}-\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T\big)$$

for $\mathbf{x} \in \mathbb{R}^k$. *In particular, if* $B \in \mathbb{B}(\mathbb{R}^k)$ *has strictly positive Lebesgue measure, then*

$$\mathbb{P}((Y_1, Y_2, \ldots, Y_k) \in B) \;\; > \;\; 0\,.$$

(*ii*) *If* $(Y_1, Y_2, \ldots, Y_k)$ *and* $(Z_1, Z_2, \ldots, Z_k)$ *are normally distributed, then* $(Y_1, Y_2, \ldots, Y_k)$ *and* $(Z_1, Z_2, \ldots, Z_k)$ *are identically distributed if and only if* $\mathbb{E}(Y_i) = \mathbb{E}(Z_i)$ *for* $i = 1, 2, \ldots, k$ *and* $\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}(Z_i, Z_j)$ *for* $i, j = 1, 2, \ldots, k$.

(*iii*) *If* $(Y_1, Y_2, \ldots, Y_k)$ *and* $(Z_1, Z_2, \ldots, Z_l)$ *are normally distributed, then* $(Y_1, Y_2, \ldots, Y_k)$ *and* $(Z_1, Z_2, \ldots, Z_l)$ *are independent if and only if* $\mathrm{Cov}(Y_i, Z_j) = 0$ *for* $i = 1, 2, \ldots, k$ *and* $j = 1, 2, \ldots, l$.

(*iv*) *If* $(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_k^{(1)})$, $(Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_k^{(2)})$, $\ldots$ *are normally distributed, then*

$$(Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_k^{(n)}) \;\; \xrightarrow{\;\mathbb{P}\;} \;\; (Y_1, Y_2, \ldots, Y_k)$$

if and only if $(Y_1, Y_2, \ldots, Y_k)$ *is normally distributed,* $\lim_{n\to\infty} \mathbb{E}(Y_i^{(n)}) = \mathbb{E}(Y_i)$ *for* $i = 1, 2, \ldots, k$ *and* $\lim_{n\to\infty} \mathrm{Cov}(Y_i^{(n)}, Y_j^{(n)}) = \mathrm{Cov}(Y_i, Y_j)$ *for* $i, j = 1, 2, \ldots, k$.

Proof. (*i*)-(*iii*) See Billingsley [17], pp. 384-385. (*iv*) is easily established by means of characteristic functions (see Theorems 26.2, 26.3 and 29.4 in Billingsley [17]). □

Next, we recall the concept of Gaussian processes.

Definition 2.5. Let $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ with $\mathbf{T} = \mathbb{R}$ or $\mathbf{T} = \mathbb{Z}$ be an $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$-valued stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$. We say $\mathbf{Y}$ *is a Gaussian process* (or $\mathbf{Y}$ *is Gaussian*) iff, for all $t_1, t_2, \ldots, t_k \in \mathbf{T}$ with $k \in \mathbb{N}$, there exist a vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and a symmetric positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$ such that

$$(Y(t_1), Y(t_2), \ldots, Y(t_k)) \text{ is distributed according to } \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The following theorem states some of the properties of Gaussian processes. We give a proof of (iv), because it is usually not included in the more applied literature.

Theorem 2.6.

(i) *If* $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ *is Gaussian, then* $\mathbf{Y}$ *is stationary if and only if* $\mathbf{Y}$ *is weakly stationary.*

(ii) *If* $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ *and* $\mathbf{Z} = (Z(t))_{t\in\mathbf{T}}$ *are Gaussian, then* $\mathbf{Y}$ *and* $\mathbf{Z}$ *have the same distribution if and only if* $\mathbb{E}(Y(t)) = \mathbb{E}(Z(t))$ *for all* $t \in \mathbf{T}$ *and* $\mathrm{Cov}(Y(s), Y(t)) = \mathrm{Cov}(Z(s), Z(t))$ *for all* $s, t \in \mathbf{T}$.

(iii) *If* $\mathbf{Y}^{(n)} = (Y^{(n)}(t))_{t\in\mathbf{T}}$ *is Gaussian for* $n = 1, 2, \ldots$, *then*

$$\mathbf{Y}^{(n)} \quad \xrightarrow{\mathbb{P}} \quad \mathbf{Y}$$

if and only if $\mathbf{Y}$ *is Gaussian,* $\lim_{n\to\infty} \mathbb{E}(Y^{(n)}(t)) = \mathbb{E}(Y(t))$ *for all* $t \in \mathbf{T}$ *and*

$$\lim_{n\to\infty} \mathrm{Cov}(Y^{(n)}(s), Y^{(n)}(t)) \quad = \quad \mathrm{Cov}(Y(s), Y(t))$$

for all $s, t \in \mathbf{T}$.

(iv) *Suppose* $\mathbf{Y} = (Y(t))_{t\in\mathbf{T}}$ *with* $\mathbf{T} = \mathbb{Z}$ *is Gaussian and stationary. Then a necessary condition for* $\mathbf{Y}$ *to be non-degenerate is given by* $|\mathrm{Cov}(Y(0), Y(k))| < \mathrm{Var}(Y(0))$ *for all* $k \in \mathbb{N}$. *A sufficient condition for* $\mathbf{Y}$ *to be non-degenerate is that* $\mathrm{Var}(Y(0)) > 0$ *and* $\mathbf{Y}$ *has a spectral density* $f_{\mathbf{Y}}$, *that is,*

$$f_{\mathbf{Y}}(\lambda) \quad := \quad \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{Cov}(Y(0), Y(k))\, e^{-i\lambda k}$$

is well-defined for all $\lambda \in [-\pi, \pi]$, *and*

$$\mathrm{Cov}(Y(0), Y(k)) \quad = \quad \int_{-\pi}^{\pi} f_{\mathbf{Y}}(\lambda)\, e^{i\lambda k} \mathrm{d}\lambda$$

for all $k \in \mathbb{Z}$.

Proof. (i) See Brockwell and Davis [19], p. 13. (ii) and (iii) are immediate consequences of Theorem 2.4 (ii) and (iv).

(iv) Let $k \in \mathbb{N}$. If $|\mathrm{Cov}(Y(0), Y(k))| = \mathrm{Var}(Y(0))$, then the covariance matrix of $(Y(0), Y(k))$ is singular and thus, according to Theorem 2.4 (i), $(Y(0), Y(k))$ is degenerate. Hence, $|\mathrm{Cov}(Y(0), Y(k))| < \mathrm{Var}(Y(0))$ is a necessary condition for $\mathbf{Y}$ to be non-degenerate.

In order to establish the sufficient condition, suppose that $\mathrm{Var}(Y(0)) > 0$ and $\mathbf{Y}$ has a spectral density $f_{\mathbf{Y}}$. Furthermore, without loss of generality, suppose that $\mathbf{Y}$ is zero-mean. Now, assume there exist $t_1 < t_2 < \ldots < t_k \in \mathbb{Z}$ with $k \in \mathbb{N}$ such that $(Y(t_1), Y(t_2), \ldots, Y(t_k))$ is degenerate. Let $\mathbf{\Sigma} := (\mathrm{Cov}(Y(t_i), Y(t_j)))_{i,j=1}^{k}$. Since $\mathbf{Y}$ is Gaussian, it follows the existence a vector $\mathbf{a} = (a_1, a_2, \ldots, a_k) \in \mathbb{R}^k$ with $\mathbf{a} \neq (0, 0, \ldots, 0)$ such that $\mathbf{a}\,\mathbf{\Sigma}\,\mathbf{a}^T = 0$. Consequently, $\mathrm{Var}(a_1 Y(t_1) + a_2 Y(t_2) + \ldots + a_k Y(t_k)) = 0$ and thus $a_1 Y(t_1) + a_2 Y(t_2) + \ldots + a_k Y(t_k) = 0$ $\mathbb{P}$-almost surely. Without loss of generality, we may assume that $a_k \neq 0$ (if $a_k = 0$, then $(Y(t_1), Y(t_2), \ldots, Y(t_{k-1}))$ is degenerate). For $s \in \mathbb{N}_0$, define

$$a(s) \;:=\; \begin{cases} -\frac{a_{k-i}}{a_k} & \text{if } i \in \{1, 2, \ldots, k-1\} \text{ is such that } s = t_k - t_{k-i} \\ 0 & \text{otherwise} \end{cases} .$$

Note that $a_1 Y(t_1) + a_2 Y(t_2) + \ldots + a_k Y(t_k) = 0$ is equivalent to

$$Y(t_k) \;=\; \sum_{s=0}^{\infty} a(s) Y(t_k - s) \,.$$

Since $\mathbf{Y}$ is stationary, we obtain that

$$Y(t) \;=\; \sum_{s=0}^{\infty} a(s) Y(t - s)$$

$\mathbb{P}$-almost surely for every $t \in \mathbb{Z}$. Let $\mu \in [0, 1]$ and consider the process $\mathbf{X} = (X(t))_{t \in \mathbb{Z}}$ defined by $X(t) := \mu Y(t) + (1 - \mu) \sum_{s=0}^{\infty} a(s) Y(t - s)$ for $t \in \mathbb{Z}$. Since $\mathbf{X}$ and $\mathbf{Y}$ are identical, $\mathbf{X}$ has a spectral density $f_{\mathbf{X}}$, and $f_{\mathbf{X}}(\lambda) = f_{\mathbf{Y}}(\lambda)$ for all $\lambda \in [0, \pi]$. On the other hand, $\mathbf{X}$ is obtained by a linear filtering of $\mathbf{Y}$, and thus

$$f_{\mathbf{X}}(\lambda) \;=\; \left| \mu + (1 - \mu) \sum_{s=0}^{\infty} a(s) e^{is\lambda} \right|^2 f_{\mathbf{Y}}(\lambda)$$

for $\lambda \in [0, \pi]$ (see Theorem 4.10.1 in Brockwell and Davis [19]). Let $N := t_k - t_1$. Since $a(s) = 0$ for $s = 0$ and $s > N$, it follows that

$$\left| \mu + (1 - \mu) \sum_{s=1}^{N} a(s) e^{is\lambda} \right| \;=\; 1$$

for all $\lambda \in [0, \pi]$ with $f_{\mathbf{Y}}(\lambda) > 0$. Because $\mu \in [0, 1]$ is arbitrary, we obtain

$$1 - \sum_{s=1}^{N} a(s) e^{is\lambda} \;=\; 0 \tag{2.1}$$

for all $\lambda \in [0, \pi)$ with $f_{\mathbf{Y}}(\lambda) > 0$. Since $\mathrm{Var}(Y(0)) = \int_{-\pi}^{\pi} f_{\mathbf{Y}}(\lambda) \mathrm{d}\lambda$ and $f_{\mathbf{Y}}(\lambda) = f_{\mathbf{Y}}(-\lambda)$ for all $\lambda \in [0, \pi)$ (see Brockwell and Davis [19], pp. 121-122), the assumption $\mathrm{Var}(Y(0)) > 0$ yields the existence of infinitely many $\lambda \in [0, \pi)$ with $f_{\mathbf{Y}}(\lambda) > 0$. Therefore, we find pairwise different $\lambda_1, \lambda_2, \ldots, \lambda_{N+1} \in [0, \pi)$ such that $f_{\mathbf{Y}}(\lambda_i) > 0$ for $i = 1, 2, \ldots, N+1$. Since the Vandermonde matrix

$$\mathbf{A} \;=\; \begin{pmatrix} 1 & e^{\lambda_1 i} & e^{2\lambda_1 i} & \ldots & e^{N\lambda_1 i} \\ 1 & e^{\lambda_2 i} & e^{2\lambda_2 i} & \ldots & e^{N\lambda_2 i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{\lambda_{N+1} i} & e^{2\lambda_{N+1} i} & \ldots & e^{N\lambda_{N+1} i} \end{pmatrix}$$

has rank $N+1$ (see Horn and Johnson [49], p. 29), the only solution of the linear equation $\mathbf{A}\mathbf{x}^T = \mathbf{0}$ is given by $\mathbf{x} = (0, 0, \ldots, 0)$ (see Horn and Johnson [49], p. 14). However, equation (2.1) holds for $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_{N+1}$, and thus $(1, -a(1), -a(2), \ldots, -a(N))$ is another solution of the linear equation. This contradiction shows that $\mathbf{Y}$ is non-degenerate. $\square$

2.2.3 Fractional Gaussian Noise

Let $\mathbf{Y} = (Y(t))_{t \in \mathbb{Z}}$ be a family of $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$-measurable mappings on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H} \in (0,1)}$. The subscript $\boldsymbol{H}$ (for instance, $\mathbb{E}_{\boldsymbol{H}}$, $\mathrm{Var}_{\boldsymbol{H}}$, etc.) indicates integration with respect to $\mathbb{P}_{\boldsymbol{H}}$.

Assume that, for $\boldsymbol{H} \in (0, 1)$, the following conditions are satisfied:

(A1) $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is zero-mean Gaussian.

(A2) The autocovariances of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ are given by

$$\mathrm{Cov}_{\boldsymbol{H}}(Y(t),\, Y(t+k)) \;=\; \rho_{\boldsymbol{H}}(k) \;:=\; \frac{1}{2}\left(|k+1|^{2\boldsymbol{H}} - 2\,|k|^{2\boldsymbol{H}} + |k-1|^{2\boldsymbol{H}} \right)$$

for $t, k \in \mathbb{Z}$.

Then $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is called (*standard*) *Fractional Gaussian Noise* (FGN) with the *Hurst parameter* $\boldsymbol{H}$. For $\boldsymbol{H} \in (0, 1)$, we denote by $\mathbf{G}_{\boldsymbol{H}}$ the distribution of FGN with the Hurst parameter $\boldsymbol{H}$. According to Theorem 2.6 (i), $\mathbf{G}_{\boldsymbol{H}}$ is uniquely determined by (A1) and (A2).

Note that $(\Omega, \mathcal{A})$, $\mathbf{Y}$ and $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H} \in (0,1)}$ such that $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is FGN with the Hurst parameter $\boldsymbol{H}$ for every $\boldsymbol{H} \in (0, 1)$ actually do exist. In particular, the existence of FGN with the Hurst parameter $\boldsymbol{H}$ follows by the existence of FBM with the Hurst parameter $\boldsymbol{H}$ (see below). Now, $(\Omega, \mathcal{A})$ can be chosen as the space of real-valued sequences $\mathbb{R}^{\mathbb{Z}}$ with the corresponding Borel σ-algebra, $\mathbf{Y}$ as the identity on $(\Omega, \mathcal{A})$, and $\mathbb{P}_{\boldsymbol{H}}$ as the probability measure on $(\Omega, \mathcal{A})$ induced by $\mathbf{G}_{\boldsymbol{H}}$.

We refer to Fractional Gaussian Noise as defined by (A1) and (A2) as *standard* FGN, because $\mathrm{Var}_{\boldsymbol{H}}(Y(t)) = \rho_{\boldsymbol{H}}(0) = 1$ for all $t \in \mathbb{Z}$. In particular,

$$\rho_{\boldsymbol{H}}(k) \;=\; \mathrm{Corr}_{\boldsymbol{H}}(Y(t),\, Y(t+k))$$

for all $t, k \in \mathbb{Z}$. More generally, one can consider FGN with the autocovariances given by $\mathrm{Cov}_{\boldsymbol{H}}(Y(t),\, Y(t+k)) = \sigma^2 \rho_{\boldsymbol{H}}(k)$ for some $\sigma^2 > 0$.

It is well-known that FGN with the Hurst parameter $\boldsymbol{H} \in (0,1)$ is stationary (see Beran [15], p. 55). Furthermore, FGN with the Hurst parameter $\boldsymbol{H} \in (0,1)$ has a spectral density (see Proposition 2.1 in Beran [15]). Thus, according to Theorem 2.6 (iv), FGN with the Hurst parameter $\boldsymbol{H} \in (0,1)$ is non-degenerate.

Note that one can also consider FGN as given by (A1) and (A2) for the Hurst parameter $\boldsymbol{H} = 1$ (see Beran [15]). In this case, $\rho_{\boldsymbol{H}}(k) = 1$ for every $k \in \mathbb{Z}$ which shows that $\mathbf{Y}$ is degenerate (see Theorem 2.6 (iv)). In particular, $\mathrm{Var}_{\boldsymbol{H}}(Y(s) - Y(t)) = 0$ for all $s, t \in \mathbb{Z}$ and thus $Y(s) = Y(t)$ $\mathbb{P}_{\boldsymbol{H}}$-almost surely.

Lemma 2.7. *The autocovariances of FGN have the following properties:*

(i) *If* $\boldsymbol{H} = \frac{1}{2}$, *then* $\rho_{\boldsymbol{H}}(k) = 0$ *for all* $k \in \mathbb{N}$.

(ii) *For every* $\boldsymbol{H} \in (0,1)$, *we have* $\rho_{\boldsymbol{H}}(k) \sim \boldsymbol{H}(2\boldsymbol{H}-1)\, k^{2\boldsymbol{H}-2}$.

(iii) *If* $\boldsymbol{H} < \frac{1}{2}$, *then* $\rho_{\boldsymbol{H}}(k) < \rho_{\boldsymbol{H}}(k+1) < 0$ *for all* $k \in \mathbb{N}$, *and*

$$\sum_{k=-\infty}^{\infty} \rho_{\boldsymbol{H}}(k) \;=\; 0\,.$$

(iv) *If* $I \subset (0, \frac{1}{2})$ *is compact, then there exists a function* $\nu : \mathbb{N} \to \mathbb{R}$ *satisfying* $\nu(1) > -1$ *and* $|\nu(k)| = o(k^{-\beta})$ *for some* $\beta > \frac{1}{2}$ *such that*

$$\rho_{\boldsymbol{H}}(k) \;\geq\; \nu(k)$$

for all $\boldsymbol{H} \in I$ *and* $k \in \mathbb{N}$. *In particular,* $\rho_{\boldsymbol{H}}(k) > -\frac{1}{2k}$ *for all* $\boldsymbol{H} \in (0, \frac{1}{2})$ *and* $k \in \mathbb{N}$.

Proof. (i) immediately follows by the definition of $\rho_{\boldsymbol{H}}(k)$.

(ii) See Proposition 3.1. (f) in Taqqu [96].

(iii) Let $\boldsymbol{H} < \frac{1}{2}$. Consider the mapping $\rho : [1, \infty) \to \mathbb{R}$ given by

$$\rho(x) \;:=\; (x+1)^{2\boldsymbol{H}} - 2x^{2\boldsymbol{H}} + (x-1)^{2\boldsymbol{H}}$$

for $x \in [1, \infty)$. Since $x \mapsto x^{2\boldsymbol{H}}$ is strictly concave on $(0, \infty)$ and continuous on $[0, \infty)$, we have $\rho(x) < 0$ for every $x \in [1, \infty)$, which shows that $\rho_{\boldsymbol{H}}(k) = \frac{1}{2}\rho(x) < 0$ for all $k \in \mathbb{N}$. Note that the first derivative of ρ on $(1, \infty)$ is given by

$$\rho'(x) \;=\; 2\boldsymbol{H}(x+1)^{2\boldsymbol{H}-1} - 4\boldsymbol{H}x^{2\boldsymbol{H}-1} + 2\boldsymbol{H}(x-1)^{2\boldsymbol{H}-1}$$

for $x \in (1,\infty)$. Since the mapping $x \mapsto x^{2\boldsymbol{H}-1}$ is strictly convex on $(0,\infty)$, we obtain $\rho'(x) > 0$ for every $x \in (1,\infty)$ and thus $\rho_{\boldsymbol{H}}(k) < \rho_{\boldsymbol{H}}(k+1)$ for all $k \in \mathbb{N}$. Finally, according to (ii), we have $\sum_{k=-\infty}^{\infty} |\rho_{\boldsymbol{H}}(k)| < \infty$ and hence

$$\begin{aligned}\sum_{k=-\infty}^{\infty} \rho_{\boldsymbol{H}}(k) &= \lim_{n\to\infty} \sum_{k=-n}^{n} \rho_{\boldsymbol{H}}(k) \\ &= \lim_{n\to\infty} \left((n+1)^{2\boldsymbol{H}} - n^{2\boldsymbol{H}}\right) = 0\,.\end{aligned}$$

(iv) For $(x,y) \in (0,1] \times I$, define

$$f(x,y) := \frac{1}{x^2}\left((1+x)^{2y} - 2 + (1-x)^{2y}\right).$$

Using the binomial series expansions of $(1+x)^{2y}$ and $(1-x)^{2y}$ (see Bronshtein and Semendyayev [20], p. 22), we obtain

$$\begin{aligned} f(x,y) &= \frac{1}{x^2}\left(\sum_{k=0}^{\infty}\binom{2y}{k}x^k - 2 + \sum_{k=0}^{\infty}\binom{2y}{k}(-1)^k x^k\right) \\ &= \sum_{k=1}^{\infty}\binom{2y}{k}x^{k-2} + \sum_{k=1}^{\infty}\binom{2y}{k}(-1)^k x^{k-2} \\ &= 2\sum_{l=0}^{\infty}\binom{2y}{2l+2}x^{2l}\,. \end{aligned}$$

The latter expression shows that, for every $y \in I$, the mapping $x \mapsto f(x,y)$ can be continuously continued to $[-1,1]$, namely, with $0^0 := 1$, the extension of f is given by $f(0,y) = 2y(2y-1)$. Since $x \mapsto f(x,y)$ is a power series, the partial derivative of f with respect to x is obtained by differentiating each term. Thus,

$$\frac{\partial f}{\partial x}(x,y) = 2\sum_{l=1}^{\infty} 2l\binom{2y}{2l+2}x^{2l-1}$$

for each interior point (x,y) of $[-1,1] \times I$. Since $I \subset (0,\frac{1}{2})$, we have

$$\binom{2y}{2l+2} = \frac{2y(2y-1)\ldots(2y-2l-1)}{(2l+2)!} < 0$$

for all $y \in I$ and $l \in \mathbb{N}$, which shows that $\frac{\partial f}{\partial x}$ is negative at each interior point (x,y) of $[0,1] \times I$. Consequently,

$$\min_{x\in[0,1]} f(x,y) = f(1,y)$$

for all $y \in I$. Now, let $\boldsymbol{H}_{\min} := \min I$ and $\boldsymbol{H}_{\max} := \max I$. According to the definition of f, we have $f(1, y) = 2^{2y} - 2$ for $y \in I$ and thus

$$\min_{(x,y)\in[0,1]\times I} f(x, y) \;=\; f(1, \boldsymbol{H}_{\min}) .$$

For $k \in \mathbb{N}$, define

$$\nu(k) \;:=\; \frac{1}{2} f(1, \boldsymbol{H}_{\min})\, k^{2\boldsymbol{H}_{\max}-2}$$

Since $f(1, \boldsymbol{H}_{\min}) < 0$, we obtain

$$\begin{aligned}
\rho_{\boldsymbol{H}}(k) &= \frac{1}{2}\left((k+1)^{2\boldsymbol{H}} - 2\,k^{2\boldsymbol{H}} + (k-1)^{2\boldsymbol{H}} \right) \\
&= \frac{1}{2} k^2 \left((1+k^{-1})^{2\boldsymbol{H}} - 2 + (1-k^{-1})^{2\boldsymbol{H}} \right) k^{2\boldsymbol{H}-2} \\
&= \frac{1}{2} f(k^{-1}, \mathbf{H})\, k^{2\boldsymbol{H}-2} \\
&\geq \nu(k)
\end{aligned}$$

for all $k \in \mathbb{N}$ and $\boldsymbol{H} \in I$. Note that $\nu(1) = \frac{1}{2} f(1, \boldsymbol{H}_{\min}) > -1$. Furthermore, because $\boldsymbol{H}_{\max} < \frac{1}{2}$, we have $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$. If $I = \{\boldsymbol{H}\}$ for some $\boldsymbol{H} \in (0, \frac{1}{2})$, then $\nu(k) = \frac{1}{2} f(1, \boldsymbol{H})\, k^{2\boldsymbol{H}-2}$ which shows that $\rho_{\boldsymbol{H}}(k) > -\frac{1}{2k}$ for all $k \in \mathbb{N}$. The proof is complete. □

Note that FGN can be considered as the increment process of Fractional Brownian Motion. Let $\mathbf{B} = (B(t))_{t\in\mathbb{R}}$ be a family of $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$-measurable mappings on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H}\in(0,1)}$. Suppose that, for $\boldsymbol{H} \in (0, 1)$, the following conditions are satisfied:

(B1) $\mathbf{B}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is zero-mean Gaussian.

(B2) The autocovariances of $\mathbf{B}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ are given by

$$\mathrm{Cov}_{\boldsymbol{H}}(B(s),\, B(t)) \;=\; \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right)$$

for $s, t \in \mathbb{R}$.

Then $\mathbf{B}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is called (*standard*) *Fractional Brownian Motion* (FBM) with the *Hurst parameter* $\boldsymbol{H}$.

In fact, such $(\Omega, \mathcal{A})$, $\mathbf{B}$ and $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H}\in(0,1)}$ do exist. For a proof of the existence of FBM with the Hurst parameter $\boldsymbol{H}$, see Proposition 2.2 in Taqqu [96]. Furthermore, for each $\boldsymbol{H}$, there exists a version of FBM with continuous sample paths (see Taqqu [96]). Thus, $(\Omega, \mathcal{A})$ can

be chosen as the set of continuous functions on $\mathbb{R}$ with the corresponding Borel σ-algebra, $\mathbf{B}$ as the identity on $(\Omega, \mathcal{A})$, and $\mathbb{P}_{\boldsymbol{H}}$ as the probability measure on $(\Omega, \mathcal{A})$ induced by the distribution of FBM with Hurst parameter $\boldsymbol{H}$. Note that some authors define FBM only for $t \in [0, \infty)$ (see, e.g., Embrechts and Maejima [36]). Here, we adopt the definition with double-side infinite time range of Taqqu [96].

Now, let $\mathbf{Y} = (Y(t))_{t \in \mathbb{Z}}$ be the process of first-order differences of $\mathbf{B}$, given by

$$Y(t) \;:=\; B(t) - B(t-1)$$

for $t \in \mathbb{Z}$. Then $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is standard FGN with the Hurst parameter $\boldsymbol{H}$ (see Taqqu [96]).

It is well-known that FBM with the Hurst parameter $\boldsymbol{H}$ is $\boldsymbol{H}$-*self-similar*, that is,

$$(B(at))_{t \in \mathbb{R}} \;\stackrel{\text{dist}}{=}\; (a^{\boldsymbol{H}} B(t))_{t \in \mathbb{R}} \tag{2.2}$$

for every $a > 0$ (see Embrechts and Maejima [36]).

Similarly as for FGN, one can also consider FBM with the Hurst parameter $\boldsymbol{H} = 1$. In this case, $\mathrm{Cov}_{\boldsymbol{H}}(B(s), B(t)) = \mathrm{Cov}_{\boldsymbol{H}}(sB(1), tB(1)) = st$ for $s, t \in \mathbb{R}$ which shows that the sample paths of $\mathbf{B}$ are almost surely straight lines, where the slope $B(1)$ is distributed according to $\mathrm{N}(0, 1)$.

2.2.4 ARFIMA(0,d,0) processes

Let $\mathbf{Y} = (Y(t))_{t \in \mathbb{Z}}$ be a family of $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$-measurable mappings on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{d}})_{\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})}$.

Assume that, for $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$, the following conditions are satisfied:

(A1) $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ is zero-mean Gaussian.

(A2) The autocovariances of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ are given by

$$\mathrm{Cov}_{\boldsymbol{d}}(Y(t),\, Y(t+k)) \;=\; \rho_{\boldsymbol{d}}(k)$$

for $t, k \in \mathbb{Z}$, where $\rho_{\boldsymbol{d}}(0) := 1$ and $\rho_{\boldsymbol{d}}(k) := 0$ for $k \in \mathbb{Z} \setminus \{0\}$ if $\boldsymbol{d} = 0$, and

$$\rho_{\boldsymbol{d}}(k) \;:=\; \frac{\Gamma(k + \boldsymbol{d})\,\Gamma(1 - \boldsymbol{d})}{\Gamma(k + 1 - \boldsymbol{d})\,\Gamma(\boldsymbol{d})}$$

for $k \in \mathbb{Z}$, otherwise.

Then $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ is called (*standard*) ARFIMA(0,d,0) *process* with the *fractional differencing parameter* $\boldsymbol{d}$.

Note that "ARFIMA" stands for *Auto-Regressive Fractionally Integrated Moving Average*. For the more general definition of ARFIMA(p,d,q) processes and a proof of their existence, see Taqqu [96]. The existence of $(\Omega, \mathcal{A})$, $\mathbf{Y}$ and $(\mathbb{P}_{\boldsymbol{d}})_{\boldsymbol{d}\in(-\frac{1}{2},\frac{1}{2})}$ follows by the same arguments as for FGN. According to Theorem 2.6 (i), the distribution of an ARFIMA(0,d,0) process is uniquely determined by (A1) and (A2). We refer to ARFIMA(0,d,0) processes as defined by (A1) and (A2) as *standard*, because $\mathrm{Var}_{\boldsymbol{d}}(Y(t)) = \rho_{\boldsymbol{d}}(0) = 1$ for all $t \in \mathbb{Z}$. In particular,

$$\rho_{\boldsymbol{d}}(k) \;=\; \mathrm{Corr}_{\boldsymbol{d}}(Y(t),\, Y(t+k))$$

for all $t, k \in \mathbb{Z}$. It is well-known that ARFIMA(0,d,0) processes with $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ are stationary (see Beran [15], p. 61). Furthermore, ARFIMA(0,d,0) processes have a spectral density (see Beran [15], p. 63). Thus, according to Theorem 2.6 (iv), ARFIMA(0,d,0) processes are non-degenerate for every $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$.

Lemma 2.8. *The autocovariances of an ARFIMA(0,d,0) process with the fractional differencing parameter $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ have the following properties:*

(i) *If $\boldsymbol{d} \neq 0$, then $\rho_{\boldsymbol{d}}(k) \sim \frac{\Gamma(1-\boldsymbol{d})}{\Gamma(\boldsymbol{d})}\, k^{2\boldsymbol{d}-1}$.*

(ii) *If $\boldsymbol{d} < 0$, then $\rho_{\boldsymbol{d}}(k) < \rho_{\boldsymbol{d}}(k+1) < 0$ for all $k \in \mathbb{N}$, and*

$$\sum_{k=-\infty}^{\infty} \rho_{\boldsymbol{d}}(k) \;=\; 0\,.$$

(iii) *If $I \subset (-\frac{1}{2}, 0)$ is compact, then there exists a function $\nu : \mathbb{N} \to \mathbb{R}$ with $\nu(1) > -1$ and $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ such that*

$$\rho_{\boldsymbol{d}}(k) \;\geq\; \nu(k)$$

for all $\boldsymbol{d} \in I$ and $k \in \mathbb{N}$.

Proof. (i) See Taqqu [96], Proposition 6.1.

(ii) Let $\boldsymbol{d} \in (-\frac{1}{2}, 0)$. It is well-known that $\Gamma(x) > 0$ for $x > 0$, $\Gamma(x) < 0$ for $x \in (-1, 0)$ and $\Gamma(1+x) = x\Gamma(x)$ for all $x \in \mathbb{R} \setminus \{0, -1, -2, \ldots\}$ (see Olver [78], pp. 32-35). Thus, $\rho_{\boldsymbol{d}}(k) < 0$ for all $k \in \mathbb{N}$. Furthermore, $\rho_{\boldsymbol{d}}(k+1) = \frac{k+\boldsymbol{d}}{k-\boldsymbol{d}+1}\, \rho_{\boldsymbol{d}}(k)$ for all $k \in \mathbb{N}$, which shows that $\rho_{\boldsymbol{d}}(k+1) > \rho_{\boldsymbol{d}}(k)$. By induction, one easily establishes

$$\frac{\Gamma(\boldsymbol{d})}{\Gamma(1-\boldsymbol{d})} \sum_{k=1}^{n} \rho_{\boldsymbol{d}}(k) \;=\; \frac{\Gamma(1-\boldsymbol{d})\,\Gamma(1+\boldsymbol{d}+n) - \boldsymbol{d}\,\Gamma(\boldsymbol{d})\,\Gamma(1-\boldsymbol{d}+n)}{2\boldsymbol{d}\,\Gamma(1-\boldsymbol{d})\,\Gamma(1-\boldsymbol{d}+n)}$$

for $n \in \mathbb{N}$. Similarly to (i), we obtain that

$$\frac{\Gamma(1+\boldsymbol{d}+n)}{\Gamma(1-\boldsymbol{d}+n)} \;=\; \frac{(\boldsymbol{d}+n)\Gamma(\boldsymbol{d}+n)}{\Gamma(1-\boldsymbol{d}+n)} \;\sim\; (\boldsymbol{d}+n)\, n^{2\boldsymbol{d}-1} \;\sim\; n^{2\boldsymbol{d}}$$

as $n \to \infty$, which shows that $\Gamma(1+\boldsymbol{d}+n)/\Gamma(1-\boldsymbol{d}+n) \to 0$ as $n \to \infty$ and thus $\sum_{k=1}^{\infty} \rho_{\boldsymbol{d}}(k) = -\frac{1}{2}$. Since $\sum_{k=-\infty}^{\infty} \rho_{\boldsymbol{d}}(k) = 1 + 2\sum_{k=1}^{\infty} \rho_{\boldsymbol{d}}(k)$, the statement follows.
(iii) Define $\boldsymbol{d}_{\min} := \min I$ and $\boldsymbol{d}_{\max} := \max I$. Furthermore, let

$$\kappa \;:=\; \max_{x \in I} \left| \frac{\Gamma(1-x)}{\Gamma(x)} \right|$$

Note that κ is well-defined because $x \mapsto \Gamma(x)$ is continuous on $\mathbb{R} \setminus \{0, -1, -2, \ldots\}$ (see Olver [78], p. 32) and I is compact. Now, define $\nu(1) := \frac{\boldsymbol{d}_{\min}}{1-\boldsymbol{d}_{\min}}$. Since $x \mapsto \frac{x}{1-x}$ is increasing on $(-\frac{1}{2}, 0)$, we have $-1 < \nu(1) \leq \rho_{\boldsymbol{d}}(1)$ for all $\boldsymbol{d} \in I$. For $k \in \mathbb{N}$ with $k > 1$, let

$$\nu(k) \;:=\; -\kappa\, \frac{\Gamma(k+\boldsymbol{d}_{\max})}{\Gamma(k+1-\boldsymbol{d}_{\max})}\,.$$

Since $x \mapsto \Gamma(x)$ is positive and increasing on $[\frac{3}{2}, \infty)$ (see Olver [78], p. 36), we obtain that $x \mapsto \frac{\Gamma(k+x)}{\Gamma(k+1-x)}$ is positive and increasing on $(-\frac{1}{2}, 0)$ for all $k \in \mathbb{N}$ with $k > 1$. Altogether, $\nu(k) \leq \rho_{\boldsymbol{d}}(k)$ for all $k \in \mathbb{N}$ and $\boldsymbol{d} \in I$. Similar to (i), we obtain

$$\frac{\Gamma(k+\boldsymbol{d}_{\max})}{\Gamma(k+1-\boldsymbol{d}_{\max})} \;\sim\; k^{2\boldsymbol{d}_{\max}-1}\,.$$

Since $\boldsymbol{d}_{\max} < 0$, we have $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$, and hence the proof is complete. □

2.2.5 AR(1) processes

Let $\mathbf{Y} = (Y(t))_{t \in \mathbb{Z}}$ be a family of $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$-measurable mappings on a measurable space $(\Omega, \mathcal{A})$, and let $(\Omega, \mathcal{A})$ be equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{a}})_{\boldsymbol{a} \in (-1,1)}$.
Assume that, for $\boldsymbol{a} \in (-1, 1)$, the following conditions are satisfied:

(A1) $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ is zero-mean Gaussian.

(A2) The autocovariances of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ are given by

$$\mathrm{Cov}_{\boldsymbol{a}}(Y(t), Y(t+k)) \;=\; \rho_{\boldsymbol{a}}(k) \;:=\; \boldsymbol{a}^{|k|}$$

for $t, k \in \mathbb{Z}$, where $0^0 := 1$.

Then $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ is called (*standard*) AR(1) *process* with the *autoregressive coefficient* $\boldsymbol{a}$.

The acronym "AR" stands for *Auto-Regressive*. For the more general definition of AR(p) processes and a proof of their existence, see Beran [15], p. 59. According to Theorem 2.6 (i), the distribution of AR(1) processes is uniquely determined by (A1) and (A2). We refer to an AR(1) process as defined by (A1) and (A2) as *standard*, because $\mathrm{Var}_{\boldsymbol{a}}(Y(t)) = \rho_{\boldsymbol{a}}(0) = 1$ for all $t \in \mathbb{Z}$. In particular,

$$\rho_{\boldsymbol{a}}(k) \;=\; \mathrm{Corr}_{\boldsymbol{a}}(Y(t), Y(t+k))$$

for all $t, k \in \mathbb{Z}$. Note that AR(1) processes with the autoregressive coefficient $\boldsymbol{a} \in (-1, 1)$ are stationary (see Beran [15], p. 59). Furthermore, AR(1) processes have a spectral density (see Beran [15], p. 61). Thus, according to Theorem 2.6 (iv), we obtain that AR(1) processes are non-degenerate for every $\boldsymbol{a} \in (-1, 1)$.

Lemma 2.9. *For every* $\boldsymbol{a} \in (-1, 1)$,

$$\sum_{k=1}^{\infty} |\rho_{\boldsymbol{a}}(k)| \;=\; \frac{|\boldsymbol{a}|}{1-|\boldsymbol{a}|} \qquad \textit{and} \qquad \sum_{k=-\infty}^{\infty} \rho_{\boldsymbol{a}}(k) \;=\; 1 + 2\,\frac{\boldsymbol{a}}{1-\boldsymbol{a}}\,.$$

Proof. The statement is obtained by the expression for the sum of a geometric series. □

Chapter 3

Estimation of ordinal pattern probabilities

In this chapter, we study the estimation of ordinal pattern probabilities in real-valued stochastic processes. Our work extends and partially generalizes the results given in Shiha [91] and Bandt and Shiha [12].

In Section 3.1, we introduce a parametric family of real-valued stochastic processes as the framework of our analysis. We assume that the increment processes are non-degenerate, stationary and zero-mean Gaussian with unit variance. In Section 3.2, we show that the distribution of ordinal patterns in such processes is stationary and any ordinal pattern has a strictly positive probability of occurrence.

Estimators of ordinal pattern probabilities are considered in Section 3.3. A simple unbiased estimator is given by the relative frequency of ordinal patterns in a sample. We show that, due to statistical symmetries of the increment processes, "better" estimators (in terms of the risk with respect to convex loss functions) are obtained by averaging the estimates of certain ordinal pattern probabilities. Conditions for strong consistency and asymptotic normality are established in Section 3.4. In Section 3.5, we apply the results to equidistant discretizations of FBM and to processes where the increments are ARFIMA(0,d,0) and AR(1).

3.1 Modelling

Throughout this chapter, the framework of our analysis is given by the following class of stochastic processes:

Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ a sequence of measurable mappings from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. Let $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ denote the process of increments of $\mathbf{X}$, given

by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$. Suppose $(\Omega, \mathcal{A})$ is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}}$ with $\boldsymbol{\Theta} \neq \emptyset$. The subscript $\boldsymbol{\vartheta}$ (for instance, $\mathbb{E}_{\boldsymbol{\vartheta}}$, $\mathrm{Var}_{\boldsymbol{\vartheta}}$, ...) indicates integration with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$. By saying that a certain property of $\mathbf{X}$ holds for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, we mean that this property holds for $\mathbf{X}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$.

We always assume that $\mathbf{Y}$ satisfies the following conditions:

(M1) $\mathbf{Y}$ is non-degenerate for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

(M2) $\mathbf{Y}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

(M3) $\mathbf{Y}$ is zero-mean Gaussian and has unit variance for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

As a consequence of (M1), the values of $(X_t)_{t \in \mathbb{Z}}$ are pairwise different $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. To see this, note that the set $\{(y_1, y_2, \ldots, y_k) \in \mathbb{R}^k \mid y_1 + y_2 + \ldots + y_k = 0\}$ has Lebesgue measure 0 for every $k \in \mathbb{N}$. Therefore, for all $t \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\vartheta}}(X_t \neq X_{t+k}) &= 1 - \mathbb{P}_{\boldsymbol{\vartheta}}(X_t = X_{t+k}) \\ &= 1 - \mathbb{P}_{\boldsymbol{\vartheta}}(Y_{t+1} + Y_{t+2} + \ldots + Y_{t+k} = 0) \\ &= 1\,. \end{aligned} \tag{3.1}$$

For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z}$, define

$$\rho_{\boldsymbol{\vartheta}}(k) := \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0, Y_k)\,. \tag{3.2}$$

Since $\mathbf{Y}$ is non-degenerate, we have

$$|\rho_{\boldsymbol{\vartheta}}(k)| < 1$$

for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z} \setminus \{0\}$ (see Theorem 2.6 (iv)).

The class of stochastic processes satisfying (M1)-(M3) includes, for example, standard Fractional Gaussian Noise with the Hurst parameter $\boldsymbol{H} \in (0, 1)$ (see Section 2.2.3), standard ARFIMA(0,d,0) processes with the fractional differencing parameter $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ (see Section 2.2.4), and standard AR(1) processes with the autoregressive coefficient $\mathbf{a} \in (-1, 1)$ (see Section 2.2.5).

As we will see below, the distribution of ordinal patterns does not depend on the variance of $\mathbf{Y}$ (as long as the variance is strictly positive, which is necessary for $\mathbf{Y}$ to be non-degenerate). The assumption that $\mathbf{Y}$ has unit variance is for sake of convenience. In particular, the autocovariances of $\mathbf{Y}$ are equal to the autocorrelations in this case.

The reason why we consider Gaussian processes is that there exist simple closed-form expressions for ordinal pattern probabilities in this case, and the variance of estimators

can be easily evaluated numerically. As we show in the following section, all statements on ordinal patterns and their distributions are valid for a larger class of processes. In particular, $(X_t)_{t\in\mathbb{Z}}$ can be replaced by the transformed process $(h(X_t))_{t\in\mathbb{Z}}$ where $h : \mathbb{R} \to \mathbb{R}$ is strictly monotonically increasing. Note that, in general, the increments of the transformed process $(h(X_t))_{t\in\mathbb{Z}}$ are not Gaussian.

3.2 Ordinal pattern probabilities

Ordinal patterns. We start with the main definition of this thesis. For $d \in \mathbb{N}$, let S_d denote the set of permutations of $\{0, 1, \ldots, d\}$, which we write as $(d+1)$-tuples containing each of the numbers $0, 1, \ldots, d$ exactly once. For instance, $S_1 = \{(0, 1), (1, 0)\}$, $S_2 = \{(0, 1, 2), (1, 0, 2), (1, 2, 0), (2, 0, 1), (0, 2, 1), (2, 1, 0)\}$, and so on. Ordinal patterns represent the order relations among a fixed number of equidistant values in a time series. If we assume that the values are pairwise different, then it is natural to identify ordinal patterns with permutations.

Let $d \in \mathbb{N}$. For $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$, let

$$\pi(\mathbf{x}) \;=\; (r_0, r_1, \ldots, r_d)$$

be the unique permutation of $\{0, 1, \ldots, d\}$ which satisfies

$$x_{r_0} \;\geq\; x_{r_1} \;\geq\; \ldots \;\geq\; x_{r_d} \tag{3.3}$$

and

$$r_{i-1} > r_i \text{ if } x_{r_{i-1}} = x_{r_i} \quad \text{for } i = 1, 2, \ldots, d. \tag{3.4}$$

Condition (3.4) is necessary to guarantee the uniqueness of $(r_0, r_1, \ldots, r_d)$ if there are equal values among $x_0, x_1, \ldots, x_d$.

We may regard $\pi(\mathbf{x})$ as a representation of the rank order of $x_0, x_1, \ldots, x_d$. If $x_i = x_j$ for $i, j \in \{0, 1, \ldots, d\}$ with $i < j$, then x_j is ranked higher than x_i. When $x_0, x_1, \ldots, x_d$ are pairwise different, then the order relation between any two components of $\mathbf{x}$ (being either $<$ or $>$) can be obtained from $\pi(\mathbf{x})$.

Figure 3.1 displays the components $x_0, x_1, \ldots, x_5$ of a vector $\mathbf{x} \in \mathbb{R}^6$. The horizontal alignment and the connecting line segments suggest that $x_0, x_1, \ldots, x_5$ are observed sequently in time. As emphasized by the dashed lines, $x_3 = 0.9$ is the largest value, $x_4 = 0.7$ is the second-largest value, and so on. This rank order is uniquely described by the permutation $\pi(\mathbf{x}) = (3, 4, 5, 2, 0, 1)$. Since there are no equal values among the components of $\mathbf{x}$, condition (3.4) does not apply here.

Now, let $d \in \mathbb{N}$ and $t \in \mathbb{Z}$. By the *(random) ordinal pattern of order d at time t in* $\mathbf{X}$, we mean the random permutation

$$\Pi(t) \;:=\; \pi(X_t, X_{t+1}, \ldots, X_{t+d})\,.$$

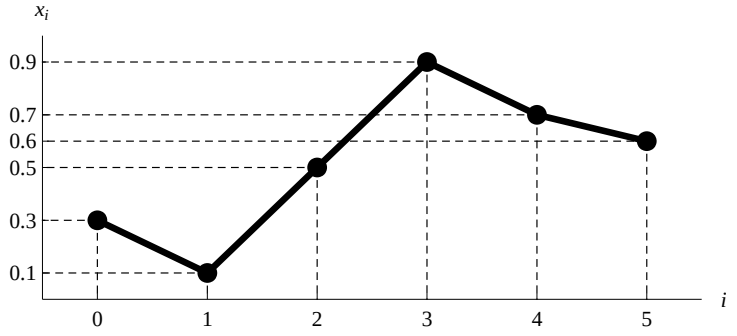

Figure 3.1: A vector $\mathbf{x} = (x_0, x_1, \ldots, x_5) \in \mathbb{R}^6$ for which $\pi(\mathbf{x}) = (3, 4, 5, 2, 0, 1)$.

According to (3.1), the random variables $X_t, X_{t+1}, \ldots, X_{t+d}$ are pairwise different $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Therefore, for every $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$, the events $\{\Pi(t) = \mathbf{r}\}$ and $\{X_{t+r_0} > X_{t+r_1} > \ldots > X_{t+r_d}\}$ are equivalent $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely (that is, the symmetric difference of both events has probability 0 with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$).

In this chapter, we study the distribution of the ordinal pattern process $(\Pi(t))_{t\in\mathbb{Z}}$ and the estimation of ordinal pattern probabilities. In Chapter 5, we focus on ordinal patterns of order $d = 2$. In Chapter 6, we introduce *delays* of ordinal patterns as an additional parameter besides the order.

Clearly, if $h : \mathbb{R} \to \mathbb{R}$ is strictly monotonically increasing, then

$$\pi(\mathbf{x}) \;=\; \pi(h(x_0), h(x_1), \ldots, h(x_d)) \tag{3.5}$$

for all $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$. Thus, the ordinal patterns in $(X_t)_{t\in\mathbb{Z}}$ and in $(h(X_t))_{t\in\mathbb{Z}}$ are identical. Note that this statement is also true when h depends on $\omega \in \Omega$. For instance, if A and B are random "offsets" and "scalings", that is, measurable mappings from $(\Omega, \mathcal{A})$ into $\mathbb{R}$ and $(0, \infty)$, respectively, then the ordinal patterns in $(X_t)_{t\in\mathbb{Z}}$ and in $(A + B \cdot X_t)_{t\in\mathbb{Z}}$ are identical.

If $h(x) = cx$ for some $c > 0$, then $(h(X_t))_{t\in\mathbb{Z}}$ has the increments $h(X_t) - h(X_{t-1}) = cY_t$ for $t \in \mathbb{Z}$. Thus, according to the model assumptions (M1)-(M3), the process of increments is non-degenerate, stationary, zero-mean Gaussian and has variance $\mathrm{Var}_{\boldsymbol{\vartheta}}(cY_t) = c^2$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. This argument shows that the distribution of ordinal patterns actually does not depend on the variance of the increment process.

Note that we could have defined $\Pi(t)$ as a causal filter $\Pi(t) = \pi(X_{t-d}, X_{t-d+1}, \ldots, X_t)$ only depending on the "past" of $\mathbf{X}$ at time t. The above "non-causal" definition is just for the sake of simpler notation in some proofs.

Stationarity. Next we show that, as a consequence of the model assumption (M2), the ordinal pattern process $(\Pi(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Let $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$. Since shifting a vector by constant does not affect any order relation among its components, we have

$$\pi(\mathbf{x}) \;=\; \pi(x_0 - x_0,\, x_1 - x_0,\, \ldots,\, x_d - x_0)\,.$$

Furthermore, for each $i \in \{1, 2, \ldots, d\}$, we can write $x_i - x_0$ as the sum of the differences $x_1 - x_0$, $x_2 - x_1$, $\ldots$, $x_i - x_{i-1}$. Therefore, with $\tilde{\pi}$ defined by

$$\tilde{\pi}(\mathbf{y}) \;:=\; \pi(0,\, y_1,\, y_1 + y_2,\, \ldots,\, y_1 + y_2 + \ldots + y_d)\,.$$

for $\mathbf{y} = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d$, we obtain

$$\pi(\mathbf{x}) \;=\; \tilde{\pi}(x_1 - x_0,\, x_2 - x_1,\, \ldots,\, x_d - x_{d-1})\,.$$

This shows that, for every $t \in \mathbb{Z}$, the ordinal pattern $\Pi(t)$ only depends on the increments $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+d}$, namely,

$$\Pi(t) \;=\; \tilde{\pi}(Y_{t+1},\, Y_{t+2},\, \ldots,\, Y_{t+d})\,.$$

Now, the following corollary is an immediate consequence of the model assumption (M2) that $\mathbf{Y}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Corollary 3.1. *The process $(\Pi(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.*

Note that Corollary 3.1 relaxes the condition for stationarity of $(\Pi(t))_{t\in\mathbb{Z}}$ given in Shiha [91], which requires the increments to be stationary and independent.
Let $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$. For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, define

$$p_{\mathbf{r}}(\boldsymbol{\vartheta}) \;:=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = \mathbf{r})\,.$$

According to Corollary 3.1, the function $p_{\mathbf{r}}(\cdot)$ does not depend on the specific time point $t \in \mathbb{Z}$ on the right hand side of the definition. We call $p_{\mathbf{r}}(\cdot)$ the *probability of the ordinal pattern* $\mathbf{r}$. The following corollary shows that the probability of any ordinal pattern is strictly positive.

Corollary 3.2. *For every $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$ with $d \in \mathbb{N}$ and every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$,*

$$0 \;<\; p_{\mathbf{r}}(\boldsymbol{\vartheta}) \;<\; 1\,.$$

Proof. Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. We only need to show that $p_{\mathbf{r}}(\boldsymbol{\vartheta}) > 0$. Since there exists an $\mathbf{s} \in S_d$ with $\mathbf{s} \neq \mathbf{r}$, it then follows that $p_{\mathbf{r}}(\boldsymbol{\vartheta}) \leq 1 - p_{\mathbf{s}}(\boldsymbol{\vartheta}) < 1$. Note that

$$\begin{aligned} p_{\mathbf{r}}(\boldsymbol{\vartheta}) &= \mathbb{P}_{\boldsymbol{\vartheta}}(X_{r_0} > X_{r_1} > \ldots > X_{r_d}) \\ &= \mathbb{P}_{\boldsymbol{\vartheta}}(X_{r_0} - X_{r_1} > 0,\, X_{r_1} - X_{r_2} > 0,\, \ldots,\, X_{r_{d-1}} - X_{r_d} > 0)\,. \end{aligned}$$

Thus, in order to prove that $p_{\mathbf{r}}(\boldsymbol{\vartheta}) > 0$, it suffices to show that the random vector $(X_{r_0} - X_{r_1},\, X_{r_1} - X_{r_2},\, \ldots,\, X_{r_{d-1}} - X_{r_d})$ is non-degenerate Gaussian. Let $k \in \{1, 2, \ldots, d\}$ and suppose $i, j \in \{0, 1, \ldots, d\}$ are the indices for which $r_i = k$ and $r_j = k - 1$. If $i < j$, then

$$Y_k \;=\; (X_{r_i} - X_{r_{i+1}}) + (X_{r_{i+1}} - X_{r_{i+2}}) + \ldots + (X_{r_{j-1}} - X_{r_j})\,,$$

and if $i > j$, we have

$$Y_k \;=\; -(X_{r_{i-1}} - X_{r_i}) - (X_{r_{i-2}} - X_{r_{i-1}}) - \ldots - (X_{r_{j-1}} - X_{r_j})\,.$$

Consequently, there exists a matrix $\mathbf{B} \in \mathbb{R}^{d \times d}$ such that

$$(Y_1,\, Y_2,\, \ldots,\, Y_d)^T \;=\; \mathbf{B}\,(X_{r_0} - X_{r_1},\, X_{r_1} - X_{r_2},\, \ldots,\, X_{r_{d-1}} - X_{r_d})^T\,.$$

Now, let $\boldsymbol{\Sigma}$ be the covariance matrix of $(Y_1,\, Y_2,\, \ldots,\, Y_d)$ and $\tilde{\boldsymbol{\Sigma}}$ the covariance matrix of $(X_{r_0} - X_{r_1},\, X_{r_1} - X_{r_2},\, \ldots,\, X_{r_{d-1}} - X_{r_d})$. Since $\mathbf{Y}$ is non-degenerate, $\boldsymbol{\Sigma}$ is strictly positive definite. Furthermore, $\boldsymbol{\Sigma} = \mathbf{B}\,\tilde{\boldsymbol{\Sigma}}\,\mathbf{B}^T$ which shows that $\mathbf{B}$ is non-singular and $\tilde{\boldsymbol{\Sigma}}$ is strictly positive definite. Consequently, $(X_{r_0} - X_{r_1},\, X_{r_1} - X_{r_2},\, \ldots,\, X_{r_{d-1}} - X_{r_d})$ is non-degenerate Gaussian, and hence the proof is complete. □

Bandt and Shiha [12] gives closed-form expressions for the probability of ordinal patterns of order $d = 2$ and $d = 3$. We will review the results for $d = 2$ in Chapter 5.

3.3 Estimators of ordinal pattern probabilities

Let $d \in \mathbb{N}$ and $\mathbf{r} \in S_d$. Furthermore, let $n \in \mathbb{N}$ and define

$$\boldsymbol{\Pi}_n \;:=\; (\Pi(0),\, \Pi(1),\, \ldots,\, \Pi(n-1))\,.$$

In this section, we study the problem of estimating the ordinal pattern probability $p_{\mathbf{r}}(\boldsymbol{\vartheta})$ given an observation of the ordinal pattern sample $\boldsymbol{\Pi}_n$ governed by $\mathbb{P}_{\boldsymbol{\vartheta}}$ with $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ unknown. A natural estimator of $p_{\mathbf{r}}(\cdot)$ is the relative frequency of observations of $\mathbf{r}$ in $\boldsymbol{\Pi}_n$, namely,

$$\hat{q}_{\mathbf{r},n} \;=\; \hat{q}_{\mathbf{r},n}(\boldsymbol{\Pi}_n) \;:=\; \frac{1}{n}\sum_{t=0}^{n-1} \mathbf{1}_{\{\Pi(t)=\mathbf{r}\}}\,.$$

According to Corollary 3.1, we have

$$\mathbb{E}_{\boldsymbol{\vartheta}}(\hat{q}_{\mathbf{r},n}) \;=\; p_{\mathbf{r}}(\boldsymbol{\vartheta})$$

for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, that is, $\hat{q}_{\mathbf{r},n}$ is an unbiased estimator of $p_{\mathbf{r}}(\cdot)$. Next we show that, due to statistical symmetries of the increment process $\mathbf{Y}$, there is a simple way for improving this estimator.

Space and time symmetry. Let $n \in \mathbb{N}$ and $t_1 < t_2 < \ldots < t_n \in \mathbb{Z}$. According to the model assumption (M3), both $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ and $(-Y_{t_1}, -Y_{t_2}, \ldots, -Y_{t_n})$ are zero-mean Gaussian for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Furthermore, because $\mathrm{Cov}_{\boldsymbol{\vartheta}}(Y_{t_i}, Y_{t_j}) = \mathrm{Cov}_{\boldsymbol{\vartheta}}(-Y_{t_i}, -Y_{t_j})$ for all $i, j \in \{1, 2, \ldots, n\}$, they have the same covariance matrix and thus they are identically distributed. Consequently,

$$(Y_t)_{t\in\mathbb{Z}} \overset{\text{dist}}{=} (-Y_t)_{t\in\mathbb{Z}} \tag{3.6}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Now, note that, because $\mathbf{Y}$ is stationary (see model assumption (M2)), we have $\mathrm{Cov}_{\boldsymbol{\vartheta}}(Y_{t_i}, Y_{t_j}) = \mathrm{Cov}_{\boldsymbol{\vartheta}}(Y_{-t_i}, Y_{-t_j})$ for all $i, j \in \{1, 2, \ldots, n\}$, which shows that $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ and $(Y_{-t_1}, Y_{-t_2}, \ldots, Y_{-t_n})$ are identically distributed. Consequently,

$$(Y_t)_{t\in\mathbb{Z}} \overset{\text{dist}}{=} (Y_{-t})_{t\in\mathbb{Z}} \tag{3.7}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. We refer to the properties (3.6) and (3.7) of $\mathbf{Y}$ as *symmetry in space* and *time*, respectively. Note that it is well-known that zero-mean stationary Gaussian processes have these properties. According to the terminology of Bandt and Shiha [12], symmetry in space is equivalent to reversibility, and symmetry in time is equivalent to rotation symmetry.

Next we show that, as a consequence of (3.6) and (3.7), the distribution of $\boldsymbol{\Pi}_n$ is invariant with respect to spatial and time reversals of ordinal pattern sequences.

Let the mappings α, β from S_d onto itself be defined by

$$\alpha(\mathbf{r}) := (r_d,\, r_{d-1},\, \ldots,\, r_0) \;\text{ and }\; \beta(\mathbf{r}) := (d - r_0,\, d - r_1,\, \ldots,\, d - r_d) \tag{3.8}$$

for $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$. Geometrically, we may regard $\alpha(\mathbf{r})$ and $\beta(\mathbf{r})$ as the spatial and time reversal of $\mathbf{r}$ (for an illustration, see Figure 3.2). In particular, if the components of $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$ are pairwise different, then

$$\alpha(\pi(\mathbf{x})) = \pi(-x_0, -x_1, \ldots, -x_d) \;\text{ and }\; \beta(\pi(\mathbf{x})) = \pi(x_d, x_{d-1}, \ldots, x_0)\,. \tag{3.9}$$

Thus, $\alpha(\pi(\mathbf{x}))$ and $\beta(\pi(\mathbf{x}))$ represent the rank orders of the vectors obtained by reversing the components of $\mathbf{x}$ in "space" and "time", respectively. In terms of the vector of increments $\mathbf{y} = (y_1, y_2, \ldots, y_d)$ given by $y_k := x_k - x_{k-1}$ for $k = 1, 2, \ldots, d$, we can write (3.9) as

$$\alpha(\tilde{\pi}(\mathbf{y})) = \tilde{\pi}(-y_1, -y_2, \ldots, -y_d) \;\text{ and }\; \beta(\tilde{\pi}(\mathbf{y})) = \tilde{\pi}(-y_d, -y_{d-1}, \ldots, -y_1)\,. \tag{3.10}$$

For $\mathbf{r} \in S_d$, consider the subset $\bar{\mathbf{r}}$ of S_d defined by

$$\bar{\mathbf{r}} := \big\{\mathbf{r},\, \alpha(\mathbf{r}),\, \beta(\mathbf{r}), \beta \circ \alpha(\mathbf{r})\big\}\,.$$

Since $\alpha \circ \beta(\mathbf{r}) = \beta \circ \alpha(\mathbf{r})$ and $\alpha \circ \alpha(\mathbf{r}) = \beta \circ \beta(\mathbf{r}) = \mathbf{r}$, the set $\bar{\mathbf{r}}$ is closed under α and β, that is, $\alpha(\bar{\mathbf{r}}) = \beta(\bar{\mathbf{r}}) = \bar{\mathbf{r}}$. Consequently, if $\mathbf{s} \in \bar{\mathbf{r}}$ for $\mathbf{r}, \mathbf{s} \in S_d$, then $\bar{\mathbf{s}} = \bar{\mathbf{r}}$. This

$\mathbf{r} = (0,2,1)$ $\quad \alpha(\mathbf{r}) = (1,2,0)$ $\quad \beta(\mathbf{r}) = (2,0,1)$ $\quad \beta \circ \alpha(\mathbf{r}) = (1,0,2)$

Figure 3.2: The spatial and time reversals of $\mathbf{r} = (0,2,1)$.

provides a partition of each S_d into classes which contain 2 or 4 elements. For $d = 1$, the only class is $S_1 = \{(0,1),(1,0)\}$; for $d = 2$, there are two classes: $\{(0,1,2),(2,1,0)\}$ and $\{(0,2,1),(2,0,1),(1,2,0),(1,0,2)\}$; for $d = 3$, there are 8 classes. In general, for $d \geq 2$, both classes of 2 and 4 elements are possible, for instance, $\sharp\,\bar{\mathbf{r}} = 2$ if $\mathbf{r} = (0,1,\ldots,d)$ and $\sharp\,\bar{\mathbf{r}} = 4$ if $\mathbf{r} = (0,2,3,\ldots,d,1)$.

Now, let $n \in \mathbb{N}$. Consider the mappings A, B from $(S_d)^n$ onto $(S_d)^n$ defined by

$$\begin{aligned} A(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n) &:= (\alpha(\mathbf{r}_1),\, \alpha(\mathbf{r}_2),\, \ldots,\, \alpha(\mathbf{r}_n)), \\ B(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n) &:= (\beta(\mathbf{r}_n),\, \beta(\mathbf{r}_{n-1}),\, \ldots,\, \beta(\mathbf{r}_1)) \end{aligned}$$

for $(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n) \in (S_d)^n$. According to the geometrical interpretation of α and β, the ordinal pattern sequences $A(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n)$ and $B(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n)$ can be regarded as the spatial and time reversal of the ordinal pattern sequence $(\mathbf{r}_1,\, \mathbf{r}_2,\, \ldots,\, \mathbf{r}_n)$.

Lemma 3.3. *For every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$,*

$$\boldsymbol{\Pi}_n \overset{\text{dist}}{=} A(\boldsymbol{\Pi}_n) \overset{\text{dist}}{=} B(\boldsymbol{\Pi}_n) \overset{\text{dist}}{=} B \circ A(\boldsymbol{\Pi}_n)\,.$$

Proof. Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Since the values in $\mathbf{X}$ are pairwise different $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely (see (3.1)), the first equation in (3.10) yields

$$\alpha(\tilde{\pi}(Y_{t+1}, Y_{t+2}, \ldots, Y_{t+d})) = \tilde{\pi}(-Y_{t+1}, -Y_{t+2}, \ldots, -Y_{t+d})$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $t \in \mathbb{Z}$. Furthermore, according to the space symmetry of $\mathbf{Y}$, the random vectors $(Y_1, Y_2, \ldots, Y_{n+d-1})$ and $(-Y_1, -Y_2, \ldots, -Y_{n+d-1})$ have the same distribution with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$. Thus,

$$\begin{aligned} \boldsymbol{\Pi}_n &= \big(\tilde{\pi}(Y_1, \ldots, Y_d),\, \tilde{\pi}(Y_2, \ldots, Y_{d+1}),\, \ldots,\, \tilde{\pi}(Y_n, \ldots, Y_{n+d-1})\big) \\ &\overset{\text{dist}}{=} \big(\tilde{\pi}(-Y_1, \ldots, -Y_d),\, \tilde{\pi}(-Y_2, \ldots, -Y_{d+1}),\, \ldots,\, \tilde{\pi}(-Y_n, \ldots, -Y_{n+d-1})\big) \\ &= A(\boldsymbol{\Pi}_n)\,, \end{aligned} \tag{3.11}$$

where the last equality holds $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Similarly, we obtain that

$$\beta(\tilde{\pi}(Y_{t+1}, Y_{t+2}, \ldots, Y_{t+d})) = \tilde{\pi}(-Y_{t+d}, -Y_{t+d-1}, \ldots, -Y_{t+1})$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $t \in \mathbb{Z}$. By the space and time symmetry of $\mathbf{Y}$, we obtain that $(Y_1, Y_2, \ldots, Y_{n+d-1})$ and $(-Y_{n+d-1}, -Y_{n+d-2}, \ldots, -Y_1)$ have the same distribution with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$, and thus

$$\begin{aligned} \boldsymbol{\Pi}_n &= \big(\tilde{\pi}(Y_1, \ldots, Y_d),\, \tilde{\pi}(Y_2, \ldots, Y_{d+1}),\, \ldots,\, \tilde{\pi}(Y_n, \ldots, Y_{n+d-1})\big) \\ &\overset{\text{dist}}{=} \big(\tilde{\pi}(-Y_{n+d-1}, \ldots, -Y_n),\, \tilde{\pi}(-Y_{n+d-2}, \ldots, -Y_{n-1}),\, \ldots,\, \tilde{\pi}(-Y_d, \ldots, -Y_1)\big) \\ &= B(\boldsymbol{\Pi}_n)\,, \end{aligned} \tag{3.12}$$

where the last equality holds $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Now, combining (3.11) and (3.12) yields equality in distribution of $\boldsymbol{\Pi}_n$ and $B \circ A(\boldsymbol{\Pi}_n)$. □

For the proof of Lemma 3.3 we have only used that $\mathbf{Y}$ is symmetric in space and time and the values of $\mathbf{X}$ are pairwise different $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Thus, the statement that $\boldsymbol{\Pi}_n$, $A(\boldsymbol{\Pi}_n)$, $B(\boldsymbol{\Pi}_n)$ and $B \circ A(\boldsymbol{\Pi}_n)$ have the same distribution is valid under much more general conditions than (M1)-(M3).

In the following paragraph, we show that when $\boldsymbol{\Pi}_n$, $A(\boldsymbol{\Pi}_n)$, $B(\boldsymbol{\Pi}_n)$ and $B \circ A(\boldsymbol{\Pi}_n)$ have the same distribution, a better estimator of $p_{\mathbf{r}}(\cdot)$ than $\hat{q}_{\mathbf{r},n}$ is obtained by averaging the relative frequencies of $\mathbf{r}$, $\alpha(\mathbf{r})$, $\beta(\mathbf{r})$ and $\beta \circ \alpha(\mathbf{r})$. This finding is of great practical importance, as many empirical time series seem to be realizations of processes for which the distributions of ordinal pattern sequences and their spatial and time reversals are identical. For instance, Bandt and Shiha [12] reports a similar frequency of ordinal patterns and their time reversals in speech signals and EEG data. Also, Keller et al. [58] observes a similar frequency of ordinal patterns and their space reversals in EEG data. As an exception, Keller et al. [58] shows "asymmetric" EEG time series with more "upwards" than "downwards" patterns.

A Rao-Blackwellization. Let $d \in \mathbb{N}$, $\mathbf{r} \in S_d$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. For $n = 1$, Lemma 3.3 shows that $\Pi(0)$, $\alpha(\Pi(0))$, $\beta(\Pi(0))$ and $\beta \circ \alpha(\Pi(0))$ have the same distribution with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$, and thus

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(0) = \mathbf{r}) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\alpha(\Pi(0)) = \mathbf{r}) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\beta(\Pi(0)) = \mathbf{r}) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\beta \circ \alpha(\Pi(0)) = \mathbf{r})\,.$$

According to the properties of α and β, the events $\{\alpha(\Pi(0)) = \mathbf{r}\}$, $\{\beta(\Pi(0)) = \mathbf{r}\}$ and $\{\beta \circ \alpha(\Pi(0)) = \mathbf{r}\}$ are equivalent to $\{\Pi(0) = \alpha(\mathbf{r})\}$, $\{\Pi(0) = \beta(\mathbf{r})\}$ and $\{\Pi(0) = \alpha \circ \beta(\mathbf{r})\}$, respectively. Therefore,

$$p_{\mathbf{r}}(\cdot) \;=\; p_{\alpha(\mathbf{r})}(\cdot) \;=\; p_{\beta(\mathbf{r})}(\cdot) \;=\; p_{\alpha \circ \beta(\mathbf{r})}(\cdot)$$

which shows that $\hat{q}_{\mathbf{r},n}$, $\hat{q}_{\alpha(\mathbf{r}),n}$, $\hat{q}_{\beta(\mathbf{r}),n}$ and $\hat{q}_{\alpha\circ\beta(\mathbf{r}),n}$ are all unbiased estimators of $p_{\mathbf{r}}(\cdot)$. By averaging them, we obtain another unbiased estimator of $p_{\mathbf{r}}(\cdot)$, namely,

$$\begin{aligned}
\hat{p}_{\mathbf{r},n} \;=\; \hat{p}_{\mathbf{r},n}(\mathbf{\Pi}_n) \;&:=\; \frac{1}{4}\Big(\hat{q}_{\mathbf{r},n}(\mathbf{\Pi}_n) + \hat{q}_{\alpha(\mathbf{r}),n}(\mathbf{\Pi}_n) + \hat{q}_{\beta(\mathbf{r}),n}(\mathbf{\Pi}_n) + \hat{q}_{\alpha\circ\beta(\mathbf{r}),n}(\mathbf{\Pi}_n)\Big) \\
&=\; \frac{1}{4n}\sum_{t=0}^{n-1}\Big(\mathbf{1}_{\{\Pi(t)=\mathbf{r}\}} + \mathbf{1}_{\{\Pi(t)=\alpha(\mathbf{r})\}} + \mathbf{1}_{\{\Pi(t)=\beta(\mathbf{r})\}} + \mathbf{1}_{\{\Pi(t)=\alpha\circ\beta(\mathbf{r})\}}\Big) \\
&=\; \frac{1}{(\sharp\,\bar{\mathbf{r}})\,n}\sum_{t=0}^{n-1}\mathbf{1}_{\{\Pi(t)\in\bar{\mathbf{r}}\}}\,. \qquad (3.13)
\end{aligned}$$

In Theorem 3.5, we show that the estimator $\hat{p}_{\mathbf{r},n}$ has lower risk than $\hat{q}_{\mathbf{r},n}$ with respect to any convex loss function. The key step in the proof is to establish a sufficient statistic for $\mathbf{\Pi}_n$ which identifies $\mathbf{\Pi}_n$, $A(\mathbf{\Pi}_n)$, $B(\mathbf{\Pi}_n)$ and $B\circ A(\mathbf{\Pi}_n)$. Then, according to the *Sufficiency Principle*, the inference on $p_{\mathbf{r}}(\cdot)$ should be the same regardless of whether $\mathbf{\Pi}_n$, $A(\mathbf{\Pi}_n)$, $B(\mathbf{\Pi}_n)$ or $B\circ A(\mathbf{\Pi}_n)$ is observed (see Casella and Berger [25], p. 272).

First, we establish the following lemma.

Lemma 3.4. *For every* $\boldsymbol{\vartheta}\in\mathbf{\Theta}$,

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}\neq\hat{q}_{\mathbf{r},n}) \;>\; 0\,.$$

Proof. Let $\boldsymbol{\vartheta}\in\mathbf{\Theta}$. We show the existence of a permutation $(s_0,s_1,\ldots,s_{n+d-1})\in S_{n+d-1}$ such that $X_{s_0}>X_{s_1}>\ldots>X_{s_{n+d-1}}$ implies $\hat{p}_{\mathbf{r},n}>0$ and $\hat{q}_{\mathbf{r},n}=0$. Then, according to Corollary 3.2,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}\neq\hat{q}_{\mathbf{r},n}) \;&\geq\; \mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}>0,\,\hat{q}_{\mathbf{r},n}=0) \\
&\geq\; \mathbb{P}_{\boldsymbol{\vartheta}}(X_{s_0}>X_{s_1}>\ldots>X_{s_{n+d-1}}) \\
&>\; 0\,.
\end{aligned}$$

Let $i,j\in\{0,1,\ldots,d\}$ be such that $r_i=d-1$ and $r_j=d$. If $i<j$, then we choose

$$(s_0,s_1,\ldots,s_{n+d-1}) \;=\; (n+d-1,n+d-2,\ldots,d+1,r_d,r_{d-1},\ldots,r_0)\,.$$

Otherwise, we choose

$$(s_0,s_1,\ldots,s_{n+d-1}) \;=\; (r_d,r_{d-1},\ldots,r_0,d+1,d+2,\ldots,n+d-1)\,.$$

In both cases, $X_{s_0}>X_{s_1}>\ldots>X_{s_{n+d-1}}$ implies that $\Pi(0)=\alpha(\mathbf{r})$ and $\Pi(t)\neq\mathbf{r}$ for $t=1,2,\ldots,n-1$, and thus $\hat{p}_{\mathbf{r},n}>0$ and $\hat{q}_{\mathbf{r},n}=0$. The proof is complete. □

Theorem 3.5. *The estimator $\hat{p}_{\mathbf{r},n}$ of $p_{\mathbf{r}}(\cdot)$ is unbiased and has lower risk than $\hat{q}_{\mathbf{r},n}$ with respect to any convex loss function, that is, for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$,*

$$\mathbb{E}_{\boldsymbol{\vartheta}}\big(\,\varphi(\hat{p}_{\mathbf{r},n},\, p_{\mathbf{r}}(\boldsymbol{\vartheta}))\,\big) \;\leq\; \mathbb{E}_{\boldsymbol{\vartheta}}\big(\,\varphi(\hat{q}_{\mathbf{r},n},\, p_{\mathbf{r}}(\boldsymbol{\vartheta}))\,\big)$$

with respect to each function $\varphi : [0,1] \times [0,1] \to [0,\infty)$ such that $\varphi(p,\, p) = 0$ and $\varphi(\cdot,\, p)$ is convex for every $p \in [0,1]$. When $\varphi(\cdot,\, p)$ is strictly convex for every $p \in [0,1]$, then $\hat{p}_{\mathbf{r},n}$ has strictly lower risk than $\hat{q}_{\mathbf{r},n}$ with respect to φ. In particular,

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) \;<\; \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{q}_{\mathbf{r},n})$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Proof. Let $\prec$ be any total order on $(S_d)^n$. Define

$$S(\boldsymbol{\Pi}_n) \;:=\; \min_{\prec}\big\{\boldsymbol{\Pi}_n,\, A(\boldsymbol{\Pi}_n),\, B(\boldsymbol{\Pi}_n),\, B\circ A(\boldsymbol{\Pi}_n)\big\}$$

where $\min_{\prec}$ denotes the minimum with respect to $\prec$. According to Lemma 3.3, when $\boldsymbol{\pi} \in (S_d)^n$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ are such that $\mathbb{P}_{\boldsymbol{\vartheta}}(S(\boldsymbol{\Pi}_n) = \boldsymbol{\pi}) \neq 0$, then the conditional distribution of $\boldsymbol{\Pi}_n$ given that $S(\boldsymbol{\Pi}_n) = \boldsymbol{\pi}$ is the equidistribution on $\{\boldsymbol{\pi},\, A(\boldsymbol{\pi}),\, B(\boldsymbol{\pi}),\, B\circ A(\boldsymbol{\pi})\}$, that is,

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\boldsymbol{\Pi}_n = \boldsymbol{\pi}' \,|\, S(\boldsymbol{\Pi}_n) = \boldsymbol{\pi}) \;=\; \frac{1}{4}$$

for $\boldsymbol{\pi}' \in \{\boldsymbol{\pi},\, A(\boldsymbol{\pi}),\, B(\boldsymbol{\pi}),\, B\circ A(\boldsymbol{\pi})\}$. Since the conditional distribution does not depend on $\boldsymbol{\vartheta}$, it follows that $S(\boldsymbol{\Pi}_n)$ is a sufficient statistic for $\boldsymbol{\Pi}_n$. Now, note that

$$\hat{p}_{\mathbf{r},n} \;=\; \frac{1}{4}\Big(\hat{q}_{\mathbf{r},n}(\boldsymbol{\Pi}_n) + \hat{q}_{\mathbf{r},n}(A(\boldsymbol{\Pi}_n)) + \hat{q}_{\mathbf{r},n}(B(\boldsymbol{\Pi}_n)) + \hat{q}_{\mathbf{r},n}(B\circ A(\boldsymbol{\Pi}_n))\Big)$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, which shows that $\hat{p}_{\mathbf{r},n}$ is a conditional expectation of $\hat{q}_{\mathbf{r},n}$ given $S(\boldsymbol{\Pi}_n)$. Since $\hat{q}_{\mathbf{r},n}$ is unbiased, Theorem 3.2.1 in Pfanzagl [79] shows that $\hat{p}_{\mathbf{r},n}$ has lower risk than $\hat{q}_{\mathbf{r},n}$. The result on strictness is also a consequence of Theorem 3.2.1 in Pfanzagl [79] and the fact that, according to Lemma 3.4, $\mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n} \neq \hat{q}_{\mathbf{r},n}) > 0$. Now, the statement on the variance follows because the function $(\cdot - p)^2$ is strictly convex for every $p \in [0,1]$. □

Remark 3.6. As the proof of Theorem 3.5 shows, the estimator obtained by averaging the relative frequencies of $\mathbf{r}$ and $\alpha(\mathbf{r})$ also has lower risk than $\hat{q}_{\mathbf{r},n}$. In particular,

$$T(\boldsymbol{\Pi}_n) \;:=\; \min_{\prec}\big\{\boldsymbol{\Pi}_n,\, A(\boldsymbol{\Pi}_n)\big\}$$

is a sufficient statistic for $\boldsymbol{\Pi}_n$, and

$$\frac{1}{2}\Big(\hat{q}_{\mathbf{r},n}(\boldsymbol{\Pi}_n) + \hat{q}_{\mathbf{r},n}(A(\boldsymbol{\Pi}_n))\Big) \;=\; \frac{1}{2n}\sum_{t=0}^{n-1}\Big(\mathbf{1}_{\{\Pi(t)=\mathbf{r}\}} + \mathbf{1}_{\{\Pi(t)=\alpha(\mathbf{r})\}}\Big)$$

is a conditional expectation of $\hat{q}_{\mathbf{r},n}$ given $T(\boldsymbol{\Pi}_n)$. □

Note that the proof of Theorem 3.5 only uses the assumptions of Lemma 3.3 and Lemma 3.4. In particular, for showing that $\hat{p}_{\mathbf{r},n}$ has lower risk than $\hat{q}_{\mathbf{r},n}$, it is sufficient that $\mathbf{Y}$ is symmetric in space and time and the values of $\mathbf{X}$ are pairwise different $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

3.4 Asymptotic properties

In this section, we study asymptotic properties of $\hat{p}_{\mathbf{r},n}$. As mentioned above, $\hat{p}_{\mathbf{r},n}$ is an unbiased estimator of $p_{\mathbf{r}}(\cdot)$, that is, $\mathbb{E}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) = p_{\mathbf{r}}(\boldsymbol{\vartheta})$ for all $n \in \mathbb{N}$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Next, we investigate conditions under which $\hat{p}_{\mathbf{r},n}$ is consistent and asymptotically normally distributed.

3.4.1 Consistency

Estimators are *weakly consistent* if they converge to the quantity of interest in probability, and *strongly consistent* if the convergence is almost surely. We will show that, under the model assumptions (M1)-(M3), a natural sufficient condition for weak consistency of $\hat{p}_{\mathbf{r},n}$ is also sufficient for strong consistency.

Weak consistency. Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. By the definition of convergence in probability, weak consistency of $\hat{p}_{\mathbf{r},n}$ is equivalent to

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\vartheta}}(\, |\hat{p}_{\mathbf{r},n} - p_{\mathbf{r}}(\boldsymbol{\vartheta})| \geq \epsilon \,) \quad = \quad 0$$

for every $\epsilon > 0$. Since $\mathbb{E}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) = p_{\mathbf{r}}(\boldsymbol{\vartheta})$, Chebyshev's inequality yields

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\, |\hat{p}_{\mathbf{r},n} - p_{\mathbf{r}}(\boldsymbol{\vartheta})| \geq \epsilon \,) \quad \leq \quad \frac{\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n})}{\epsilon^2} \, .$$

Therefore, a sufficient condition for weak consistency of $\hat{p}_{\mathbf{r},n}$ is $\lim_{n\to\infty} \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) = 0$. For $k \in \mathbb{Z}$, define

$$\gamma_{\boldsymbol{\vartheta}}(k) \quad := \quad \mathrm{Cov}_{\boldsymbol{\vartheta}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}}, \, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}}) \, .$$

Because $(\Pi(t))_{t\in\mathbb{Z}}$ is stationary, we have $\mathrm{Cov}_{\boldsymbol{\vartheta}}(\mathbf{1}_{\{\Pi(t)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(t+k)\in\bar{\mathbf{r}}\}}) = \gamma_{\boldsymbol{\vartheta}}(k)$ for $t, k \in \mathbb{Z}$, and thus

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) &= \frac{1}{n^2\,(\sharp\,\bar{\mathbf{r}})^2} \sum_{s,t=0}^{n-1} \mathrm{Cov}_{\boldsymbol{\vartheta}}(\mathbf{1}_{\{\Pi(s)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(t)\in\bar{\mathbf{r}}\}}) \\
&= \frac{1}{n^2\,(\sharp\,\bar{\mathbf{r}})^2} \Big(n\,\gamma_{\boldsymbol{\vartheta}}(0) + 2\sum_{k=1}^{n-1} (n-k)\,\gamma_{\boldsymbol{\vartheta}}(k) \Big) \\
&\le \frac{2}{n} \sum_{k=0}^{n-1} \gamma_{\boldsymbol{\vartheta}}(k)\,.
\end{aligned}$$

Therefore, a sufficient condition for $\lim_{n\to\infty} \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) = 0$ is $\lim_{k\to\infty} \gamma_{\boldsymbol{\vartheta}}(k) = 0$. Note that, for $k \in \mathbb{Z}$,

$$\begin{aligned}
\gamma_{\boldsymbol{\vartheta}}(k) &= \mathbb{P}_{\boldsymbol{\vartheta}}(\,\Pi(0)\in\bar{\mathbf{r}},\, \Pi(k)\in\bar{\mathbf{r}}\,) - \mathbb{P}_{\boldsymbol{\vartheta}}(\,\Pi(0)\in\bar{\mathbf{r}}\,)\,\mathbb{P}_{\boldsymbol{\vartheta}}(\,\Pi(k)\in\bar{\mathbf{r}}\,) \\
&= \mathbb{P}_{\boldsymbol{\vartheta}}(\,(Y_1, Y_2, \ldots, Y_d)\in B,\, (Y_{k+1}, Y_{k+2}, \ldots, Y_{k+d})\in B) \\
&\qquad - (\mathbb{P}_{\boldsymbol{\vartheta}}(\,(Y_1, Y_2, \ldots, Y_d)\in B))^2\,,
\end{aligned}$$

where $B = \{(y_1, y_2, \ldots, y_d) \in \mathbb{R}^d \,|\, \tilde{\pi}(y_1, y_2, \ldots, y_d) \in \bar{\mathbf{r}}\}$. Furthermore, a necessary condition for $(y_1, y_2, \ldots, y_d), (y_1', y_2', \ldots, y_d') \in \mathbb{R}^d$ to satisfy $\tilde{\pi}(y_1, y_2, \ldots, y_d) \in \bar{\mathbf{r}}$ and $\tilde{\pi}(y_1', y_2', \ldots, y_d') \notin \bar{\mathbf{r}}$ is the existence of a binary vector $(b_1, b_2, \ldots, b_d) \in \{0,1\}^d \setminus \{\mathbf{0}\}$ such that

$$\sum_{k=1}^{d} b_k y_k \ge 0 \;\text{ and }\; \sum_{k=1}^{d} b_k y_k' < 0\,, \quad \text{or} \quad \sum_{k=1}^{d} b_k y_k < 0 \;\text{ and }\; \sum_{k=1}^{d} b_k y_k' \ge 0\,.$$

This argument shows that the boundary of B in $\mathbb{R}^d$ is contained in the union of all sets $B_{\boldsymbol{b}}$ with $\boldsymbol{b} = (b_1, b_2, \ldots, b_d) \in \{0,1\}^d \setminus \{\mathbf{0}\}$, given by

$$B_{\boldsymbol{b}} := \Big\{ (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d \;\Big|\; \sum_{k=1}^{d} b_k y_k = 0 \Big\}\,.$$

Since each of these finitely many sets is a hyperplane in $\mathbb{R}^d$, the boundary of B in $\mathbb{R}^d$ has Lebesgue measure 0. Hence, a sufficient condition for $\lim_{k\to\infty} \gamma_{\boldsymbol{\vartheta}}(k) = 0$ is that $(Y_1, Y_2, \ldots, Y_d, Y_{k+1}, Y_{k+2}, \ldots, Y_{k+d})$ converges in distribution to a random vector $(Z_1, Z_2, \ldots, Z_d, Z_{d+1}, Z_{d+2}, \ldots, Z_{2d})$ with $(Z_1, Z_2, \ldots, Z_d)$ and $(Z_{d+1}, Z_{d+2}, \ldots, Z_{2d})$ being independent and both having the same distribution as $(Y_1, Y_2, \ldots, Y_d)$. According to Theorem 2.4 (*iii*), (*iv*) and the model assumption that $\mathbf{Y}$ is stationary and zero-mean Gaussian with unit variance, we obtain the following result.

Theorem 3.7. *If $\rho_{\boldsymbol{\vartheta}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, then $\hat{p}_{\mathbf{r},n}$ is a weakly consistent estimator of $p_{\mathbf{r}}(\cdot)$.*

As we show in the next paragraph, the assumptions of Theorem 3.7 are also sufficient for strong consistency of $\hat{p}_{\mathbf{r},n}$.

Strong consistency. For establishing strong consistency, we use well-known results from ergodic theory. For more background, we refer to Cornfeld et al. [29]. Let τ denote the shift operator, given by

$$\tau(\mathbf{z}) \;=\; (z_{t+1})_{t\in\mathbb{Z}}$$

for $\mathbf{z} = (z_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$. For $j \in \mathbb{N}$, we define $\tau^j(\mathbf{z}) := \tau^{j-1}(\tau(\mathbf{z}))$, where $\tau^0(\mathbf{z}) := \mathbf{z}$ is the identity on $\mathbb{R}^{\mathbb{Z}}$. A real-valued stationary stochastic process $\mathbf{Z} = (Z_t)_{t\in\mathbb{Z}}$ on a probability space $(\Omega', \mathcal{A}', \mathbb{P})$ is called *ergodic* iff $\mathbb{P}(\mathbf{Z} \in B) = 0$ or $\mathbb{P}(\mathbf{Z} \in B) = 1$ for every set $B \in \mathbb{B}(\mathbb{R}^{\mathbb{Z}})$ satisfying $\mathbb{P}(\tau^{-1}(B)\,\Delta\,B) = 0$. The Birkhoff-Khinchin Ergodic Theorem states that, if $\mathbf{Z}$ is ergodic,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\tau^j(\mathbf{Z})) \;=\; \mathbb{E}(f(\mathbf{Z}))$$

$\mathbb{P}$-almost surely for every measurable mapping $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ with $\mathbb{E}(\,|f(\mathbf{Z})|\,) < \infty$ (see Cornfeld et al. [29], Chapter 1, §2, Theorem 1). This property of ergodic processes is often referred to by saying that "time averages are equal to space averages". A stronger property than ergodicity is the *mixing property*, which means that, for any two measurable mappings $f, g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ satisfying $\mathbb{E}((f(\mathbf{Z}))^2) < \infty$ and $\mathbb{E}((g(\mathbf{Z}))^2) < \infty$, respectively,

$$\lim_{j\to\infty} \mathbb{E}\big(f(\tau^j(\mathbf{Z}))\, g(\mathbf{Z})\big) \;=\; \mathbb{E}(f(\mathbf{Z}))\,\mathbb{E}(g(\mathbf{Z}))\,.$$

According to Theorem 2 in Chapter 14, §2 in Cornfeld et al. [29], a necessary and sufficient condition for a stationary Gaussian process to have the mixing property is that the autocorrelations tend to zero as the lag tends to infinity.

The following theorem gives sufficient conditions under which estimators of continuous functions of ordinal pattern probabilities are strongly consistent and asymptotically unbiased.

Theorem 3.8.

(*i*) *If* $\rho_{\boldsymbol{\vartheta}}(k) \to 0$ *as* $k \to \infty$ *for every* $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ *and* $h : [0,1] \to \mathbb{R}$ *is continuous on an open set containing* $p_{\mathbf{r}}(\boldsymbol{\Theta})$, *then* $h(\hat{p}_{\mathbf{r},n})$ *is a strongly consistent estimator of* $h(p_{\mathbf{r}}(\cdot))$, *that is,*

$$\lim_{n\to\infty} h(\hat{p}_{\mathbf{r},n}) \;=\; h(p_{\mathbf{r}}(\boldsymbol{\vartheta}))$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$*-almost surely for every* $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

(*ii*) *If, additionally to the conditions in* (*i*), h *is bounded on* $[0,1]$, *then* $h(\hat{p}_{\mathbf{r},n})$ *is an asymptotically unbiased estimator of* $h(p_{\mathbf{r}}(\cdot))$, *that is,*

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\vartheta}}(h(\hat{p}_{\mathbf{r},n})) \;=\; h(p_{\mathbf{r}}(\boldsymbol{\vartheta}))$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Proof. (i) Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. For $\mathbf{y} = (y_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$, define

$$f(\mathbf{y}) \;:=\; \begin{cases} 1 & \text{if } \tilde{\pi}(y_1, y_2, \ldots, y_d) \in \bar{\mathbf{r}} \\ 0 & \text{otherwise} \end{cases} .$$

According to the definition of τ,

$$f(\tau^j(\mathbf{y})) \;=\; \begin{cases} 1 & \text{if } \tilde{\pi}(y_{j+1}, y_{j+2}, \ldots, y_{j+d}) \in \bar{\mathbf{r}} \\ 0 & \text{otherwise} \end{cases}$$

for $j = 0, 1, 2, \ldots$, and hence $f(\tau^j(\mathbf{Y})) = \mathbf{1}_{\{\Pi(j)\in\bar{\mathbf{r}}\}}$. Under the assumptions, $\mathbf{Y}$ has the mixing property, and thus $\mathbf{Y}$ is ergodic. According to the Birkhoff-Khinchin Ergodic Theorem, we obtain

$$\begin{aligned} \lim_{n\to\infty} \hat{p}_{\mathbf{r},n} \;&=\; \frac{1}{\sharp\,\bar{\mathbf{r}}} \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\tau^j(\mathbf{Y})) \\ &=\; \frac{1}{\sharp\,\bar{\mathbf{r}}}\, \mathbb{E}_{\boldsymbol{\vartheta}}(f(\mathbf{Y})) \;=\; p_{\mathbf{r}}(\boldsymbol{\vartheta}) \end{aligned}$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Since h is continuous on an open set containing $p_{\mathbf{r}}(\boldsymbol{\Theta})$, there exists a $\delta > 0$ such that h is continuous on $(p_{\mathbf{r}}(\boldsymbol{\vartheta}) - \delta,\, p_{\mathbf{r}}(\boldsymbol{\vartheta}) + \delta)$. Thus,

$$\lim_{n\to\infty} h(\hat{p}_{\mathbf{r},n}) \;=\; h(p_{\mathbf{r}}(\boldsymbol{\vartheta}))$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely, which shows that $h(\hat{p}_{\mathbf{r},n})$ is strongly consistent.

(ii) Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Since h is bounded on $[0, 1]$, there exists a $c < \infty$ with $\mathbb{E}_{\boldsymbol{\vartheta}}(\,|h(\hat{p}_{\mathbf{r},n})|\,) < c$ for every $n \in \mathbb{N}$. According to the Dominated Convergence Theorem and the strong consistency of $h(\hat{p}_{\mathbf{r},n})$, it follows that

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\vartheta}}(h(\hat{p}_{\mathbf{r},n})) \;=\; \mathbb{E}_{\boldsymbol{\vartheta}}(\,\lim_{n\to\infty} h(\hat{p}_{\mathbf{r},n})\,) \;=\; h(p_{\mathbf{r}}(\boldsymbol{\vartheta}))\,.$$

The proof is complete. □

Remark 3.9. According to Corollary 3.2, we have $p_{\mathbf{r}}(\boldsymbol{\Theta}) \subset (0, 1)$ for every $\mathbf{r} \in S_d$ with $d \in \mathbb{N}$. Therefore, an open set as in (i) does always exist. □

3.4.2 Asymptotic normality

Next, we derive sufficient conditions for asymptotic normality of $\hat{p}_{\mathbf{r},n}$. The result is obtained by the Limit Theorem of Arcones [7], which we present in the following paragraph. Note that the result can also be established by a limit theorem of Ho and Sun [48].

The Limit Theorem of Arcones. Recall the concept of Hermite ranks: For $\mathbf{l} = (l_1, l_2, \ldots, l_d) \in \mathbb{N}_0^d$, define $|\mathbf{l}| := l_1 + l_2 + \ldots + l_d$. Furthermore, for $\mathbf{z} = (z_1, z_2, \ldots, z_d) \in \mathbb{R}^d$, let $\mathbf{z}^{\mathbf{l}} := \prod_{i=1}^d z_i^{l_i}$. Suppose $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$ is a Gaussian random vector on a probability space $(\Omega', \mathcal{A}', \mathbb{P})$ and $g : \mathbb{R}^d \to \mathbb{R}$ is a mapping satisfying $\mathbb{E}((g(\mathbf{Z}))^2) < \infty$. The *Hermite rank* of g with respect to $\mathbf{Z}$ is defined by

$$\text{rank}(g) \;:=\; \inf\Big\{\kappa \in \mathbb{N} \;\Big|\; \text{There exists an } \mathbf{l} \in \mathbb{N}_0^d \text{ with } |\mathbf{l}| = \kappa \text{ and } \mathbb{E}(\,[\,g(\mathbf{Z}) - \mathbb{E}(g(\mathbf{Z}))\,]\,\mathbf{Z}^{\mathbf{l}}\,) \neq 0\Big\}\,, \tag{3.14}$$

where the infimum of the empty set is infinity. As we show in the following paragraph, the definition given here is equivalent to the usual definition of Hermite ranks in terms of Hermite polynomials.

Let $(\mathbf{Z}(t))_{t\in\mathbb{Z}}$ with $\mathbf{Z}(t) = (Z_1(t), Z_2(t), \ldots, Z_d(t))$ be a stationary and non-degenerate sequence of Gaussian random vectors on $(\Omega', \mathcal{A}', \mathbb{P})$. Define

$$r^{(i,j)}(k) \;:=\; \text{Corr}(Z_i(0),\, Z_j(k))$$

for $i, j \in \{1, 2, \ldots, d\}$ and $k \in \mathbb{Z}$. Theorem 4 in Arcones [7] states that, if g has Hermite rank $\kappa < \infty$ with respect to $\mathbf{Z}(0)$ and

$$\sum_{k=1}^{\infty} \left|r^{(i,j)}(k)\right|^{\kappa} \;<\; \infty \tag{3.15}$$

for all $i, j \in \{1, 2, \ldots, d\}$, then

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \big(\, g(\mathbf{Z}(t)) - \mathbb{E}[g(\mathbf{Z}(0))]\,\big) \;\xrightarrow{\mathbb{P}}\; \mathrm{N}(0, \sigma^2)\,, \tag{3.16}$$

where

$$\sigma^2 \;:=\; \text{Var}\big(\, g(\mathbf{Z}(0))\,\big) \;+\; 2 \sum_{k=1}^{\infty} \text{Cov}\big(\, g(\mathbf{Z}(0)),\, g(\mathbf{Z}(k))\,\big)\,. \tag{3.17}$$

Arcones also gives a generalization in the multidimensional case. Let $m \in \mathbb{N}$. Assume that each of $g_1, g_2, \ldots, g_m$ has Hermite rank κ or greater with respect to $\mathbf{Z}(0)$, where κ satisfies condition (3.15). It is easy to see that $a_1 g_1 + a_2 g_2 + \ldots + a_m g_m$ has Hermite rank κ or greater with respect to $\mathbf{Z}(0)$ for all $a_1, a_2, \ldots, a_m \in \mathbb{R}$ and thus, by the Limit Theorem,

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \sum_{j=1}^{m} a_j \,\big(\, g_j(\mathbf{Z}(t)) - \mathbb{E}[g_j(\mathbf{Z}(0))]\,\big) \;\xrightarrow{\mathbb{P}}\; \mathrm{N}\Big(0,\, \sum_{i,j=1}^{m} a_i a_j s(i,j)\Big)\,,$$

where

$$\begin{aligned} s(i,j) &:= \operatorname{Cov}\big(\, g_i(\mathbf{Z}(0)),\, g_j(\mathbf{Z}(0))\,\big) \\ &\quad + \sum_{k=1}^{\infty} \Big(\operatorname{Cov}\big(\, g_i(\mathbf{Z}(0)),\, g_j(\mathbf{Z}(k))\,\big) + \operatorname{Cov}\big(\, g_i(\mathbf{Z}(k)),\, g_j(\mathbf{Z}(0))\,\big) \Big) \end{aligned}$$

for $i, j = 1, 2, \ldots, m$. According to the Cramér-Wold Theorem (see Theorem 29.4 in Billingsley [17]), we obtain

$$\frac{1}{\sqrt{n}} \Big(\sum_{t=0}^{n-1} g_1(\mathbf{Z}(t)),\, \sum_{t=0}^{n-1} g_2(\mathbf{Z}(t)),\, \ldots,\, \sum_{t=0}^{n-1} g_m(\mathbf{Z}(t)) \Big) \xrightarrow{\mathbb{P}} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.18}$$

where $\boldsymbol{\mu} = \big(\mathbb{E}[\, g_1(\mathbf{Z}(0))],\ \mathbb{E}[\, g_2(\mathbf{Z}(0))],\ \ldots,\ \mathbb{E}[\, g_m(\mathbf{Z}(0))] \big)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^{m}$ is given by $\sigma_{ij} := s(i,j)$ for $i, j = 1, 2, \ldots, m$.

Background: Hermite expansions. Next we provide some background on Hermite ranks and the derivation of Arcones' Limit Theorem. The reader can skip this part and continue with the following paragraph, where we apply the Limit Theorem to estimators of ordinal pattern probabilities. Recall the definition of Hermite polynomials: For $n \in \mathbb{N}_0$, let $H_n : \mathbb{R} \to \mathbb{R}$ be defined by

$$H_n(z) := (-1)^n e^{z^2/2} \frac{\mathrm{d}^n}{\mathrm{d}z^n} e^{-z^2/2}$$

for $z \in \mathbb{R}$. Indeed, each H_n is a polynomial of degree n, called the nth Hermite polynomial. For instance, $H_0(z) = 1$, $H_1(z) = z$, and $H_2(z) = z^2 - 1$ for $z \in \mathbb{R}$.

Suppose that $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$ is a standard normal random vector on $(\Omega', \mathcal{A}', \mathbb{P})$. Let $\mathbb{L}^2(\mathbf{Z})$ be the set containing any measurable function $g : \mathbb{R}^d \to \mathbb{R}$ satisfying $\mathbb{E}(g(\mathbf{Z})) = 0$ and $\mathbb{E}((g(\mathbf{Z}))^2) < \infty$. By $L^2(\mathbf{Z})$ we denote the set obtained by identifying functions $f, g \in \mathbb{L}^2(\mathbf{Z})$ with $f(\mathbf{Z}) = g(\mathbf{Z})$ $\mathbb{P}$-almost surely. Define $\langle g, h \rangle := \operatorname{Cov}(g(\mathbf{Z}), h(\mathbf{Z}))$ for $g, h \in L^2(\mathbf{Z})$. Note that $L^2(\mathbf{Z})$ equipped with the inner product $\langle \cdot\,, \cdot \rangle$ is a Hilbert space.

For $\mathbf{l} = (l_1, l_2, \ldots, l_d) \in \mathbb{N}_0^d$, let $\mathbf{l}! := l_1!\, l_2! \ldots l_d!$. Furthermore, define

$$b_{\mathbf{l}}(\mathbf{z}) := \prod_{i=1}^{d} H_{l_i}(z_i)$$

for $\mathbf{z} = (z_1, z_2, \ldots, z_d) \in \mathbb{R}^d$. The family of mappings $b_{\mathbf{l}}$ with $\mathbf{l} \in \mathbb{N}_0^d \setminus \{(0, 0, \ldots, 0)\}$ forms an orthogonal basis of $L^2(\mathbf{Z})$. The *Hermite expansion* of $g \in L^2(\mathbf{Z})$ is given by

$$g(\mathbf{z}) = \sum_{\substack{\mathbf{l} \in \mathbb{N}_0^d \\ \mathbf{l} \neq (0,0,\ldots,0)}} \frac{1}{\mathbf{l}!} \langle g, b_{\mathbf{l}} \rangle\, b_{\mathbf{l}}(\mathbf{z})$$

for $\mathbf{z} \in \mathbb{R}^d$ (see Arcones [7], Doukhan [34]).

Now, suppose $(\mathbf{Z}(t))_{t\in\mathbb{Z}}$ is a stationary sequence of standard normal random vectors on $(\Omega', \mathcal{A}', \mathbb{P})$. For the proof of his Limit Theorem, Arcones [7] first shows that when g has Hermite rank $\kappa < \infty$ with respect to $\mathbf{Z}(0)$ and condition (3.15) is satisfied,

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} g(\mathbf{Z}(t)) \quad \xrightarrow{\mathbb{P}} \quad \mathrm{N}(0, \sigma^2) \tag{3.19}$$

with σ^2 as given in (3.17). The general result for possibly non-centered functions of arbitrary non-degenerate Gaussian random vectors is obtained as follows: If $\mathbb{E}(g(\mathbf{Z})) \neq 0$, then consider the mapping $g(\cdot) - \mathbb{E}(g(\mathbf{Z}))$ instead of $g(\cdot)$. Furthermore, for every symmetric and strictly positive definite matrix $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ and every $\boldsymbol{\mu} \in \mathbb{R}^d$, there exists a matrix $\mathbf{L} \in \mathbb{R}^{d\times d}$ such that $g(\cdot)$ applied to a Gaussian random vector with means $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ has the same distribution as $g(\boldsymbol{\mu} + \mathbf{L}\,\cdot)$ applied to a standard normal random vector (see Horn and Johnson [49], p. 406).

The basic step in the proof of Arcones is to insert the Hermite expansions of $g(\mathbf{Z}(t))$ for $t = 0, 1, \ldots, n-1$ into the left hand side of (3.19), namely,

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} g(\mathbf{Z}(t)) \;=\; \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \sum_{\substack{\mathbf{l}\in\mathbb{N}_0^d \\ \mathbf{l}\neq(0,0,\ldots,0)}} \frac{1}{\mathbf{l}!} \langle g, b_{\mathbf{l}} \rangle\, b_{\mathbf{l}}(\mathbf{Z}(t)) \,, \tag{3.20}$$

and showing that the moments of the expression on the right hand side of (3.20) converge to the corresponding moments of a normal distribution. Note that, for $m \in \mathbb{N}$, the mth moment of the expression on the right hand side of (3.20) can be written as the weighted sum of terms

$$\mathbb{E}\big(H_{l_1}(Z_{i_1}(t_1))\, H_{l_2}(Z_{i_2}(t_2)) \,\ldots\, H_{l_r}(Z_{i_r}(t_r)) \big)$$

with $r = md$ and $l_j \in \mathbb{N}_0$, $i_j \in \{1, 2, \ldots, d\}$, $t_j \in \{0, 1, \ldots, n-1\}$ for $j = 1, 2, \ldots, r$. By using the *Diagram Formula* (see Surgailis [94]), each of these expressions can be written as a sum of products of correlations among $Z_{i_1}(t_1), Z_{i_2}(t_2), \ldots, Z_{i_r}(t_r)$. The products are determined by the edges of *diagrams*, which are certain undirected graphs describing the stochastic dependencies among $H_{l_1}(Z_{i_1}(t_1)), H_{l_2}(Z_{i_2}(t_2)), \ldots, H_{l_r}(Z_{i_r}(t_r))$.

It turns out that the limiting distribution of $\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} g(\mathbf{Z}(t))$ is essentially determined by the lowest order basis terms $b_{\mathbf{l}}$ for which $\langle g,\, b_{\mathbf{l}} \rangle \neq 0$. This is the reason for introducing the *Hermite rank* of g,

$$\mathrm{rank}(g) \;:=\; \inf\big\{ \kappa \in \mathbb{N} \;\big|\; \text{There exists an } \mathbf{l} \in \mathbb{N}_0^d \text{ with } |\mathbf{l}| = \kappa \text{ and } \langle g,\, b_{\mathbf{l}} \rangle \neq 0 \big\} \,.$$

Let us verify that this definition is equivalent to the one given in (3.15). Suppose there exists a $\kappa \in \mathbb{N}$ such that $\langle g,\, b_{\mathbf{l}} \rangle = \mathbb{E}(g(\mathbf{Z})\, b_{\mathbf{l}}(\mathbf{Z})) \neq 0$ for some $\mathbf{l} \in \mathbb{N}_0^d$ with $|\mathbf{l}| = \kappa$ and

$\langle g, b_{\mathbf{l}} \rangle = 0$ for every $\mathbf{l} \in \mathbb{N}_0^d$ with $|\mathbf{l}| < \kappa$. It is easy to see that $\mathbb{E}(g(\mathbf{Z})\,\mathbf{Z}^{\mathbf{l}}) \neq 0$ for some $\mathbf{l} \in \mathbb{N}_0^d$ with $|\mathbf{l}| = \kappa$ and $\mathbb{E}(g(\mathbf{Z})\,\mathbf{Z}^{\mathbf{l}}) = 0$ for all $\mathbf{l} \in \mathbb{N}_0^d$ with $|\mathbf{l}| < \kappa$. (In fact, this is a consequence of both $b_{\mathbf{l}}$ and $\mathbf{z} \mapsto \mathbf{z}^{\mathbf{l}}$ with $|\mathbf{l}| \leq \kappa$ being a basis of the polynomials $\mathbb{R}^d \to \mathbb{R}$ of degree κ.) Therefore, an equivalent expression for $\operatorname{rank}(g)$ is given by

$$\operatorname{rank}(g) \;=\; \inf\left\{\kappa \in \mathbb{N} \;\middle|\; \text{There exists an } \mathbf{l} \in \mathbb{N}_0^d \text{ with } |\mathbf{l}| = \kappa \text{ and } \mathbb{E}(g(\mathbf{Z})\,\mathbf{Z}^{\mathbf{l}}) \neq 0\right\}.$$

Now, if $g(\mathbf{Z})$ is non-centered, replacing it by the centered random variable $g(\mathbf{Z}) - \mathbb{E}(g(\mathbf{Z}))$ yields the statement in (3.15).

Application to ordinal patterns. Let us apply Arcones' Limit Theorem to derive conditions for asymptotic normality of the estimators of ordinal pattern probabilities. For $t \in \mathbb{Z}$, define

$$\mathbf{Y}(t) \;=\; (Y_1(t), Y_2(t), \ldots, Y_d(t)) \;:=\; (Y_{t+1}, Y_{t+1}, \ldots, Y_{t+d}).$$

Furthermore, let

$$r_{\boldsymbol{\vartheta}}^{(i,j)}(k) \;:=\; \operatorname{Corr}_{\boldsymbol{\vartheta}}(Y_i(0), Y_j(k)) \tag{3.21}$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, $k \in \mathbb{Z}$ and $i, j \in \{1, 2, \ldots, d\}$.

Theorem 3.10. *If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$, then*

$$\sqrt{n}\,(\hat{p}_{\mathbf{r},n} - p_{\mathbf{r}}(\boldsymbol{\vartheta})) \;\xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}}\; \mathrm{N}(0, \sigma_{\boldsymbol{\vartheta}}^2),$$

where

$$\sigma_{\boldsymbol{\vartheta}}^2 \;:=\; \gamma_{\boldsymbol{\vartheta}}(0) \,+\, 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{\vartheta}}(k)$$

and $\gamma_{\boldsymbol{\vartheta}}(k) := \frac{1}{(\sharp\,\bar{\mathbf{r}})^2} \operatorname{Cov}_{\boldsymbol{\vartheta}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}}, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}})$ for $k \in \mathbb{Z}$.

Proof. Let $g : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$g(\mathbf{z}) \;:=\; \begin{cases} \frac{1}{\sharp\,\bar{\mathbf{r}}} & \text{if } \tilde{\pi}(\mathbf{z}) \in \bar{\mathbf{r}} \\ 0 & \text{otherwise} \end{cases}$$

for $\mathbf{z} \in \mathbb{R}^d$. Note that $g(\mathbf{Y}(t)) = \frac{1}{\sharp\,\bar{\mathbf{r}}}\,\mathbf{1}_{\{\Pi(t)\in\bar{\mathbf{r}}\}}$ for every $t \in \mathbb{Z}$. Therefore, according to the definition of $\hat{p}_{\mathbf{r},n}$ (see (3.13)), we obtain

$$\sqrt{n}\,(\hat{p}_{\mathbf{r},n} - p_{\mathbf{r}}(\boldsymbol{\vartheta})) \;=\; \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \big(\, g(\mathbf{Y}(t)) - \mathbb{E}[g(\mathbf{Y}(0))] \,\big)$$

for every $n \in \mathbb{N}$. Now, let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$ be a standard normal random vector on $(\Omega', \mathcal{A}', \mathbb{P})$ and note that $\mathbb{E}((g(\mathbf{Z}))^2) < \infty$. We show that g has Hermite rank $\kappa \geq 2$ with respect to $\mathbf{Z}$. Let $i \in \{1, 2, \ldots, d\}$. According to (3.15), it suffices to show that

$$\mathbb{E}([g(\mathbf{Z}) - \mathbb{E}(g(\mathbf{Z}))]\, Z_i) \;=\; 0\,.$$

Since Z_i is zero-mean Gaussian, we have $\mathbb{E}(g(\mathbf{Z}))\,\mathbb{E}(Z_i) = 0$ and thus

$$\mathbb{E}([g(\mathbf{Z}) - \mathbb{E}(g(\mathbf{Z}))]\, Z_i) \;=\; \mathbb{E}(g(\mathbf{Z})\, Z_i)\,.$$

Furthermore, because $\mathbf{Z}$ is non-degenerate, using the same argument as in the proof of Lemma 3.3 shows that $\mathbf{1}_{\{\tilde{\pi}(-\mathbf{Z})=\alpha(\mathbf{s})\}} = \mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\mathbf{s}\}}$ $\mathbb{P}$-almost surely for every $\mathbf{s} \in S_d$. Since $\mathbf{Z}$ is zero-mean Gaussian, $\mathbf{Z}$ and $-\mathbf{Z}$ are identically distributed (compare to the discussion of space symmetry in Section 3.3), and thus

$$\mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\alpha(\mathbf{s})\}}\, Z_i) \;=\; \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(-\mathbf{Z})=\alpha(\mathbf{s})\}}\,(-Z_i)) \;=\; -\,\mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\mathbf{s}\}}\, Z_i)\,.$$

In the case $\sharp\,\bar{\mathbf{r}} = 2$, where $g(\mathbf{Z}) = \frac{1}{2}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\mathbf{r}\}} + \mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\alpha(\mathbf{r})\}})$, we obtain

$$2\,\mathbb{E}(g(\mathbf{Z})\, Z_i) \;=\; \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\mathbf{r}\}}\, Z_i) \,+\, \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\alpha(\mathbf{r})\}}\, Z_i) \;=\; 0\,.$$

Analogously, in the case $\sharp\,\bar{\mathbf{r}} = 4$, we have

$$\begin{aligned} 4\,\mathbb{E}(g(\mathbf{Z})\, Z_i) \;&=\; \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\mathbf{r}\}}\, Z_i) + \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\alpha(\mathbf{r})\}}\, Z_i) \\ &+\; \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\beta(\mathbf{r})\}}\, Z_i) + \mathbb{E}(\mathbf{1}_{\{\tilde{\pi}(\mathbf{Z})=\alpha\circ\beta(\mathbf{r})\}}\, Z_i) \;=\; 0\,. \end{aligned}$$

Altogether, $\mathbb{E}([g(\mathbf{Z}) - \mathbb{E}(g(\mathbf{Z}))]\, Z_i) = 0$ which shows that g has Hermite rank $\kappa \geq 2$. Note that we have only used that $\mathbf{Z}$ is non-degenerate zero-mean Gaussian, so g has Hermite rank $\kappa \geq 2$ also with respect to $\mathbf{Y}(0)$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Now, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and suppose $\rho_{\boldsymbol{\vartheta}}(k) = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$. According to (3.21),

$$r_{\boldsymbol{\vartheta}}^{(i,j)}(k) \;=\; \rho_{\boldsymbol{\vartheta}}(k + i - j)$$

for $k \in \mathbb{Z}$ and $i, j \in \{1, 2, \ldots, d\}$. Since $(k + i - j)^{-\beta} \sim k^{-\beta}$ for all $i, j \in \{1, 2, \ldots, d\}$, we have $r_{\boldsymbol{\vartheta}}^{(i,j)}(k) = o(k^{-\beta})$ and thus

$$\sum_{k=1}^{\infty} |r_{\boldsymbol{\vartheta}}^{(i,j)}(k)|^{\kappa} \;<\; \infty\,.$$

By the Limit Theorem of Arcones (see (3.15)-(3.17)), we obtain that $\hat{p}_{\mathbf{r},n}$ is asymptotically normally distributed, where the expression for $\sigma^2_{\boldsymbol{\vartheta}}$ follows from

$$\mathrm{Cov}_{\boldsymbol{\vartheta}}\big(\, g(\mathbf{Y}(0)),\, g(\mathbf{Y}(k))\,\big) \;=\; \frac{1}{(\sharp\,\bar{\mathbf{r}})^2}\,\mathrm{Cov}_{\boldsymbol{\vartheta}}\big(\, \mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}}\,\big)$$

for $k \in \mathbb{Z}$. The proof is complete. □

Remark 3.11. According to the multidimensional version of Arcones' Limit Theorem (see (3.18)), we obtain that the joint estimate $(\hat{p}_{\mathbf{r}_1,n},\, \hat{p}_{\mathbf{r}_2,n},\, \ldots,\, \hat{p}_{\mathbf{r}_m,n})$ of the probabilities of $\mathbf{r}_1 \in S_{d_1}, \mathbf{r}_2 \in S_{d_2}, \ldots, \mathbf{r}_m \in S_{d_m}$ with $d_1, d_2, \ldots, d_m \in \mathbb{N}$ is asymptotically normally distributed if the assumptions of Theorem 3.10 are satisfied. □

Remark 3.12. The limiting distribution $\mathrm{N}(0, \sigma^2_{\boldsymbol{\vartheta}})$ in the conclusion of Theorem 3.10 can be degenerate. In particular, if $\mathbf{r} = (1,0)$ or $\mathbf{r} = (0,1)$, then $\bar{\mathbf{r}} = \{(1,0),(0,1)\}$ and thus

$$\hat{p}_{\mathbf{r},n} \;=\; \frac{1}{2n} \sum_{t=0}^{n-1} \left(\mathbf{1}_{\{X_{t+1} \geq X_t\}} + \mathbf{1}_{\{X_t < X_{t+1}\}} \right),$$

which shows that $\hat{p}_{\mathbf{r},n}$ is consistently equal to $\frac{1}{2}$ for every $n \in \mathbb{N}$. □

Remark 3.13. If there does not exist a $\beta > \frac{1}{2}$ such that $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$, then, in general, the conclusion of Theorem 3.10 does not hold. In particular, for ordinal patterns of order $d = 2$, the quantity $\sigma^2_{\boldsymbol{\vartheta}}$ is infinite in this case (see Chapter 5). It remains an open problem whether a different scaling than $\sqrt{n}$ yields convergence to a non-degenerate normal distribution. □

Remark 3.14. The condition on the rate of decrease of $k \mapsto |\rho_{\boldsymbol{\vartheta}}(k)|$ cannot be relaxed in general. In particular, the Hermite rank of the mapping g defined in the proof of Theorem 3.10 may be exactly equal to 2. To see this, let $\mathbf{Z} = (Z_1, Z_2)$ be standard normal and $\mathbf{r} = (0,1,2)$. Since $\bar{\mathbf{r}} = \{(0,1,2),(2,1,0)\}$, we obtain

$$g(\mathbf{Z}) \;=\; \frac{1}{2} \left(\mathbf{1}_{\{Z_1>0,\, Z_2>0\}} + \mathbf{1}_{\{Z_1<0,\, Z_2<0\}} \right) \;=\; \frac{1}{2}\, \mathbf{1}_{\{Z_1 Z_2 > 0\}}$$

$\mathbb{P}$-almost surely, so g has the same Hermite rank with respect to $\mathbf{Z}$ as the mapping h given by $h(z_1,z_2) := 1$ if $z_1 z_2 > 0$ and $h(z_1,z_2) := 0$, otherwise. Note that

$$\begin{aligned} \mathbb{E}(\,[\,h(\mathbf{Z}) - \mathbb{E}(h(\mathbf{Z}))\,]\,Z_1 Z_2\,) \;&=\; \mathbb{E}(\,[1 - \mathbb{E}(h(\mathbf{Z}))] Z_1 Z_2 \mid h(\mathbf{Z}) = 1\,)\,\mathbb{P}(h(\mathbf{Z}) = 1) \\ &+\; \mathbb{E}(\,[0 - \mathbb{E}(h(\mathbf{Z}))] Z_1 Z_2 \mid h(\mathbf{Z}) = 0\,)\,\mathbb{P}(h(\mathbf{Z}) = 0)\,. \end{aligned}$$

Moreover,

$$\begin{aligned} \mathbb{E}(\,[1 - \mathbb{E}(h(\mathbf{Z}))] Z_1 Z_2 \mid h(\mathbf{Z}) = 1\,) \;&=\; (1 - \mathbb{P}(h(\mathbf{Z}) = 1))\,\mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 > 0)\,, \\ \mathbb{E}(\,[0 - \mathbb{E}(h(\mathbf{Z}))] Z_1 Z_2 \mid h(\mathbf{Z}) = 0\,) \;&=\; -\,\mathbb{P}(h(\mathbf{Z}) = 1)\,\mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 < 0)\,, \end{aligned}$$

and with $\mathbb{P}(h(\mathbf{Z}) = 1) = \mathbb{P}(Z_1 Z_2 > 0)$ and $\mathbb{P}(h(\mathbf{Z}) = 0) = (1 - \mathbb{P}(Z_1 Z_2 > 0))$, we obtain

$$\begin{aligned} \mathbb{E}(\,[\,h(\mathbf{Z}) - \mathbb{E}(h(\mathbf{Z}))\,]\,Z_1 Z_2\,) \;&=\; [1 - \mathbb{P}(Z_1 Z_2 > 0)]\,\mathbb{P}(Z_1 Z_2 > 0) \\ &\quad \times \left(\mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 > 0) - \mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 < 0) \right). \end{aligned}$$

Since $\mathbb{P}(Z_1 Z_2 > 0) = \frac{1}{2}$ and $\mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 > 0) = -\mathbb{E}(Z_1 Z_2 \mid Z_1 Z_2 < 0) > 0$ we have $\mathbb{E}(\,[\,h(\mathbf{Z}) - \mathbb{E}(h(\mathbf{Z}))\,]\,Z_1 Z_2\,) > 0$ and thus $\mathrm{rank}(h) = 2$. □

Remark 3.15. The assumptions of Theorem 3.10 are sufficient also for the asymptotic normality of the estimator obtained by averaging the relative frequencies of $\mathbf{r}$ and $\alpha(\mathbf{r})$. In particular, as the proof of Theorem 3.10 shows, the mapping h given by $h(\mathbf{z}) := 1$ if $\tilde{\pi}(\mathbf{z}) \in \{\mathbf{r}, \alpha(\mathbf{r})\}$ and $h(\mathbf{z}) := 0$, otherwise, has Hermite rank greater than 1. □

Remark 3.16. According to the Limit Theorem of Arcones, a sufficient condition for the estimator $\hat{q}_{\mathbf{r},n}$ to be asymptotically normal is given by $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > 1$. In particular, with the mapping h defined by

$$h(\mathbf{z}) \;:=\; \begin{cases} \frac{1}{\sharp\,\tilde{\mathbf{r}}} & \text{if } \tilde{\pi}(\mathbf{z}) = \mathbf{r} \\ 0 & \text{otherwise} \end{cases}$$

for $\mathbf{z} \in \mathbb{R}^d$, we have

$$\sqrt{n}\,(\hat{q}_{\mathbf{r},n} - p_{\mathbf{r}}(\boldsymbol{\vartheta})) \;=\; \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \big(h(\mathbf{Y}(t)) - \mathbb{E}[h(\mathbf{Y}(0))] \big)\,.$$

Moreover, h has Hermite rank 1 with respect to any non-degenerate zero-mean Gaussian random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$. This can be seen as follows: Corollary 3.2 implies $\mathbb{P}(h(\mathbf{Z}) = 1) > 0$. Furthermore, we have either $\mathbb{P}(Z_1 > 0 \mid h(\mathbf{Z}) = 1) = 1$ or $\mathbb{P}(Z_1 < 0 \mid h(\mathbf{Z}) = 1) = 1$, and hence

$$\begin{aligned} \mathbb{E}([h(\mathbf{Z}) - \mathbb{E}(h(\mathbf{Z}))]\,Z_1) &= \mathbb{E}(h(\mathbf{Z})\,Z_1) \\ &= \mathbb{E}(Z_1 \mid h(\mathbf{Z}) = 1)\,\mathbb{P}(h(\mathbf{Z}) = 1) \;\neq\; 0\,. \end{aligned}$$

Since $\mathbf{Y}(0)$ is non-degenerate zero-mean Gaussian, we obtain that h has Hermite rank 1 with respect to $\mathbf{Y}(0)$. □

By applying the *Delta Method* (Lehmann [68], Theorem 2.5.2), we obtain the following statement on the limiting distribution of certain differentiable functions of $\hat{p}_{\mathbf{r},n}$.

Corollary 3.17. *If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ and $h : [0,1] \to \mathbb{R}$ has a non-vanishing first derivative at $p_{\mathbf{r}}(\boldsymbol{\vartheta})$, then*

$$\sqrt{n}\,\big(h(\hat{p}_{\mathbf{r},n}) - h(p_{\mathbf{r}}(\boldsymbol{\vartheta})) \big) \;\xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}}\; \mathrm{N}(\,0,\, \sigma^2_{\boldsymbol{\vartheta}}[h'(p_{\mathbf{r}}(\boldsymbol{\vartheta}))]^2\,)\,,$$

with $\sigma^2_{\boldsymbol{\vartheta}}$ as given in Theorem 3.10.

Similar to Remark 3.11, we also obtain a statement for the multdimensional case: Let $\mathbf{r}_1 \in S_{d_1}, \mathbf{r}_2 \in S_{d_2}, \ldots, \mathbf{r}_m \in S_{d_m}$ with $d_1, d_2, \ldots, d_m \in \mathbb{N}$, and suppose $h_i : [0,1] \to \mathbb{R}$ has a non-vanishing first derivative at $p_{\mathbf{r}_i}(\boldsymbol{\vartheta})$ for $i = 1, 2, \ldots, m$. If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$, then $(h_1(\hat{p}_{\mathbf{r}_1,n}),\, h_2(\hat{p}_{\mathbf{r}_2,n}),\, \ldots,\, h_m(\hat{p}_{\mathbf{r}_m,n}))$ is asymptotically normally distributed. In particular, since linear combinations of jointly normal random variables are normally distributed, the empirical permutation entropy, given by

$$\hat{P}_n \;:=\; -\sum_{\mathbf{r} \in S_d} \hat{p}_{\mathbf{r},n} \ln \hat{p}_{\mathbf{r},n}\,,$$

is asymptotically normally distributed if $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$.

3.5 Examples

Equidistant discretizations of FBM. Next, we apply the previous results to the estimation of ordinal pattern probabilities in equidistant discretizations of Fractional Brownian Motion (FBM). Suppose $\mathbf{B} = (B(t))_{t\in\mathbb{R}}$ is a family of measurable mappings from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$, where $(\Omega, \mathcal{A})$ is equipped with a family $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H}\in(0,1)}$ of probability measures such that $\mathbf{B}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is standard FBM with the Hurst parameter $\boldsymbol{H}$ (see Section 2.2.3 for the definition of FBM and the existence of such $(\Omega, \mathcal{A})$, $\mathbf{B}$ and $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H}\in(0,1)}$).

Let the process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be defined by

$$X_t \;:=\; B(t)$$

for $t \in \mathbb{Z}$. We may regard $\mathbf{X}$ as an *equidistant discretization* of FBM with the *sampling interval length* $\delta = 1$. As first observed in Bandt and Shiha [12], the sampling interval length does not have any effect on the distribution of ordinal patterns. In particular, by the self-similarity of FBM (see (2.2), p. 37),

$$(B(t\delta))_{t\in\mathbb{Z}} \;\overset{\text{dist}}{=}\; (\delta^{\boldsymbol{H}} B(t))_{t\in\mathbb{Z}}$$

for $\boldsymbol{H} \in (0,1)$ and $\delta > 0$. Since $\pi(\mathbf{x}) = \pi(\delta^{\boldsymbol{H}} \mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^{d+1}$, the ordinal patterns in $(B(t\delta))_{t\in\mathbb{Z}}$ and $(\delta^{\boldsymbol{H}} B(t))_{t\in\mathbb{Z}}$ are identical, and thus they have the same distribution.

Let $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ with $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$ be the increment process of $\mathbf{X}$. Note that $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is standard Fractional Gaussian Noise (FGN) with the Hurst parameter $\boldsymbol{H}$ (see Section 2.2.3). According to the properties of standard FGN, $\mathbf{Y}$ is non-degenerate, stationary, zero-mean Gaussian and has unit variance for every $\boldsymbol{H} \in (0,1)$. Thus, with $\boldsymbol{\Theta} := (0,1)$ and $\boldsymbol{\vartheta} := \boldsymbol{H}$, we have a class of stochastic processes as in Section 3.1, particularly with $\mathbf{Y}$ satisfying the model assumptions (M1)-(M3) for every $\boldsymbol{H} \in (0,1)$. We do not consider equidistant discretizations of FBM with the Hurst parameter $\boldsymbol{H} = 1$, because $\mathbf{Y}$ would be degenerate in this case (see Section 2.2.3) and thus not meet the model assumption (M1).

The following corollary summarizes properties of $\hat{p}_{\mathbf{r},n}$.

Corollary 3.18. *Let $d \in \mathbb{N}$ and $\mathbf{r} \in S_d$.*

(*i*) *$\hat{p}_{\mathbf{r},n}$ is an unbiased estimator of $p_{\mathbf{r}}(\cdot)$.*

(*ii*) *If $h : [0,1] \to \mathbb{R}$ is continuous on an open set containing the image of $(0,1)$ under $p_{\mathbf{r}}(\cdot)$, then $h(\hat{p}_{\mathbf{r},n})$ is a strongly consistent estimator of $h(p_{\mathbf{r}}(\cdot))$. If, additionally, h is bounded on $[0,1]$, then $h(\hat{p}_{\mathbf{r},n})$ is an asymptotically unbiased estimator of $h(p_{\mathbf{r}}(\cdot))$.*

(iii) If $\boldsymbol{H} < \frac{3}{4}$ *and* $h : [0,1] \to \mathbb{R}$ *has a non-vanishing first derivative at* $p_{\mathbf{r}}(\boldsymbol{H})$*, then*

$$\sqrt{n}\,\big(\,h(\hat{p}_{\mathbf{r},n}) - h(p_{\mathbf{r}}(\boldsymbol{H}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{H}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{H}}[h'(p_{\mathbf{r}}(\boldsymbol{H}))]^2\,)\,,$$

where

$$\sigma^2_{\boldsymbol{H}} \;:=\; \gamma_{\boldsymbol{H}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{H}}(k)$$

and $\gamma_{\boldsymbol{H}}(k) := \frac{1}{(\sharp\,\bar{\mathbf{r}})^2}\,\mathrm{Cov}_{\boldsymbol{H}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}})$ *for* $k \in \mathbb{Z}$*.*

Proof. (i) follows by Theorem 3.5.

(ii) is a consequence of Theorem 3.8 and the fact that $\rho_{\boldsymbol{H}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{H} \in (0,1)$ (see Lemma 2.7 (ii)).

(iii) If $\boldsymbol{H} < \frac{3}{4}$, then there exists a $\beta > \frac{1}{2}$ with $|\rho_{\boldsymbol{H}}(k)| = o(k^{-\beta})$ (for instance, we can choose $\beta := \frac{5}{4} - \boldsymbol{H}$, see Lemma 2.7 (ii)). Thus, the statement follows by Corollary 3.17. □

In the following, let $\boldsymbol{\Pi}_{\boldsymbol{H}}$ with $\boldsymbol{H} \in (0,1)$ denote the distribution of the ordinal pattern process $(\Pi(t))_{t\in\mathbb{Z}}$ in an equidistant discretization of FBM with the Hurst parameter $\boldsymbol{H}$.

ARFIMA(0,d,0) processes. Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a family of measurable mappings from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. Suppose $(\mathbb{P}_{\boldsymbol{d}})_{\boldsymbol{d}\in(-\frac{1}{2},\frac{1}{2})}$ is a family of probability measures on $(\Omega, \mathcal{A})$ such that the process of increments $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$, given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$, measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ is a standard ARFIMA(0,d,0) process with the fractional differencing parameter $\boldsymbol{d}$.

According to the properties of ARFIMA(0,d,0) processes discussed in Section 2.2.4, $\mathbf{Y}$ is non-degenerate, stationary, zero-mean Gaussian with unit variance for every $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$. Thus, with $\boldsymbol{\Theta} := (-\frac{1}{2}, \frac{1}{2})$ and $\boldsymbol{\vartheta} := \boldsymbol{d}$, we have a class of stochastic process as defined in Section 3.1. In particular, $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) for every $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$. Therefore, we obtain the following statement.

Corollary 3.19. *Let* $d \in \mathbb{N}$ *and* $\mathbf{r} \in S_d$*.*

(i) $\hat{p}_{\mathbf{r},n}$ *is an unbiased estimator of* $p_{\mathbf{r}}(\cdot)$*.*

(ii) If $h : [0,1] \to \mathbb{R}$ *is continuous on an open set containing the image of* $(-\frac{1}{2}, \frac{1}{2})$ *under* $p_{\mathbf{r}}(\cdot)$*, then* $h(\hat{p}_{\mathbf{r},n})$ *is a strongly consistent estimator of* $h(p_{\mathbf{r}}(\cdot))$*. If, additionally,* h *is bounded on* $[0,1]$*, then* $h(\hat{p}_{\mathbf{r},n})$ *is an asymptotically unbiased estimator of* $h(p_{\mathbf{r}}(\cdot))$*.*

(*iii*) *If* $\boldsymbol{d} < \frac{1}{4}$ *and* $h : [0,1] \to \mathbb{R}$ *has a non-vanishing first derivative at* $p_{\mathbf{r}}(\boldsymbol{d})$, *then*

$$\sqrt{n}\,\big(\,h(\hat{p}_{\mathbf{r},n}) - h(p_{\mathbf{r}}(\boldsymbol{d}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{d}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{d}}[h'(p_{\mathbf{r}}(\boldsymbol{d}))]^2\,)\,,$$

where

$$\sigma^2_{\boldsymbol{d}} \quad := \quad \gamma_{\boldsymbol{d}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{d}}(k)$$

and $\gamma_{\boldsymbol{d}}(k) := \frac{1}{(\sharp\,\bar{\mathbf{r}})^2}\,\mathrm{Cov}_{\boldsymbol{d}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}})$ *for* $k \in \mathbb{Z}$.

Proof. The proof is similar to that of Corollary 3.18. In particular, according to Lemma 2.8 (*i*), we have $\rho_{\boldsymbol{d}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$, and if $\boldsymbol{d} < \frac{1}{4}$, then $|\rho_{\boldsymbol{d}}(k)| = o(k^{-\beta})$ for $\beta := \frac{3}{4} - \boldsymbol{d}$. □

AR(1) processes. Now, let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a family of measurable mappings from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. Suppose $(\mathbb{P}_{\boldsymbol{a}})_{\boldsymbol{a}\in(-1,1)}$ is a family of probability measures on $(\Omega, \mathcal{A})$ such that the increment process $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$, given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$, measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ is a standard AR(1) process with the autoregressive coefficient $\boldsymbol{a}$.

According to the properties of AR(1) processes discussed in Section 2.2.5, $\mathbf{Y}$ is non-degenerate, stationary, zero-mean Gaussian and has unit variance for every $\boldsymbol{a} \in (-1, 1)$. Thus, with $\boldsymbol{\Theta} := (-1, 1)$ and $\boldsymbol{\vartheta} := \boldsymbol{a}$, we have a class of stochastic process as defined in Section 3.1. In particular, $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) for every $\boldsymbol{a} \in (-1, 1)$.

Corollary 3.20. *Let* $d \in \mathbb{N}$ *and* $\mathbf{r} \in S_d$.

(*i*) $\hat{p}_{\mathbf{r},n}$ *is an unbiased estimator of* $p_{\mathbf{r}}(\cdot)$.

(*ii*) *If* $h : [0,1] \to \mathbb{R}$ *is continuous on an open set containing the image of* $(-1,1)$ *under* $p_{\mathbf{r}}(\cdot)$, *then* $h(\hat{p}_{\mathbf{r},n})$ *is a strongly consistent estimator of* $h(p_{\mathbf{r}}(\cdot))$. *If, additionally,* h *is bounded on* $[0,1]$, *then* $h(\hat{p}_{\mathbf{r},n})$ *is an asymptotically unbiased estimator of* $h(p_{\mathbf{r}}(\cdot))$.

(*iii*) *For every* $\boldsymbol{a} \in (-1,1)$, *if* $h : [0,1] \to \mathbb{R}$ *has a non-vanishing first derivative at* $p_{\mathbf{r}}(\boldsymbol{a})$, *then*

$$\sqrt{n}\,\big(\,h(\hat{p}_{\mathbf{r},n}) - h(p_{\mathbf{r}}(\boldsymbol{a}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{a}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{a}}[h'(p_{\mathbf{r}}(\boldsymbol{a}))]^2\,)\,,$$

where

$$\sigma^2_{\boldsymbol{a}} \quad := \quad \gamma_{\boldsymbol{a}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{a}}(k)$$

and $\gamma_{\boldsymbol{a}}(k) := \frac{1}{(\sharp\,\bar{\mathbf{r}})^2}\,\mathrm{Cov}_{\boldsymbol{a}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}})$ *for* $k \in \mathbb{Z}$.

Proof. The proof is similar to that of Corollary 3.18. In particular, $|\rho_{\boldsymbol{a}}(k)| \to 0$ as $k \to \infty$ with an exponential rate of decrease for every $\boldsymbol{a} \in (-1, 1)$ (see Section 2.2.5). Thus, $|\rho_{\boldsymbol{a}}(k)| = o(k^{-\beta})$ for any $\beta > \frac{1}{2}$. □

Chapter 4

Covariances of zero crossings

In this chapter we investigate the covariances of zero crossing indicator variables. The framework of analysis is the same as in Chapter 3, namely, $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a family of real-valued measurable mappings defined on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a non-empty family of probability measures $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}$. By $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ we denote the process of increments given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$. **We always assume that $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) on p. 42.** For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z}$, let $\rho_{\boldsymbol{\vartheta}}(k) = \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0, Y_k)$ denote the autocorrelations of $\mathbf{Y}$.

As we show in Chapter 5, any ordinal pattern probability of order $d = 2$ is an affine function of the probability of a zero crossing. The results of this chapter can be used for evaluating the variance of the estimators of ordinal pattern probabilities.

Zero crossings. The indicator for a *zero crossing* in $\mathbf{Y}$ at time $t \in \mathbb{Z}$ is given by

$$C(t) \;:=\; \mathbf{1}_{\{Y_{t+1}\leq 0,\, Y_{t+2}>0\}} + \mathbf{1}_{\{Y_{t+1}>0,\, Y_{t+2}\leq 0\}}\,. \tag{4.1}$$

Note that, according to assumption (M1), the probability that $Y_{t+1} = 0$ or $Y_{t+2} = 0$ is equal to 0 for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. As a consequence of assumption (M2), the process $(C(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Therefore, the probability for a zero crossing, given by

$$c(\boldsymbol{\vartheta}) \;:=\; \mathbb{P}_{\boldsymbol{\vartheta}}(C(t) = 1)$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, does not depend on the specific time point t on the right hand side of the definition. An estimator for $c(\cdot)$ is given by

$$\hat{c}_n \;:=\; \frac{1}{n}\sum_{t=0}^{n-1} C(t)\,. \tag{4.2}$$

Statistical properties of $\hat{c}_n$ will be further investigated in Chapter 5. Here, we focus on the evaluation of the variance of $\hat{c}_n$. Clearly, by the stationarity of $(C(t))_{t\in\mathbb{Z}}$, we obtain

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;=\; \frac{1}{n^2}\Big(n\,\gamma_{\boldsymbol{\vartheta}}(0) + 2\sum_{k=1}^{n-1}(n-k)\,\gamma_{\boldsymbol{\vartheta}}(k)\Big) \tag{4.3}$$

for $n \in \mathbb{N}$, where

$$\gamma_{\boldsymbol{\vartheta}}(k) \;:=\; \mathrm{Cov}_{\boldsymbol{\vartheta}}(C(0), C(k))$$

for $k \in \mathbb{Z}$. In Section 4.1, we derive closed-form expressions for $\gamma_{\boldsymbol{\vartheta}}(0)$ and $\gamma_{\boldsymbol{\vartheta}}(1)$. For $k > 1$, we give a representation of $\gamma_{\boldsymbol{\vartheta}}(k)$ in terms of four-dimensional normal orthant probabilities. The numerical evaluation of these expressions is investigated in Section 4.2. We derive asymptotic properties of the zero crossing covariances in Section 4.3, and provide bounds and approximations in Section 4.4. Asymptotic properties of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ are investigated in Sections 4.5 and 4.6.

4.1 Closed-form expressions

Normal orthant probabilities. Let $n \in \mathbb{N}$ and suppose $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ is symmetric and strictly positive definite. The n-dimensional *normal orthant probability* with respect to $\boldsymbol{\Sigma}$ is given by

$$\Phi(\boldsymbol{\Sigma}) \;:=\; \int_{[0,\infty)^n} \phi(\mathbf{0},\, \boldsymbol{\Sigma},\, \mathbf{x})\,\mathrm{d}\mathbf{x}\,,$$

where $\phi(\mathbf{0},\, \boldsymbol{\Sigma},\, \cdot)$ denotes the Lebesgue density of the n-dimensional normal distribution with means $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ (see Theorem 2.4 (i)).

Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ be a non-degenerate zero-mean Gaussian random vector with $\mathrm{Cov}(\mathbf{Z}) = \boldsymbol{\Sigma}$. (Throughout this chapter, we assume any random vector $\mathbf{Z}$, $\mathbf{Z}'$, ... is defined on some probability space $(\Omega', \mathcal{A}', \mathbb{P})$.) Clearly,

$$\Phi(\boldsymbol{\Sigma}) \;=\; \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, \ldots,\, Z_n \geq 0)\,.$$

Let $a_1, a_2, \ldots, a_n > 0$ and $\mathbf{A} = \mathbf{diag}(\sqrt{a_1},\, \sqrt{a_2},\, \ldots,\, \sqrt{a_n}\,)$. Since

$$\mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, \ldots,\, Z_n \geq 0) \;=\; \mathbb{P}(\sqrt{a_1}\,Z_1 \geq 0,\, \sqrt{a_2}\,Z_2 \geq 0,\, \ldots,\, \sqrt{a_n}\,Z_n \geq 0)$$

and $\mathrm{Cov}(\sqrt{a_1}\,Z_1,\, \sqrt{a_2}\,Z_2,\, \ldots,\, \sqrt{a_n}\,Z_n) = \mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}$, we have $\Phi(\boldsymbol{\Sigma}) = \Phi(\mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A})$. By choosing $a_i = (\mathrm{Var}(Z_i))^{-1}$ for $i = 1, 2, \ldots, n$, we obtain

$$\Phi(\mathrm{Cov}(\mathbf{Z})) \;=\; \Phi(\mathrm{Corr}(\mathbf{Z}))\,.$$

Thus, only the correlation structure of a non-degenerate zero-mean Gaussian random vector is relevant for the probability that all components simultaneously exceed the level 0.

The following expressions for two- and three-dimensional normal orthant probabilities are well-known (see Bacon [8]).

Lemma 4.1. *Let (Z_1, Z_2, Z_3) be a zero-mean non-degenerate Gaussian random vector and $\rho_{ij} = \mathrm{Corr}(Z_i, Z_j)$ for $i, j \in \{1, 2, 3\}$. Then*

$$\begin{aligned} \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0) &= \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{12}\,, \\ \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, Z_3 \geq 0) &= \frac{1}{8} + \frac{1}{4\pi} \arcsin \rho_{12} + \frac{1}{4\pi} \arcsin \rho_{13} + \frac{1}{4\pi} \arcsin \rho_{23}\,. \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1) &= \mathbb{P}_{\boldsymbol{\vartheta}}(-Y_1 \geq 0,\, Y_2 \geq 0) + \mathbb{P}_{\boldsymbol{\vartheta}}(Y_1 \leq 0,\, -Y_2 \leq 0) \\ &= 2\,\mathbb{P}_{\boldsymbol{\vartheta}}(Y_1 \geq 0,\, -Y_2 \geq 0)\,, \end{aligned}$$

where the first equality follows because $\mathbb{P}_{\boldsymbol{\vartheta}}(Y_1 = 0) = \mathbb{P}_{\boldsymbol{\vartheta}}(Y_2 = 0) = 0$, and the second one because $(-Y_1, Y_2)$ and $(Y_1, -Y_2)$ are zero-mean Gaussian with the same covariance structure and hence identically distributed. Since $x \mapsto \arcsin x$ is an odd function and $\mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_1, -Y_2) = -\rho_{\boldsymbol{\vartheta}}(1)$, Lemma 4.1 yields

$$c(\boldsymbol{\vartheta}) = \frac{1}{2} - \frac{1}{\pi} \arcsin \rho_{\boldsymbol{\vartheta}}(1)\,. \tag{4.4}$$

Clearly, the variance of $C(0)$ is given by

$$\begin{aligned} \gamma_{\boldsymbol{\vartheta}}(0) &= \mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1)\,(1 - \mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1)) \\ &= \frac{1}{4} - \frac{1}{\pi^2} (\arcsin \rho_{\boldsymbol{\vartheta}}(1))^2\,. \end{aligned} \tag{4.5}$$

By the same argument as above, $(Y_1, -Y_2, Y_3)$ and $(-Y_1, Y_2, -Y_3)$ are identically distributed. Thus,

$$\mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1,\, C(1) = 1) = 2\,\mathbb{P}_{\boldsymbol{\vartheta}}(Y_1 > 0,\, -Y_2 > 0,\, Y_3 > 0)\,.$$

Since $\mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_1, -Y_2) = \mathrm{Corr}_{\boldsymbol{\vartheta}}(-Y_2, Y_3) = -\rho_{\boldsymbol{\vartheta}}(1)$, we obtain

$$\begin{aligned} \gamma_{\boldsymbol{\vartheta}}(1) &= \mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1,\, C(1) = 1) - (\mathbb{P}_{\boldsymbol{\vartheta}}(C(0) = 1))^2 \\ &= \frac{1}{2\pi} \arcsin \rho_{\boldsymbol{\vartheta}}(2) - \frac{1}{\pi^2} (\arcsin \rho_{\boldsymbol{\vartheta}}(1))^2\,. \end{aligned} \tag{4.6}$$

The four-dimensional case. For $k > 1$, we can express $\gamma_{\vartheta}(k)$ as the sum and product, respectively, of two- and four-dimensional normal orthant probabilities: Note that

$$\begin{aligned} \gamma_{\vartheta}(k) &= \mathrm{Cov}_{\vartheta}(1 - C(0), 1 - C(k)) \\ &= \mathbb{P}_{\vartheta}(C(0) = 0,\, C(k) = 0) - (\mathbb{P}_{\vartheta}(C(0) = 0))^2 \,. \end{aligned}$$

Using the equivalent expressions $\{Y_1 \le 0,\, Y_2 \le 0\} \cup \{Y_1 > 0,\, Y_2 > 0\}$ for $\{C(0) = 0\}$ and $\{Y_{k+1} \le 0,\, Y_{k+2} \le 0\} \cup \{Y_{k+1} > 0,\, Y_{k+2} > 0\}$ for $\{C(k) = 0\}$, respectively, and identifying equal probabilities, we obtain

$$\begin{aligned} \gamma_{\vartheta}(k) &= 2\,\mathbb{P}_{\vartheta}(Y_1 \ge 0,\, Y_2 \ge 0,\, Y_{k+1} \ge 0,\, Y_{k+2} \ge 0) \\ &+ 2\,\mathbb{P}_{\vartheta}(Y_1 \ge 0,\, Y_2 \ge 0,\, -Y_{k+1} \ge 0,\, -Y_{k+2} \ge 0) \\ &- 4\,(\mathbb{P}_{\vartheta}(Y_1 \ge 0,\, Y_2 \ge 0))^2 \,. \end{aligned} \tag{4.7}$$

Up to now, no closed-form expression is known for normal orthant probabilities of dimension $n \ge 4$. Abrahamson [2] derives a formula involving two-dimensional integrals for the special case of orthoscheme probabilities, where certain entries of the covariance matrix $\mathbf{\Sigma}$ are equal to 0. As Abrahamson [2] shows, any four-dimensional normal orthant probability can be written as a linear combination of six orthoscheme probabilities. Cheng [27] proposes an expression involving the dilogarithm function in the case where certain entries of $\mathbf{\Sigma}$ are identical.

Damsleth and El-Shaarawi [31] gives a formula involving two-dimensional integrals when $\mathbf{\Sigma}$ is the covariance matrix of four variables in a stationary process (that is, $\mathbf{\Sigma}$ is a principal submatrix of a symmetric and strictly positive definite Toeplitz matrix). For the special case where the stationary process is an AR(1) process, Damsleth and El-Shaarawi [31] proposes an approximation which seems to work well as long as the autoregressive coefficient is not too large. Recent approaches to the evaluation of normal orthant probabilities of dimension four and higher use Monte Carlo sampling (see Craig [30] for an overview). Note that, according to a recursive formula given in David [32], five-dimensional normal orthant probabilities can be expressed as the sum of at most four-dimensional normal orthant probabilities.

Framework of analysis. We consider the problem of evaluating $\gamma_{\vartheta}(k)$ for $k > 1$ in the following context: By $\mathcal{R}$ we denote the set of $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in [-1, 1]^6$ for which the matrix

$$\mathbf{\Sigma}(\mathbf{r}) \;:=\; \begin{pmatrix} 1 & r_1 & r_2 & r_3 \\ r_1 & 1 & r_4 & r_5 \\ r_2 & r_4 & 1 & r_6 \\ r_3 & r_5 & r_6 & 1 \end{pmatrix}$$

is strictly positive definite, that is, $\mathbf{x}\,\boldsymbol{\Sigma}(\mathbf{r})\,\mathbf{x}^T > 0$ for all $\mathbf{x} \in \mathbb{R}^4 \setminus \{\mathbf{0}\}$. Note that $\boldsymbol{\Sigma}(\mathcal{R})$ is the set of correlation matrices of four-dimensional non-degenerate Gaussian random vectors (see Theorem 2.4 (i)). Thus, $\mathbf{r} \in \mathcal{R}$ implies that all components of $\mathbf{r}$ lie in $(-1, 1)$. Furthermore, if $\mathbf{r}, \mathbf{s} \in \mathcal{R}$, then

$$\mathbf{x}\,\boldsymbol{\Sigma}(h \cdot \mathbf{r} + (1-h) \cdot \mathbf{s})\,\mathbf{x}^T \;=\; h\,\mathbf{x}\,\boldsymbol{\Sigma}(\mathbf{r})\,\mathbf{x}^T + (1-h)\,\mathbf{x}\,\boldsymbol{\Sigma}(\mathbf{s})\,\mathbf{x}^T \;>\; 0$$

for all $\mathbf{x} \in \mathbb{R}^4 \setminus \{\mathbf{0}\}$ and $h \in [0, 1]$, which shows that $\mathcal{R}$ is convex.

For $h \in [-1, 1]$, define

$$\mathbf{I}_h \;:=\; \mathbf{diag}(1, h, h, h, h, 1).$$

We also use the symbol $\mathbf{I}_h$ to denote the mapping $\mathbf{r} \mapsto \mathbf{I}_h\,\mathbf{r}$ from $\mathcal{R}$ onto itself. Next, we establish closedness of $\mathcal{R}$ with respect to $\mathbf{I}_h$ for all $h \in [-1, 1]$. Let $\mathbf{r} \in \mathcal{R}$. We first show that $\mathbf{I}_{-1}\,\mathbf{r} \in \mathcal{R}$. Let $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)$ be a zero-mean non-degenerate Gaussian random vector with $\mathrm{Corr}(\mathbf{Z}) = \boldsymbol{\Sigma}(\mathbf{r})$. Define $\mathbf{Z}' := (Z_1, Z_2, -Z_3, -Z_4)$ and note that $\mathrm{Corr}(\mathbf{Z}') = \boldsymbol{\Sigma}(\mathbf{I}_{-1}\,\mathbf{r})$. Since $\mathbf{Z}'$ is also non-degenerate Gaussian, $\boldsymbol{\Sigma}(\mathbf{I}_{-1}\,\mathbf{r})$ is strictly positive definite and thus $\mathbf{I}_{-1}\,\mathbf{r} \in \mathcal{R}$. Now, let $h \in [-1, 1]$. Since $\mathbf{I}_1\,\mathbf{r} = \mathbf{r}$ and

$$\mathbf{I}_h\,\mathbf{r} \;=\; \frac{1+h}{2}\,\mathbf{I}_1\,\mathbf{r} + \frac{1-h}{2}\,\mathbf{I}_{-1}\,\mathbf{r}\,,$$

we obtain $\mathbf{I}_h\,\mathbf{r} \in \mathcal{R}$ by the convexity of $\mathcal{R}$.

Since $\boldsymbol{\Sigma}(\mathbf{r})$ is strictly positive definite and symmetric, $\Phi(\boldsymbol{\Sigma}(\mathbf{r}))$ is well-defined for every $\mathbf{r} \in \mathcal{R}$. To simplify notation, we write

$$\Phi(\mathbf{r}) \;=\; \Phi(\boldsymbol{\Sigma}(\mathbf{r}))\,.$$

Now, consider the mapping $\Psi : \mathcal{R} \to \mathbb{R}$ given by

$$\Psi(\mathbf{r}) \;:=\; 2\,\Phi(\mathbf{r}) + 2\,\Phi(\mathbf{I}_{-1}\,\mathbf{r}) - 4\,\Phi(\mathbf{I}_0\,\mathbf{r}) \tag{4.8}$$

for $\mathbf{r} \in \mathcal{R}$. Since $\mathbf{I}_h\,\mathbf{r} \in \mathcal{R}$ for all $h \in [-1, 1]$, Ψ is well-defined. The reason for introducing Ψ is the following: Let $\mathbf{r} \in \mathcal{R}$ and suppose (Z_1, Z_2, Z_3, Z_4) is a zero-mean non-degenerate Gaussian random vector with correlation matrix $\boldsymbol{\Sigma}(\mathbf{r})$. Then

$$\Phi(\mathbf{r}) \;=\; \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, Z_3 \geq 0,\, Z_4 \geq 0)\,.$$

Furthermore, $\boldsymbol{\Sigma}(\mathbf{I}_{-1}\,\mathbf{r})$ is the correlation matrix of $(Z_1, Z_2, -Z_3, -Z_4)$ and hence

$$\Phi(\mathbf{I}_{-1}\,\mathbf{r}) \;=\; \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, -Z_3 \geq 0,\, -Z_4 \geq 0).$$

Finally, $\boldsymbol{\Sigma}(\mathbf{I}_0\,\mathbf{r})$ is the correlation matrix of (Z_1', Z_2', Z_3', Z_4') with (Z_1', Z_2') and (Z_3', Z_4') being uncorrelated (and thus independent) and having the same distribution as (Z_1, Z_2) and (Z_3, Z_4), respectively. Therefore,

$$\Phi(\mathbf{I}_0\,\mathbf{r}) \;=\; \mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0)\,\mathbb{P}(Z_3 \geq 0,\, Z_4 \geq 0)\,.$$

Putting it all together, we have

$$
\begin{aligned}
(4.9) \qquad \Psi(\mathbf{r}) &= 2\,\mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, Z_3 \geq 0,\, Z_4 \geq 0) \\
&+ 2\,\mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0,\, -Z_3 \geq 0,\, -Z_4 \geq 0) \\
&- 4\,\mathbb{P}(Z_1 \geq 0,\, Z_2 \geq 0)\,\mathbb{P}(Z_3 \geq 0,\, Z_4 \geq 0)\,.
\end{aligned}
$$

Now, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k > 1$. By comparing (4.9) with (4.7) and noting that $\boldsymbol{\Sigma}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(k+1),\, \rho_{\boldsymbol{\vartheta}}(k-1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(1))$ is the correlation matrix of $(Y_1, Y_2, Y_{k+1}, Y_{k+2})$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$, we obtain

$$
(4.10) \qquad \gamma_{\boldsymbol{\vartheta}}(k) = \Psi(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(k+1),\, \rho_{\boldsymbol{\vartheta}}(k-1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(1))\,.
$$

Therefore, we study properties of Ψ and Φ in the following.

Note that if our purpose was only evaluating $\gamma_{\boldsymbol{\vartheta}}(k)$, we could restrict our investigations to $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$ satisfying $r_1 = r_6$ and $r_2 = r_5$. However, the problem of evaluating four-dimensional orthant probabilities is interesting in its own right, therefore, we study it in a more general context.

Basic properties. The following Lemma establishes basic equations for Ψ and closed-form expressions for Ψ and Φ in some special cases.

Lemma 4.2.

(*i*) *For every* $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$,

$$
\begin{aligned}
\Psi(\mathbf{I}_{-1}\,\mathbf{r}) = \Psi(\mathbf{r}) &= \Psi(-r_1,\, -r_2,\quad r_3,\quad r_4,\, -r_5,\, -r_6) \\
&= \Psi(-r_1,\quad r_2,\, -r_3,\, -r_4,\quad r_5,\, -r_6)\,.
\end{aligned}
$$

(*ii*) *If* $r_2 = r_3 = r_4 = r_5 = 0$, *then* $\Psi(\mathbf{r}) = 0$ *and*

$$
\Phi(\mathbf{r}) = \Big(\frac{1}{4} + \frac{1}{2\pi}\arcsin r_1\Big)\Big(\frac{1}{4} + \frac{1}{2\pi}\arcsin r_6\Big)\,.
$$

Proof. (*i*) The first equation is obtained by the definition of Ψ and because $\mathbf{I}_0\,\mathbf{I}_{-1} = \mathbf{I}_0$. In order to show the second equation, let (Z_1, Z_2, Z_3, Z_4) be zero-mean Gaussian with correlation matrix $\boldsymbol{\Sigma}(\mathbf{r})$. Furthermore, let $\mathbf{r}' := (-r_1, -r_2, r_3, r_4, -r_5, -r_6)$. Since $\boldsymbol{\Sigma}(\mathbf{r}')$ is the correlation matrix of $(Z_1, -Z_2, -Z_3, Z_4)$, we have

$$
\begin{aligned}
\Psi(\mathbf{r}) &= \mathrm{Cov}(\mathbf{1}_{\{Z_1 \geq 0, Z_2 \geq 0\}} + \mathbf{1}_{\{Z_1 \leq 0, Z_2 \leq 0\}},\, \mathbf{1}_{\{Z_3 \geq 0, Z_4 \geq 0\}} + \mathbf{1}_{\{Z_3 \leq 0, Z_4 \leq 0\}}) \\
&= \mathrm{Cov}(\mathbf{1}_{\{Z_1 \geq 0, -Z_2 \leq 0\}} + \mathbf{1}_{\{Z_1 \leq 0, -Z_2 \geq 0\}},\, \mathbf{1}_{\{Z_3 \geq 0, -Z_4 \leq 0\}} + \mathbf{1}_{\{Z_3 \leq 0, -Z_4 \geq 0\}}) \\
&= \Psi(\mathbf{r}')\,,
\end{aligned}
$$

and the result follows. Applying $\Psi(\mathbf{r}) = \Psi(\mathbf{I}_{-1}\,\mathbf{r})$ to $\mathbf{r} = (-r_1, -r_2, r_3, r_4, -r_5, -r_6)$ yields the third equation.

(*ii*) Under the assumptions, $\mathbf{r} = \mathbf{I}_{-1}\,\mathbf{r} = \mathbf{I}_0\,\mathbf{r}$ which shows that $\Psi(\mathbf{r}) = 0$. Furthermore, if (Z_1, Z_2, Z_3, Z_4) is Gaussian with the correlation matrix $\mathbf{\Sigma}(\mathbf{r})$, then (Z_1, Z_2) and (Z_3, Z_4) are independent and hence $\Phi(\mathbf{r}) = \mathbb{P}(Z_1 \geq 0, Z_2 \geq 0)\,\mathbb{P}(Z_3 \geq 0, Z_4 \geq 0)$. Since $\mathrm{Corr}(Z_1, Z_2) = r_1$ and $\mathrm{Corr}(Z_3, Z_4) = r_6$, the result follows by Lemma 4.1. □

Note that bounds for $\Psi(\mathbf{r})$ can be obtained by the Berman-inequality, namely,

$$|\Psi(\mathbf{r})| \;\leq\; \frac{2}{\pi} \sum_{k=2}^{5} \frac{|r_k|}{\sqrt{1 - r_k}}\,.$$

(see Theorem 2.6.1 in Berman [16] and Theorem C.2 in Piterbarg [80]).

4.2 Numerical evaluation

Partial derivatives. In this section, we derive representations of Ψ and Φ by one-dimensional integrals which can easily be evaluated numerically. We begin our analysis by determining partial derivatives of Φ and Ψ.

Lemma 4.3. *For every* $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$,

$$\begin{aligned}
\frac{\partial \Phi}{\partial r_2}(\mathbf{r}) &= \frac{1}{2\pi\sqrt{1 - r_2^2}} \Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{\sigma_{24}(\mathbf{r})}{\sqrt{\sigma_{22}(\mathbf{r})\sigma_{44}(\mathbf{r})}} \Big)\,, \\
\frac{\partial \Phi}{\partial r_3}(\mathbf{r}) &= \frac{1}{2\pi\sqrt{1 - r_3^2}} \Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{\sigma_{23}(\mathbf{r})}{\sqrt{\sigma_{22}(\mathbf{r})\sigma_{33}(\mathbf{r})}} \Big)\,, \\
\frac{\partial \Phi}{\partial r_4}(\mathbf{r}) &= \frac{1}{2\pi\sqrt{1 - r_4^2}} \Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{\sigma_{14}(\mathbf{r})}{\sqrt{\sigma_{11}(\mathbf{r})\sigma_{44}(\mathbf{r})}} \Big)\,, \\
\frac{\partial \Phi}{\partial r_5}(\mathbf{r}) &= \frac{1}{2\pi\sqrt{1 - r_5^2}} \Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{\sigma_{13}(\mathbf{r})}{\sqrt{\sigma_{11}(\mathbf{r})\sigma_{33}(\mathbf{r})}} \Big)\,,
\end{aligned}$$

where

$$\begin{aligned}
\sigma_{11}(\mathbf{r}) &= 1 - r_4^2 - r_5^2 - r_6^2 + 2r_4r_5r_6\,, \\
\sigma_{22}(\mathbf{r}) &= 1 - r_2^2 - r_3^2 - r_6^2 + 2r_2r_3r_6\,, \\
\sigma_{33}(\mathbf{r}) &= 1 - r_1^2 - r_3^2 - r_5^2 + 2r_1r_3r_5\,, \\
\sigma_{44}(\mathbf{r}) &= 1 - r_1^2 - r_2^2 - r_4^2 + 2r_1r_2r_4\,, \\
\sigma_{13}(\mathbf{r}) &= r_2 - r_1r_4 + r_3r_4r_5 - r_2r_5^2 - r_3r_6 + r_1r_5r_6\,, \\
\sigma_{14}(\mathbf{r}) &= r_3 - r_1r_5 + r_2r_4r_5 - r_3r_4^2 - r_2r_6 + r_1r_4r_6\,, \\
\sigma_{23}(\mathbf{r}) &= r_4 - r_1r_2 + r_2r_3r_5 - r_4r_3^2 - r_5r_6 + r_1r_3r_6\,, \\
\sigma_{24}(\mathbf{r}) &= r_5 - r_1r_3 + r_2r_3r_4 - r_5r_2^2 - r_4r_6 + r_1r_2r_6\,.
\end{aligned}$$

Proof. Let $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$. For $i, j \in \{1, 2, 3, 4\}$, let $\sigma'_{ij}(\mathbf{r})$ denote the (i, j)-th component of $(\mathbf{\Sigma}(\mathbf{r}))^{-1}$. As a well-known fact, the inverse and any principal submatrix of a symmetric strictly positive definite matrix are symmetric and strictly positive definite (see Horn and Johnson [49], pp. 169, 397). Let $k \in \{2, 3, \ldots, 5\}$. Suppose $\{i, j\}$ with $i \neq j$ is the unique subset of $\{1, 2, 3, 4\}$ such that r_k does not lie in the i-th row and j-th column of $\mathbf{\Sigma}(\mathbf{r})$. By the so-called reduction formula for normal orthant probabilities (see Plackett [81], Berman [16], p. 31), we obtain

$$\frac{\partial \Phi}{\partial r_k}(\mathbf{r}) = \frac{1}{2\pi\sqrt{1 - r_k^2}} \Phi\left(\begin{pmatrix} \sigma'_{ii}(\mathbf{r}) & \sigma'_{ij}(\mathbf{r}) \\ \sigma'_{ij}(\mathbf{r}) & \sigma'_{jj}(\mathbf{r}) \end{pmatrix}^{-1} \right) . \tag{4.11}$$

Note that the matrix in the argument of Φ on the right hand side of (4.11) is the inverse of a principal submatrix of $(\mathbf{\Sigma}(\mathbf{r}))^{-1}$ and hence symmetric and strictly positive definite. Computing the inverse yields

$$\begin{pmatrix} \sigma'_{ii}(\mathbf{r}) & \sigma'_{ij}(\mathbf{r}) \\ \sigma'_{ij}(\mathbf{r}) & \sigma'_{jj}(\mathbf{r}) \end{pmatrix}^{-1} = \frac{1}{\sigma'_{ii}(\mathbf{r})\sigma'_{jj}(\mathbf{r}) - (\sigma'_{jj}(\mathbf{r}))} \begin{pmatrix} \sigma'_{jj}(\mathbf{r}) & -\sigma'_{ij}(\mathbf{r}) \\ -\sigma'_{ij}(\mathbf{r}) & \sigma'_{ii}(\mathbf{r}) \end{pmatrix} .$$

It is easy to see that if (Z_1, Z_2) is a random vector with the expression on the right hand side of the previous equation as covariance matrix, the correlation coefficient is given by

$$\mathrm{Corr}(Z_1, Z_2) = -\frac{\sigma'_{ij}(\mathbf{r})}{\sqrt{\sigma'_{ii}(\mathbf{r})\sigma'_{jj}(\mathbf{r})}} .$$

Thus, according to Lemma 4.1, we obtain

$$\Phi\left(\begin{pmatrix} \sigma'_{ii}(\mathbf{r}) & \sigma'_{ij}(\mathbf{r}) \\ \sigma'_{ij}(\mathbf{r}) & \sigma'_{jj}(\mathbf{r}) \end{pmatrix}^{-1} \right) = \frac{1}{4} - \frac{1}{2\pi} \arcsin \frac{\sigma'_{ij}(\mathbf{r})}{\sqrt{\sigma'_{ii}(\mathbf{r})\sigma'_{jj}(\mathbf{r})}} . \tag{4.12}$$

Now, for $i, j \in \{1, 2, 3, 4\}$, let $\sigma_{ij}(\mathbf{r})$ be given by $\sigma_{ij}(\mathbf{r}) := -\det(\mathbf{\Sigma}(\mathbf{r}))\, \sigma'_{ij}(\mathbf{r})$ if $i \neq j$, and by $\sigma_{ij}(\mathbf{r}) := \det(\mathbf{\Sigma}(\mathbf{r}))\, \sigma'_{ij}(\mathbf{r})$, otherwise. By combining (4.11) and (4.12), we obtain the expressions for $\frac{\partial \Phi}{\partial r_k}(\mathbf{r})$. Note that $\sigma_{ij}(\mathbf{r})$ is equal to the determinant of the matrix which is obtained by deleting the ith row and the jth column of $\mathbf{\Sigma}(\mathbf{r})$, multiplied with $(-1)^{i+j+1}$ if $i \neq j$ (see Horn and Johnson [49], p. 20). Thus, by elementary calculations, we obtain the expressions for $\sigma_{ij}(\mathbf{r})$. □

As an immediate consequence of Lemma 4.3, any partial derivative of first order of Φ with respect to r_2, r_3, r_4, r_5 is continuous on $\mathcal{R}$. Furthermore, any partial derivative of higher order can be written as the sum, the product and the composition, respectively, of rational functions and derivatives of $x \mapsto \arcsin x$. Therefore, we obtain the following corollary:

Corollary 4.4. *Any partial derivative of Φ with respect to r_2, r_3, r_4, r_5 exists and is continuous on $\mathcal{R}$.*

The following lemma gives partial derivatives of Ψ.

Lemma 4.5. *For every $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$,*

$$\begin{aligned}
\frac{\partial \Psi}{\partial r_2}(\mathbf{r}) &= \frac{1}{\pi^2 \sqrt{1 - r_2^2}} \arcsin \frac{\sigma_{24}(\mathbf{r})}{\sqrt{\sigma_{22}(\mathbf{r}) \sigma_{44}(\mathbf{r})}}\,, \\
\frac{\partial \Psi}{\partial r_3}(\mathbf{r}) &= \frac{1}{\pi^2 \sqrt{1 - r_3^2}} \arcsin \frac{\sigma_{23}(\mathbf{r})}{\sqrt{\sigma_{22}(\mathbf{r}) \sigma_{33}(\mathbf{r})}}\,, \\
\frac{\partial \Psi}{\partial r_4}(\mathbf{r}) &= \frac{1}{\pi^2 \sqrt{1 - r_4^2}} \arcsin \frac{\sigma_{14}(\mathbf{r})}{\sqrt{\sigma_{11}(\mathbf{r}) \sigma_{44}(\mathbf{r})}}\,, \\
\frac{\partial \Psi}{\partial r_5}(\mathbf{r}) &= \frac{1}{\pi^2 \sqrt{1 - r_5^2}} \arcsin \frac{\sigma_{13}(\mathbf{r})}{\sqrt{\sigma_{11}(\mathbf{r}) \sigma_{33}(\mathbf{r})}}\,,
\end{aligned}$$

with $\sigma_{11}(\mathbf{r}), \sigma_{22}(\mathbf{r}), \ldots, \sigma_{24}(\mathbf{r})$ as given in Lemma 4.3.

Proof. Let $\mathbf{r} \in \mathcal{R}$ and $k = 2$. For $k = 3, 4, 5$, the proof is similar. For $h \in [-1, 1]$, let $\mathbf{I}_h$ denote the mapping $\mathbf{r} \mapsto \mathbf{I}_h\, \mathbf{r}$ from $\mathcal{R}$ onto itself. Clearly,

$$\frac{\partial \Psi}{\partial r_2}(\mathbf{r}) = 2\, \frac{\partial \Phi}{\partial r_2}(\mathbf{r}) + 2\, \frac{\partial (\Phi \circ \mathbf{I}_{-1})}{\partial r_2}(\mathbf{r}) - 4\, \frac{\partial (\Phi \circ \mathbf{I}_0)}{\partial r_2}(\mathbf{r})\,. \tag{4.13}$$

According to Lemma 4.2 (*ii*), we have $(\Phi \circ \mathbf{I}_0)(\mathbf{r}) = \big(\frac{1}{4} + \frac{1}{2\pi} \arcsin r_1\big)\big(\frac{1}{4} + \frac{1}{2\pi} \arcsin r_6\big)$. Thus, $\mathbf{r} \mapsto (\Phi \circ \mathbf{I}_0)(\mathbf{r})$ is constant in r_2 and the last term on the right side of (4.13) is equal to 0. Now, note that $\frac{\partial \mathbf{I}_{-1}}{\partial r_2}(\mathbf{r}) = -1$ and hence, by the chain rule of differentiation,

$$\frac{\partial (\Phi \circ \mathbf{I}_{-1})}{\partial r_2}(\mathbf{r}) = -\frac{\partial \Phi}{\partial r_2}(\mathbf{I}_{-1}(\mathbf{r}))\,.$$

Since $\sigma_{22}(\mathbf{I}_{-1}(\mathbf{r})) = \sigma_{22}(\mathbf{r})$, $\sigma_{44}(\mathbf{I}_{-1}(\mathbf{r})) = \sigma_{44}(\mathbf{r})$ and $\sigma_{24}(\mathbf{I}_{-1}(\mathbf{r})) = -\sigma_{24}(\mathbf{r})$, inserting the expressions for $\frac{\partial \Phi}{\partial r_2}(\mathbf{r})$ and $\frac{\partial \Phi}{\partial r_2}(\mathbf{I}_{-1}(\mathbf{r}))$ from Lemma 4.3 into (4.13) yields the result. □

Integral representation. Next, we state the main result of this section. We give representations of $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$ by one-dimensional integrals which can be evaluated numerically using standard quadrature rules. Furthermore, we establish a simple expression for the difference between $\Phi(\mathbf{r})$ and $\Phi(\mathbf{I}_{-1}\, \mathbf{r})$. Note that a similar representation of $\Psi(\mathbf{r})$ as in (4.14) is used for the proof of the Berman inequality (see above).

Theorem 4.6. *For every* $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$,

$$\Psi(\mathbf{r}) \;=\; \sum_{k=2}^{5} r_k \int_0^1 \frac{\partial \Psi}{\partial r_k}(\mathbf{I}_h\, \mathbf{r})\, \mathrm{d}h\,, \tag{4.14}$$

$$\Phi(\mathbf{r}) \;=\; \Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin r_1\Big)\Big(\frac{1}{4} + \frac{1}{2\pi} \arcsin r_6\Big) + \frac{1}{8\pi} \sum_{k=2}^{5} \arcsin r_k + \frac{\Psi(\mathbf{r})}{4}\,, \tag{4.15}$$

$$\Phi(\mathbf{r}) - \Phi(\mathbf{I}_{-1}\, \mathbf{r}) \;=\; \frac{1}{4\pi} \sum_{k=2}^{5} \arcsin r_k\,. \tag{4.16}$$

Proof. Let $\mathbf{r} \in \mathcal{R}$. For $h \in [0,1]$, define $u(h) := \Psi(\mathbf{I}_h\, \mathbf{r})$. By the chain rule of differentiation, we obtain

$$u'(h) \;=\; \sum_{i=2}^{5} r_i \frac{\partial \Psi}{\partial r_i}(\mathbf{I}_h\, \mathbf{r})\,.$$

Since $u(1) = \Psi(\mathbf{r})$, the Fundamental Theorem of Calculus yields

$$\Psi(\mathbf{r}) \;=\; u(0) \,+\, \int_0^1 u'(h)\, \mathrm{d}h\,.$$

According to Lemma 4.2 (ii), we have $u(0) = 0$ and hence (4.14) follows. Analogously, let $v(h) := \Phi(\mathbf{I}_h\, \mathbf{r})$ for $h \in [0,1]$ and note that

$$\Phi(\mathbf{r}) \;=\; v(0) \,+\, \int_0^1 v'(h)\, \mathrm{d}h\,.$$

Comparing the partial derivatives of Φ and Ψ given in Lemma 4.3 and 4.5, respectively, we obtain

$$\frac{\partial \Phi}{\partial r_k}(\mathbf{r}) \;=\; \frac{1}{8\pi\sqrt{1 - r_k^2}} + \frac{1}{4} \frac{\partial \Psi}{\partial r_k}(\mathbf{r})$$

for $k = 2, 3, 4, 5$. Thus,

$$\begin{aligned}
\int_0^1 v'(h)\, \mathrm{d}h \;&=\; \sum_{k=2}^{5} r_k \int_0^1 \frac{\partial \Phi}{\partial r_k}(\mathbf{I}_h\, \mathbf{r})\, \mathrm{d}h \\
&=\; \frac{1}{8\pi} \sum_{k=2}^{5} r_k \int_0^1 \frac{1}{\sqrt{1 - r_k^2 h^2}}\, \mathrm{d}h \,+\, \frac{1}{4} \sum_{k=2}^{5} r_k \int_0^1 \frac{\partial \Psi}{\partial r_k}(\mathbf{I}_h\, \mathbf{r})\, \mathrm{d}h \\
&=\; \frac{1}{8\pi} \sum_{k=2}^{5} \arcsin r_k + \frac{\Psi(\mathbf{r})}{4}\,,
\end{aligned}$$

where the last equality follows from (4.14) and

$$r_k \int_0^1 \frac{1}{\sqrt{1-r_k^2h^2}}\,\mathrm{d}h \;=\; \int_0^{r_k} \frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t \;=\; \arcsin r_k\,.$$

According to Lemma 4.2 (*ii*), we have

$$v(0) \;=\; \Big(\frac{1}{4}+\frac{1}{2\pi}\arcsin r_1\Big)\Big(\frac{1}{4}+\frac{1}{2\pi}\arcsin r_6\Big),$$

and thus (4.15) follows. Now (4.16) is an immediate consequence of (4.15) and the fact that $x \mapsto \arcsin x$ is an odd function. □

Note that, for $k = 2,3,4,5$, the derivative of $h \mapsto \frac{\partial\Psi}{\partial r_k}(\mathbf{I}_h\,\mathbf{r})$ is continuous on $[0,1]$ and hence bounded. Furthermore, for fixed $\mathbf{r} \in \mathcal{R}$, upper and lower bounds can be given in a closed form, which allows to evaluate the integrals in (4.14) numerically to any desired precision.

4.3 Asymptotic properties

Next, we relate the asymptotics of special sequences in $\mathcal{R}$ to the asymptotics of the corresponding values of Ψ. At the end of this section, we specialize the results to derive asymptotics of zero crossing covariances.

In order to state the main result in Theorem 4.8, we define asymptotic equivalence of vector-valued sequences: Let $(\mathbf{r}(k))_{k\in\mathbb{N}}$ and $(\mathbf{s}(k))_{k\in\mathbb{N}}$ be sequences of vectors in $\mathbb{R}^n$, where $\mathbf{r}(k) = (r_1(k), r_2(k), \ldots, r_n(k))$ and $\mathbf{s}(k) = (s_1(k), s_2(k), \ldots, s_n(k))$ for $k \in \mathbb{N}$. We write

$$\mathbf{r}(k) \;\sim\; \mathbf{s}(k)$$

and say $(\mathbf{r}(k))_{k\in\mathbb{N}}$ *and* $(\mathbf{s}(k))_{k\in\mathbb{N}}$ *are asymptotically equivalent* iff $r_i(k) \sim s_i(k)$ for all $i \in \{1,2,\ldots,n\}$. For the proof of Theorem 4.8, we need the following lemma.

Lemma 4.7. *For every* $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$,

$$\begin{aligned}
\frac{\partial^2\Phi}{\partial^2 r_2}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial^2 r_3}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial^2 r_4}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial^2 r_5}(\mathbf{I}_0\,\mathbf{r}) \;&=\; \frac{r_1 r_6}{4\pi^2\sqrt{(1-r_1^2)(1-r_6^2)}}\,,\\
\frac{\partial^2\Phi}{\partial r_2\partial r_3}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial r_4\partial r_5}(\mathbf{I}_0\,\mathbf{r}) \;&=\; \frac{-r_1}{4\pi^2\sqrt{(1-r_1^2)(1-r_6^2)}}\,,\\
\frac{\partial^2\Phi}{\partial r_2\partial r_4}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial r_3\partial r_5}(\mathbf{I}_0\,\mathbf{r}) \;&=\; \frac{-r_6}{4\pi^2\sqrt{(1-r_1^2)(1-r_6^2)}}\,,\\
\frac{\partial^2\Phi}{\partial r_2\partial r_5}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{\partial^2\Phi}{\partial r_3\partial r_4}(\mathbf{I}_0\,\mathbf{r}) \;&=\; \frac{1}{4\pi^2\sqrt{(1-r_1^2)(1-r_6^2)}}\,.
\end{aligned}$$

Proof. Let $k,l \in \{2,3,4,5\}$. According to Lemma 4.3, there exist unique $i,j \in \{1,2,3,4\}$ such that $\frac{\partial\Phi}{\partial r_k}(\mathbf{r}) = \frac{1}{2\pi}\, f(\mathbf{r})\left(\frac{1}{4} + \frac{1}{2\pi}\, g(\mathbf{r})\right)$, where

$$f(\mathbf{r}) := \frac{1}{\sqrt{1-r_k^2}} \quad \text{and} \quad g(\mathbf{r}) := \arcsin \frac{\sigma_{ij}(\mathbf{r})}{\sqrt{\sigma_{ii}(\mathbf{r})\sigma_{jj}(\mathbf{r})}}\,.$$

Note that $f(\mathbf{I}_0\,\mathbf{r}) = 1$ and $\frac{\partial f}{\partial r_l}(\mathbf{I}_0\,\mathbf{r}) = 0$. Consequently,

$$\frac{\partial^2\Phi}{\partial r_k \partial r_l}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{1}{4\pi^2}\,\frac{\partial g}{\partial r_l}(\mathbf{I}_0\,\mathbf{r})\,.$$

Since $\sigma_{ij}(\mathbf{I}_0\,\mathbf{r}) = 0$ and the first derivative of $x \mapsto \arcsin x$ in 0 is equal to 1, we obtain

$$\frac{\partial^2\Phi}{\partial r_k \partial r_l}(\mathbf{I}_0\,\mathbf{r}) \;=\; \frac{1}{4\pi^2\sqrt{\sigma_{ii}(\mathbf{I}_0\,\mathbf{r})\sigma_{jj}(\mathbf{I}_0\,\mathbf{r})}}\,\frac{\partial\sigma_{ij}}{\partial r_l}(\mathbf{I}_0\,\mathbf{r})$$

With the expressions for σ_{ii}, σ_{jj} and σ_{ij} given in Lemma 4.3, the statement follows. □

Theorem 4.8. *Let $(\mathbf{r}(k))_{k\in\mathbb{N}}$ be a sequence in $\mathcal{R}$. If there exists a function $f : \mathbb{N} \to \mathbb{R}$ with $\lim_{k\to\infty} f(k) = 0$ and a vector $\boldsymbol{\alpha} = (\alpha_1,\, \alpha_2,\, \alpha_3,\, \alpha_4,\, \alpha_5,\, \alpha_6) \in \mathbb{R}^6$ with $|\alpha_1|, |\alpha_6| < 1$ such that $\mathbf{r}(k) \sim (\alpha_1,\, \alpha_2\, f(k),\, \alpha_3\, f(k),\, \alpha_4\, f(k),\, \alpha_5\, f(k),\, \alpha_6)$, then*

$$\Psi(\mathbf{r}(k)) \;\sim\; \frac{(f(k))^2\, q(\boldsymbol{\alpha})}{2\pi^2\sqrt{(1-\alpha_1^2)(1-\alpha_6^2)}} \,+\, O\big((f(k))^4\big)\,,$$

where $q(\boldsymbol{\alpha}) := \alpha_1\alpha_6 \sum_{i=2}^5 \alpha_i^2 - 2\alpha_1(\alpha_2\alpha_3 + \alpha_4\alpha_5) - 2\alpha_6(\alpha_2\alpha_4 + \alpha_3\alpha_5) + 2(\alpha_2\alpha_5 + \alpha_3\alpha_4)$.

Proof. Let $\mathbf{r}(k) = (r_1(k),\, r_2(k),\, r_3(k),\, r_4(k),\, r_5(k),\, r_6(k))$. According to Corollary 4.4, any partial derivative of Φ with respect to r_2, r_3, r_4, r_5 exists on the entire $\mathcal{R}$. Therefore, according to Taylor's Theorem, we find for each $k \in \mathbb{N}$ an $h_1(k) \in [0,1]$ such that

$$\begin{aligned}\Phi(\mathbf{r}(k)) \;=\; \Phi(\mathbf{I}_0\,\mathbf{r}(k)) \,+\, \sum_{i=2}^5 r_i(k)\frac{\partial\Phi}{\partial r_i}(\mathbf{I}_0\,\mathbf{r}(k)) \,+\, \frac{1}{2}\sum_{i,j=2}^5 r_i(k)r_j(k)\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k))\\ +\,\frac{1}{6}\sum_{i,j,l=2}^5 r_i(k)r_j(k)r_l(k)\frac{\partial^3\Phi}{\partial r_i\partial r_j\partial r_l}(\mathbf{I}_0\,\mathbf{r}(k))\\ +\,\frac{1}{24}\sum_{i,j,l,m=2}^5 r_i(k)r_j(k)r_l(k)r_m(k)\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}\big((\mathbf{I}_0 + h_1(k)(\mathbf{I}_1 - \mathbf{I}_0))\,\mathbf{r}(k)\big)\,.\end{aligned}$$

Furthermore, using the fact that $\mathbf{I}_0\,\mathbf{I}_{-1} = \mathbf{I}_0$, we find an $h_2(k) \in [-1,0]$ such that

$$\begin{aligned}\Phi(\mathbf{I}_{-1}\,\mathbf{r}(k)) \;=\; \Phi(\mathbf{I}_0\,\mathbf{r}(k)) \;-\; \sum_{i=2}^{5} r_i(k)\frac{\partial\Phi}{\partial r_i}(\mathbf{I}_0\,\mathbf{r}(k)) \;+\; \frac{1}{2}\sum_{i,j=2}^{5} r_i(k)r_j(k)\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k))\\ -\frac{1}{6}\sum_{i,j,l=2}^{5} r_i(k)r_j(k)r_l(k)\frac{\partial^3\Phi}{\partial r_i\partial r_j\partial r_l}(\mathbf{I}_0\,\mathbf{r}(k))\\ +\frac{1}{24}\sum_{i,j,l,m=2}^{5} r_i(k)r_j(k)r_l(k)r_m(k)\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}\big(\,(\mathbf{I}_0 - h_2(k)(\mathbf{I}_{-1}-\mathbf{I}_0))\,\mathbf{r}(k)\,\big)\,.\end{aligned}$$

Since $\mathbf{I}_0 + h(\mathbf{I}_1 - \mathbf{I}_0) = \mathbf{I}_h$ and $\mathbf{I}_0 + h(\mathbf{I}_{-1} - \mathbf{I}_0) = \mathbf{I}_{-h}$ for all $h \in [0,1]$, we obtain

$$\begin{aligned}2\,\Phi(\mathbf{r}(k)) + 2\,\Phi(\mathbf{I}_{-1}\,\mathbf{r}(k)) \;=\; 4\,\Phi(\mathbf{I}_0\,\mathbf{r}(k)) + 2\sum_{i,j=2}^{5} r_i(k)r_j(k)\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k))\\ +\;\frac{1}{12}\sum_{i,j,l,m=2}^{5} r_i(k)r_j(k)r_l(k)r_m(k)\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_1(k)}\,\mathbf{r}(k))\\ +\;\frac{1}{12}\sum_{i,j,l,m=2}^{5} r_i(k)r_j(k)r_l(k)r_m(k)\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_2(k)}\,\mathbf{r}(k))\,.\end{aligned} \tag{4.17}$$

Note that $r_{i_1}(k)\,r_{i_2}(k)\ldots r_{i_n}(k) \sim \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_n}(f(k))^n$ for all $i_1, i_2, \ldots, i_n \in \{2,3,4,5\}$ with $n \in \mathbb{N}$. Thus, inserting the asymptotically equivalent expressions for $r_2(k)$, $r_3(k)$, $r_4(k)$, $r_5(k)$ into (4.17) yields

$$\begin{aligned}2\,\Phi(\mathbf{r}(k)) + 2\,\Phi(\mathbf{I}_{-1}\,\mathbf{r}(k)) \;\sim\; 4\,\Phi(\mathbf{I}_0\,\mathbf{r}(k)) + 2\,(f(k))^2\sum_{i,j=2}^{5}\alpha_i\alpha_j\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k))\\ +\;\frac{1}{12}\,(f(k))^4\sum_{i,j,l,m=2}^{5}\alpha_i\alpha_j\alpha_l\alpha_m\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_1(k)}\,\mathbf{r}(k))\\ +\;\frac{1}{12}\,(f(k))^4\sum_{i,j,l,m=2}^{5}\alpha_i\alpha_j\alpha_l\alpha_m\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_2(k)}\,\mathbf{r}(k))\,.\end{aligned}$$

According to the definition of Ψ, we obtain

$$\Psi(\mathbf{r}(k)) \;\sim\; 2\,(f(k))^2\sum_{i,j=2}^{5}\alpha_i\alpha_j\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k)) + (f(k))^4\,R(k)\,, \tag{4.18}$$

where

$$R(k) \;=\; \frac{1}{12}\sum_{i,j,l,m=2}^{5} \alpha_i\alpha_j\alpha_l\alpha_m \left(\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_1(k)}\,\mathbf{r}(k)) + \frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\mathbf{I}_{h_2(k)}\,\mathbf{r}(k))\right).$$

Note that $|\alpha_1|, |\alpha_6| < 1$ implies $\mathbf{I}_0\,\boldsymbol{\alpha} = (\alpha_1, 0, 0, 0, 0, \alpha_6) \in \mathcal{R}$. Hence, according to Corollary 4.4, the partial derivatives of second order of Φ exist and are continuous at $\mathbf{I}_0\,\boldsymbol{\alpha}$. Since $\lim_{k\to\infty}\mathbf{I}_0\,\mathbf{r}(k) = \mathbf{I}_0\,\boldsymbol{\alpha}$, we obtain

$$\lim_{k\to\infty}\frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\mathbf{r}(k)) \;=\; \frac{\partial^2\Phi}{\partial r_i\partial r_j}(\mathbf{I}_0\,\boldsymbol{\alpha})$$

for all $i, j \in \{2, 3, 4, 5\}$. Thus, $\mathbf{I}_0\,\mathbf{r}(k)$ in (4.18) can be replaced by $\mathbf{I}_0\,\boldsymbol{\alpha}$. Inserting the expressions from Lemma 4.7 with $\mathbf{r} = \boldsymbol{\alpha}$ into (4.18) yields

$$\Psi(\mathbf{r}(k)) \;\sim\; \frac{2(f(k))^2\,q(\boldsymbol{\alpha})}{4\pi^2\sqrt{(1-\alpha_1^2)(1-\alpha_6^2)}} + (f(k))^4\,R(k)\,,$$

with $q(\boldsymbol{\alpha})$ as given in the statement of the theorem. Now, it only remains to show that $(f(k))^4\,R(k) = O\big((f(k))^4\big)$ or, equivalently, $\sup_{k\in\mathbb{N}} R(k) < \infty$. Since the sequences $(\mathbf{r}(k))_{k\in\mathbb{N}}$ and $(\mathbf{I}_{-1}\,\mathbf{r}(k))_{k\in\mathbb{N}}$ both have the limit $\mathbf{I}_0\,\boldsymbol{\alpha}$, the set

$$\mathcal{S} \;:=\; \{\mathbf{I}_0\,\boldsymbol{\alpha}\} \cup \bigcup_{k\in\mathbb{N}}\{\mathbf{r}(k),\, \mathbf{I}_{-1}\,\mathbf{r}(k)\}$$

is closed in $\mathbb{R}^6$. Furthermore, because $\mathcal{S} \subset \mathcal{R} \subset [-1, 1]^6$, we obtain that $\mathcal{S}$ is compact. Consequently, also the convex hull $\bar{\mathcal{S}}$ of $\mathcal{S}$ (that is, the set containing all convex combination of points in $\mathcal{S}$) is compact. According to Corollary 4.4, the partial derivatives of fourth order of Φ with respect to r_2, r_3, r_4, r_5 are continuous on the entire $\mathcal{R}$. Hence,

$$\sup\,\frac{\partial^4\Phi}{\partial r_i\partial r_j\partial r_l\partial r_m}(\,\bar{\mathcal{S}}\,) \;<\; \infty$$

for all $i, j, l, m \in \{2, 3, 4, 5\}$. Now, note that, for all $k \in \mathbb{N}$ and $h \in [-1, 1]$, we have $\mathbf{I}_h\,\mathbf{r}(k) = \frac{1+h}{2}\mathbf{r}(k) + \frac{1-h}{2}\mathbf{I}_{-1}\,\mathbf{r}(k)$. Thus, $\mathbf{I}_h\,\mathbf{r}(k)$ is a convex combination of $\mathbf{r}(k)$ and $\mathbf{I}_{-1}\,\mathbf{r}(k)$, which shows that $\mathbf{I}_h\,\mathbf{r}(k)$ is an element of $\bar{\mathcal{S}}$. The proof is complete. □

A special case. In the remaining part of this section, we apply Theorem 4.8 to vectors $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6) \in \mathcal{R}$ satisfying $r_1 = r_6$ and $r_2 = r_5$.

For $\mathbf{r} = (r_1, r_2, r_3, r_4) \in (-1, 1)^4$, define $\pi^*(\mathbf{r}) := (r_1, r_2, r_3, r_4, r_2, r_1)$. Furthermore, let

$$\mathcal{R}^* \;:=\; \big\{\mathbf{r} \in (-1, 1)^4 \;\big|\; \pi^*(\mathbf{r}) \in \mathcal{R}\big\}\,.$$

For $\mathbf{r} \in \mathcal{R}^*$, define

$$\Psi^*(\mathbf{r}) \;:=\; \Psi(\pi^*(\mathbf{r}))\,. \tag{4.19}$$

Note that $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \mathcal{R}^*$ if and only if

$$\mathbf{\Sigma}(\pi^*(\mathbf{r})) \;=\; \begin{pmatrix} 1 & r_1 & r_2 & r_3 \\ r_1 & 1 & r_4 & r_2 \\ r_2 & r_4 & 1 & r_1 \\ r_3 & r_2 & r_1 & 1 \end{pmatrix}$$

is strictly positive definite. Furthermore, $\pi^*((1-h)\,\mathbf{r} + h\,\mathbf{s}) = (1-h)\,\pi^*(\mathbf{r}) + h\,\pi^*(\mathbf{s})$ for all $\mathbf{r}, \mathbf{s} \in \mathcal{R}^*$ and $h \in [0,1]$. Since $\mathcal{R}$ is convex, it follows that $\mathcal{R}^*$ is convex.

The reason for introducing $\mathcal{R}^*$ and Ψ^* is that, for all $\boldsymbol{\vartheta} \in \mathbf{\Theta}$ and $k > 1$, the correlation matrix of $(Y_1, Y_2, Y_{k+1}, Y_{k+2})$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ is an element of $\mathbf{\Sigma}(\pi^*(\mathcal{R}^*))$. In particular, according to (4.10),

$$\gamma_{\boldsymbol{\vartheta}}(k) \;=\; \Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(k+1),\, \rho_{\boldsymbol{\vartheta}}(k-1))\,. \tag{4.20}$$

In order to evaluate $\Psi^*(\mathbf{r})$ numerically, we can apply Theorem 4.6 to the evaluation of $\Psi(\pi^*(\mathbf{r}))$. The next corollary is a special case of Theorem 4.8.

Corollary 4.9. *Let $(\mathbf{r}(k))_{k\in\mathbb{N}}$ be a sequence in $\mathcal{R}^*$ and assume $f : \mathbb{N} \to \mathbb{R}$ is a function with $\lim_{k\to\infty} f(k) = 0$.*

(i) If $\mathbf{r}(k) \sim (\alpha_1,\, \alpha_2\, f(k),\, \alpha_3\, f(k),\, \alpha_4\, f(k))$ for some vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ with $|\alpha_1| < 1$, then

$$\Psi^*(\mathbf{r}(k)) \;\sim\; \frac{(f(k))^2\, q(\boldsymbol{\alpha})}{2\pi^2(1-\alpha_1^2)} \,+\, O\big((f(k))^4\big)\,,$$

where $q(\boldsymbol{\alpha}) := \alpha_1^2(2\alpha_2^2 + \alpha_3^2 + \alpha_4^2) - 4\alpha_1\alpha_2(\alpha_3 + \alpha_4) + 2(\alpha_2^2 + \alpha_3\alpha_4)$.

(ii) If $f(k+1) \sim \beta f(k)$ for some $\beta \neq 0$ and there exists an α with $|\alpha| < 1$ such that $\mathbf{r}(k) \sim (\alpha,\, f(k),\, f(k+1),\, f(k-1))$, then

$$\Psi^*(\mathbf{r}(k)) \;\sim\; \frac{(f(k))^2\,(2 - \alpha(\beta + \beta^{-1}))^2}{2\pi^2(1-\alpha^2)} \,+\, O\big((f(k))^4\big)\,.$$

(iii) If the assumptions of (ii) hold with $\beta = 1$, then

$$\Psi^*(\mathbf{r}(k)) \;\sim\; \frac{2\,(f(k))^2\,(1-\alpha)}{\pi^2(1+\alpha)} \,+\, O\big((f(k))^4\big)\,.$$

Proof. (i) follows by Theorem 4.8 and the fact that, under the assumptions, $\pi^*(\mathbf{r}(k))$ is asymptotically equivalent to $(\alpha_1,\, \alpha_2\, f(k),\, \alpha_3\, f(k),\, \alpha_4\, f(k),\, \alpha_2\, f(k), \alpha_1)$.

(ii) is a special case of (i) where $\mathbf{r}(k) \sim (\alpha,\, f(k),\, \beta\, f(k),\, f(k)/\beta)$ and thus

$$\begin{aligned} q(\alpha,\, 1,\, \beta,\, 1/\beta) &= \alpha^2(2+\beta^2+\beta^{-2}) - 4\alpha(\beta+\beta^{-1}) + 4 \\ &= (2-\alpha(\beta+\beta^{-1}))^2\,. \end{aligned}$$

Now, (iii) is obvious. □

4.4 Bounds and approximations

Bounds. Under certain conditions on $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \mathcal{R}^*$, Theorem 4.11 below shows that lower and upper bounds for $\Psi^*(\mathbf{r})$ are obtained by setting r_2, r_3, r_4 equal to r_3 and r_4, respectively. In the context of zero crossing covariances, where $(r_1, r_2, r_3, r_4) = (\rho_{\boldsymbol{\vartheta}}(1), \rho_{\boldsymbol{\vartheta}}(k), \rho_{\boldsymbol{\vartheta}}(k+1), \rho_{\boldsymbol{\vartheta}}(k-1))$, the bounds are obtained by setting the correlations between (Y_1, Y_2) and (Y_{k+1}, Y_{k+2}) equal to the correlations between Y_1 and Y_{k+2} and between Y_2 and Y_{k+1}, respectively.

For the proof of Theorem 4.11, we need the following lemma.

Lemma 4.10. *For every* $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \mathcal{R}^*$,

$$\begin{aligned} \frac{\partial \Psi^*}{\partial r_2}(\mathbf{r}) &= \frac{2}{\pi^2\sqrt{1-r_2^2}} \arcsin \frac{\sigma_{13}(\pi^*(\mathbf{r}))}{\sqrt{\sigma_{11}(\pi^*(\mathbf{r}))\sigma_{22}(\pi^*(\mathbf{r}))}}\,, \\ \frac{\partial \Psi^*}{\partial r_3}(\mathbf{r}) &= \frac{1}{\pi^2\sqrt{1-r_3^2}} \arcsin \frac{\sigma_{23}(\pi^*(\mathbf{r}))}{\sigma_{22}(\pi^*(\mathbf{r}))}\,, \\ \frac{\partial \Psi^*}{\partial r_4}(\mathbf{r}) &= \frac{1}{\pi^2\sqrt{1-r_4^2}} \arcsin \frac{\sigma_{14}(\pi^*(\mathbf{r}))}{\sigma_{11}(\pi^*(\mathbf{r}))} \end{aligned}$$

and

$$\begin{aligned} \sigma_{11}(\pi^*(\mathbf{r})) &= 1 - r_1^2 - r_2^2 - r_4^2 + 2r_1r_2r_4\,, \\ \sigma_{22}(\pi^*(\mathbf{r})) &= 1 - r_1^2 - r_2^2 - r_3^2 + 2r_1r_2r_3\,, \\ \sigma_{13}(\pi^*(\mathbf{r})) &= r_2 - r_1r_3 + r_2r_3r_4 - r_2^3 - r_1r_4 + r_2r_1^2\,, \\ \sigma_{14}(\pi^*(\mathbf{r})) &= r_3 - 2r_1r_2 + r_4r_2^2 - r_3r_4^2 + r_4r_1^2\,, \\ \sigma_{23}(\pi^*(\mathbf{r})) &= r_4 - 2r_1r_2 + r_3r_2^2 - r_4r_3^2 + r_3r_1^2\,. \end{aligned}$$

Proof. The expressions for $\sigma_{ij}(\pi^*(\mathbf{r}))$ are special cases of the expressions given in Lemma 4.3. By the chain rule of differentiation, we obtain

$$\frac{\partial \Psi^*}{\partial r_2}(\mathbf{r}) = \frac{\partial \Psi}{\partial r_2}(\pi^*(\mathbf{r})) + \frac{\partial \Psi}{\partial r_5}(\pi^*(\mathbf{r})) \quad \text{and} \quad \frac{\partial \Psi^*}{\partial r_k}(\mathbf{r}) = \frac{\partial \Psi}{\partial r_k}(\pi^*(\mathbf{r})) \text{ for } k = 3, 4.$$

Since $\sigma_{11}(\pi^*(\mathbf{r})) = \sigma_{44}(\pi^*(\mathbf{r}))$, $\sigma_{22}(\pi^*(\mathbf{r})) = \sigma_{33}(\pi^*(\mathbf{r}))$ and $\sigma_{13}(\pi^*(\mathbf{r})) = \sigma_{24}(\pi^*(\mathbf{r}))$ (see Lemma 4.3), we obtain the result according to Lemma 4.5. □

Theorem 4.11. *Let* $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \mathcal{R}^*$ *with* $r_4, r_2 \geq r_3 \geq 0$. *For* $h \in [0,1]$, *define* $\mathbf{s}_h := (1-h)\cdot\mathbf{r} + h\cdot(r_1, r_3, r_3, r_3)$ *and* $\mathbf{t}_h := (1-h)\cdot\mathbf{r} + h\cdot(r_1, r_4, r_4, r_4)$.

(*i*) *If* $1 + r_1 - 2r_3 > 0$ *and* $\sigma_{13}(\pi^*(\mathbf{s}_h))$, $\sigma_{14}(\pi^*(\mathbf{s}_h)) \geq 0$ *for all* $h \in [0,1]$, *then*

$$\Psi^*(\mathbf{r}) \;\; \geq \;\; \Psi^*(r_1, r_3, r_3, r_3)\,.$$

(*ii*) *If* $1 + r_1 - 2r_4 > 0$ *and* $\sigma_{13}(\pi^*(\mathbf{t}_h))$, $\sigma_{23}(\pi^*(\mathbf{t}_h)) \geq 0$ *for all* $h \in [0,1]$, *then*

$$\Psi^*(r_1, r_4, r_4, r_4) \;\; \geq \;\; \Psi^*(\mathbf{r})\,.$$

Proof. (*i*) Note that the set of eigenvalues of $\mathbf{\Sigma}(\pi^*(r_1, r_3, r_3, r_3))$ is given by $\{1 - r_1,\ 1 + r_1 + 2r_3,\ 1 + r_1 - 2r_3\}$. Under the assumptions, each eigenvalue is strictly larger than 0, which shows that $\mathbf{\Sigma}(\pi^*(r_1, r_3, r_3, r_3))$ is strictly positive definite (see Horn and Johnson [49], p. 398) and hence $(r_1, r_3, r_3, r_3) \in \mathcal{R}^*$. Since $\mathcal{R}^*$ is convex, we have $\mathbf{s}_h \in \mathcal{R}^*$ for all $h \in [0,1]$, and thus $f(h) := \Psi^*(\mathbf{s}_h)$ is well-defined for all $h \in [0,1]$. Because $f(0) = \Psi^*(\mathbf{r})$ and $f(1) = \Psi^*(r_1, r_3, r_3, r_3)$, it is sufficient to show that $h \mapsto f(h)$ is monotonically decreasing on $[0,1]$ or, equivalently,

$$f'(h) \;\; = \;\; (r_3 - r_2)\,\frac{\partial\Psi^*(\mathbf{s}_h)}{\partial r_2} \;+\; (r_3 - r_4)\,\frac{\partial\Psi^*(\mathbf{s}_h)}{\partial r_4} \;\; \leq \;\; 0$$

for all $h \in [0,1]$. Since $r_3 - r_2 \leq 0$ and $r_3 - r_4 \leq 0$, Lemma 4.10 shows that a sufficient condition for this inequality to hold is $\sigma_{13}(\pi^*(\mathbf{s}_h)) \geq 0$ and $\sigma_{14}(\pi^*(\mathbf{s}_h)) \geq 0$ for all $h \in [0,1]$.
(*ii*) Analogously, define $g(h) := \Psi^*(\mathbf{t}_h)$ and note that a sufficient condition for

$$g'(h) \;\; = \;\; (r_4 - r_2)\,\frac{\partial\Psi(\mathbf{t}_h)}{\partial r_2} \;+\; (r_4 - r_3)\,\frac{\partial\Psi(\mathbf{t}_h)}{\partial r_3} \;\; \geq \;\; 0$$

to hold is given by $\sigma_{13}(\pi^*(\mathbf{t}_h)) \geq 0$ and $\sigma_{23}(\pi^*(\mathbf{t}_h)) \geq 0$ for all $h \in [0,1]$. □

Remark 4.12. As the proof of Theorem 4.11 shows, a sufficient condition for strict inequality in (*i*) is given by $r_4 > r_3$ and $\sigma_{14}(\pi^*(\mathbf{s}_h)) > 0$ for some $h \in [0,1]$, or $r_2 > r_3$ and $\sigma_{13}(\pi^*(\mathbf{s}_h)) > 0$ for some $h \in [0,1]$. Similarly, a sufficient condition for strict inequality in (*ii*) is given by $r_4 > r_3$ and $\sigma_{23}(\pi^*(\mathbf{t}_h)) > 0$ for some $h \in [0,1]$, or $r_4 > r_2$ and $\sigma_{13}(\pi^*(\mathbf{t}_h)) > 0$ for some $h \in [0,1]$.

The next lemma gives easily verifiable conditions for the assumptions of Theorem 4.11.

Lemma 4.13. *Let* $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \mathcal{R}^*$ *and* $\mathbf{s}_h$, $\mathbf{t}_h$ *as defined in Theorem 4.11. If* $r_1 \leq 0$ *and* $r_2, r_3, r_4 \geq 0$, *then* $\sigma_{13}(\mathbf{s}_h)$, $\sigma_{14}(\mathbf{s}_h) > 0$ *and* $\sigma_{13}(\mathbf{t}_h)$, $\sigma_{23}(\mathbf{t}_h) > 0$ *for all* $h \in [0,1]$.

Proof. Let $h \in [0,1]$ and $(s_1, s_2, s_3, s_4) = \mathbf{s}_h$. Note that $s_1 \leq 0$ and $s_2,\ s_3,\ s_4 \in [0,1)$. Since $\sigma_{13}(\pi^*(\mathbf{s}_h)) \geq s_2 - s_2^3$ and $\sigma_{14}(\pi^*(\mathbf{s}_h)) \geq s_3 - s_3 s_4^2$ (see Lemma 4.10), we obtain $\sigma_{13}(\mathbf{s}_h),\ \sigma_{14}(\mathbf{s}_h) > 0$. For $(t_1, t_2, t_3, t_4) = \mathbf{t}_h$, we have $t_1 \leq 0$ and $t_2,\ t_3,\ t_4 \in [0,1)$. Because $\sigma_{13}(\pi^*(\mathbf{t}_h)) \geq t_2 - t_2^3$ and $\sigma_{23}(\pi^*(\mathbf{t}_h)) \geq t_4 - t_4 t_3^2$, the statement follows. □

Approximations. Next, we analyze approximations of the lower and upper bounds for $\Psi^*(\mathbf{r})$ established by Theorem 4.11. Let $\mathcal{R}^{**}$ be the set of $\mathbf{r} = (r_1, r_2) \in (-1,1)^2$ such that $\pi^{**}(\mathbf{r}) := (r_1, r_2, r_2, r_2) \in \mathcal{R}^*$ or, equivalently,

$$\mathbf{\Sigma}(\pi^{**}(\mathbf{r})) \;=\; \begin{pmatrix} 1 & r_1 & r_2 & r_2 \\ r_1 & 1 & r_2 & r_2 \\ r_2 & r_2 & 1 & r_1 \\ r_2 & r_2 & r_1 & 1 \end{pmatrix}$$

is strictly positive definite. Note that the set of eigenvalues of $\mathbf{\Sigma}(\pi^{**}(\mathbf{r}))$ is given by $\{1 - r_1,\ 1 + r_1 + 2r_2,\ 1 + r_1 - 2r_2\}$. Since a matrix is strictly positive definite if and only if all its eigenvalues are greater than 0 (see Horn and Johnson [49], p. 398), $\mathcal{R}^{**}$ is given by

$$\mathcal{R}^{**} \;=\; \left\{ (r_1, r_2) \in (-1,1)^2 \;\middle|\; 2\,|r_2| < 1 + r_1 \right\}. \tag{4.21}$$

For $\mathbf{r} \in \mathcal{R}^{**}$, define

$$\Phi^{**}(\mathbf{r}) \;:=\; \Phi(\pi^{**}(\mathbf{r})) \quad \text{and} \quad \Psi^{**}(\mathbf{r}) \;:=\; \Psi(\pi^{**}(\mathbf{r}))\,.$$

If $(\mathbf{r}(k))_{k \in \mathbb{N}}$ is a sequence in $\mathcal{R}^{**}$ with $\mathbf{r}(k) \sim (\alpha_1,\ \alpha_2\, f(k))$ for a function $f : \mathbb{N} \to \mathbb{R}$ with $\lim_{k\to\infty} f(k) = 0$ and $\alpha_1, \alpha_2 \in \mathbb{R}$ with $|\alpha_1| < 1$, then the asymptotics of $\Psi^{**}(\mathbf{r}(k))$ is obtained by Corollary 4.9 (*iii*), namely,

$$\Psi^{**}(\mathbf{r}(k)) \;\sim\; \frac{2\,(f(k))^2\,(1-\alpha)}{\pi^2(1+\alpha)} \,+\, O\big((f(k))^4\big)\,. \tag{4.22}$$

Let $\mathbf{r} \in \mathcal{R}^{**}$. In order to evaluate $\Psi^{**}(\mathbf{r})$ numerically, we can use the integral representation for $\Psi(\pi^{**}(\mathbf{r}))$ given in (4.14). Because of the special structure of $\pi^{**}(\mathbf{r})$, this integral representation has a very simple form: Note that $\sigma_{ii}(\pi^{**}(\mathbf{r})) = (1 - r_1)(1 + r_1 - 2r_2^2)$ for $i = 1, 2, 3, 4$, and $\sigma_{13}(\pi^{**}(\mathbf{r})) = \sigma_{14}(\pi^{**}(\mathbf{r})) = \sigma_{23}(\pi^{**}(\mathbf{r})) = \sigma_{24}(\pi^{**}(\mathbf{r})) = r_2(1 - r_1)^2$ (see Lemma 4.3). Therefore, according to Lemma 4.5,

$$\frac{\partial \Psi}{\partial r_k}(\pi^{**}(\mathbf{r})) \;=\; \frac{1}{\sqrt{1 - r_2^2}} \arcsin \frac{r_2(1 - r_1)}{1 + r_1 - 2r_2^2}$$

for $k = 2, 3, 4, 5$. Inserting these expressions into (4.14) yields

$$\begin{aligned} \Psi^{**}(\mathbf{r}) \;&=\; \frac{4\,r_2}{\pi^2} \int_0^1 \frac{1}{\pi^2 \sqrt{1 - r_2^2}} \arcsin \frac{r_2(1 - r_1)\,h}{1 + r_1 - 2r_2^2 h^2}\; \mathrm{d}h \\ &=\; \frac{4}{\pi^2} \int_0^{r_2} \frac{1}{\sqrt{1 - t^2}} \arcsin \frac{(1 - r_1)\,t}{1 + r_1 - 2t^2}\; \mathrm{d}t\,. \end{aligned} \tag{4.23}$$

By analyzing the expression on the right hand side of (4.23), we immediately obtain the following statement.

Corollary 4.14.

(*i*) *For every* $\mathbf{r} = (r_1, r_2) \in \mathcal{R}^{**}$,

$$\Psi^{**}(\mathbf{r}) \geq 0 \quad \textit{and} \quad \Psi^{**}(r_1, r_2) = \Psi^{**}(r_1, -r_2)\,.$$

(*ii*) *If* $(r_1, r_2), (r_1', r_2') \in \mathcal{R}^{**}$ *satisfy* $r_1' \leq r_1$ *and* $r_2' \leq r_2 < 0$, *then*

$$\Psi^{**}(r_1,\, r_2) \;\leq\; \Psi^{**}(r_1',\, r_2')\,.$$

As the following theorem shows, $\Psi^{**}(\mathbf{r})$ can be approximated monotonically from below by successively adding further terms of the Taylor expansion of $\Psi^{**}(\mathbf{r})$ at $(r_1, 0)$. For the proof, we need the partial derivative of Φ^{**} with respect to r_2. By the expressions for the partial derivatives of Φ given in Lemma 4.3, we obtain

$$\frac{\partial \Phi}{\partial r_k}(\pi^{**}(\mathbf{r})) \;=\; \frac{1}{2\pi\sqrt{1-r_2^2}} \left(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{r_2(1-r_1)}{1+r_1-2r_2^2} \right)$$

for $k = 2, 3, 4, 5$. Hence, according to the chain rule of differentation,

$$\begin{aligned} \frac{\partial \Phi^{**}}{\partial r_2}(\mathbf{r}) &= \frac{\partial \Phi}{\partial r_2}(\pi^{**}(\mathbf{r})) + \frac{\partial \Phi}{\partial r_3}(\pi^{**}(\mathbf{r})) + \frac{\partial \Phi}{\partial r_4}(\pi^{**}(\mathbf{r})) + \frac{\partial \Phi}{\partial r_5}(\pi^{**}(\mathbf{r})) \\ &= \frac{2}{\pi\sqrt{1-r_2^2}} \left(\frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{r_2(1-r_1)}{1+r_1-2r_2^2} \right)\,. \end{aligned} \tag{4.24}$$

Theorem 4.15. *For every* $\mathbf{r} = (r_1, r_2) \in \mathcal{R}^{**}$,

$$\frac{\partial^l \Phi^{**}}{\partial^l r_2}(r_1, 0) \;\geq\; 0 \quad \textit{for all } l \in \mathbb{N}_0\,, \tag{4.25}$$

$$\Psi^{**}(\mathbf{r}) \;=\; 4 \sum_{l=1}^{\infty} \frac{r_2^{2l}}{(2l)!} \frac{\partial^{2l} \Phi^{**}}{\partial^{2l} r_2}(r_1, 0)\,. \tag{4.26}$$

Proof. Let $\mathbf{r} = (r_1, r_2) \in \mathcal{R}^{**}$. Since the orthant probability $\Phi^{**}(r_1, 0)$ is non-negative, (4.25) is true for $l = 0$. Now, let $f(x) := \frac{1}{2\pi} \arcsin x$ for $x \in (-1, 1)$. Furthermore, define $g_1(x) := x(1 - r_1)$, $g_2(x) := \frac{1}{1+r_1-2x^2}$, $g(x) := g_1(x)g_2(x)$ and $h(x) := f(g(x))$ for $x \in (-\frac{1+r_1}{2}, \frac{1+r_1}{2})$. According to (4.21), we have $|r_2| < \frac{1+r_1}{2}$, so (4.24) yields

$$\frac{\partial \Phi^{**}}{\partial r_2}(\mathbf{r}) \;=\; f'(r_2) + 4f'(r_2)h(r_2)\,.$$

By Leibniz's rule, we obtain

$$\frac{\partial^l \Phi^{**}}{\partial^l r_2}(r_1, 0) = f^{(l)}(0) + 4 \sum_{k=1}^{l} \binom{l-1}{k-1} f^{(k)}(0)\, h^{(l-k)}(0)$$

for $l \in \mathbb{N}$. As a well-known fact, $x \mapsto \arcsin x$ has the power series expansion

$$\arcsin x = \sum_{n=0}^{\infty} \frac{3 \cdot 5 \cdot \ldots \cdot (2n-1)}{2 \cdot 4 \cdot \ldots \cdot (2n) \cdot (2n+1)}\, x^{2n+1}$$

for $x \in (-1, 1)$ (see Bronshtein and Semendyayev [20], p. 24), so $f^{(l)}(0) \geq 0$ for all $l \in \mathbb{N}$. Therefore, in order to prove (4.25), it is sufficient to show that $h^{(l)}(0) \geq 0$ for all $l \in \mathbb{N}$.

Let $g_2(x) = f_2(f_1(x))$ with $f_1(x) := 1 + r_1 - 2x^2$ and $f_2(x) := \frac{1}{x}$. For each $l \in \mathbb{N}$, we can write $g_2^{(l)}(0) = (f_2 \circ f_1)^{(l)}(0)$ as the sum of terms $f_2^{(k)}(f_1(0)) \cdot f_1^{(i_1)}(0) \cdot f_1^{(i_2)}(0) \cdot \ldots \cdot f_1^{(i_k)}(0)$ with $k, i_1, i_2, \ldots, i_k \in \mathbb{N}$ which satisfy $i_1 + i_2 + \ldots + i_k = l$. Because $f_1^{(l)}(0) \neq 0$ only if $l \in \{0, 2\}$, each term can only be non-zero if $i_1 = i_2 = \ldots = i_k = 2$. Since $f_2^{(k)}(f_1(0)) = (-1)^k k!(1 + r_1)^{-(k+1)}$ and $f_1^{(2)}(0) = -4$, we obtain $g_2^{(l)}(0) \geq 0$. Now, because $g_1^{(k)}(0) \neq 0$ only if $k = 1$, applying Leibniz's rule yields $g^{(l)}(0) = l \cdot g_1^{(1)}(0) \cdot g_2^{(l-1)}(0) = l \cdot (1 - r_1) \cdot g_2^{(l-1)}(0)$ and hence $g^{(l)}(0) \geq 0$ for all $l \in \mathbb{N}$. For each $l \in \mathbb{N}$, we can write $h^{(l)}(0) = (f \circ g)^{(l)}(0)$ as the sum of products consisting of factors of the form $f^{(k)}(g(0)) = f^{(k)}(0)$ and $g^{(m)}(0)$ with $k, m \in \mathbb{N}$. Thus, $h^{(l)}(0) \geq 0$.

In order to prove (4.26), note that f and g have power series expansions at 0 with the radius of convergence 1 and $\frac{1+r_1}{2}$, respectively, and the image of the interval $(-\frac{1+r_1}{2}, \frac{1+r_1}{2})$ under g is a subset of $(-1, 1)$. By elementary properties of power series, the mapping $\frac{\partial \Phi^{**}}{\partial r_2}(r_1, \cdot) = f'(\cdot) + 4f'(\cdot)f(g(\cdot))$ has a power series expansion at 0 with the radius of convergence $\frac{1+r_1}{2}$, and the same holds for $\Phi^{**}(r_1, \cdot)$. Since $r_2 \in (-\frac{1+r_1}{2}, \frac{1+r_1}{2})$, we obtain

$$\begin{aligned}
\Psi^{**}(\mathbf{r}) &= 2\,\Phi^{**}(\mathbf{r}) + 2\,\Phi^{**}(r_1, -r_2) - 4\,\Phi^{**}(r_1, 0) \\
&= 2 \sum_{l=0}^{\infty} \frac{r_2^l}{l!} \frac{\partial^l \Phi^{**}}{\partial^l r_2}(r_1, 0) + 2 \sum_{l=0}^{\infty} \frac{(-r_2)^l}{l!} \frac{\partial^l \Phi^{**}}{\partial^l r_2}(r_1, 0) - 4\,\Phi^{**}(r_1, 0) \\
&= 4 \sum_{l=1}^{\infty} \frac{r_2^{2l}}{(2l)!} \frac{\partial^{2l} \Phi^{**}}{\partial^{2l} r_2}(r_1, 0)\,.
\end{aligned}$$

The proof is complete. □

When the order is small, closed-form expressions for the partial derivatives of Φ^{**} at $(r_1, 0)$ can be obtained manually. For larger orders, we can use computer algebra systems. For instance, the derivatives of second, fourth and sixth order are given by

$$\frac{\partial^2 \Phi^{**}}{\partial^2 r_2}(r_1, 0) = \frac{1 - r_1}{\pi^2 (1 + r_1)}, \tag{4.27}$$

$$\frac{\partial^4 \Phi^{**}}{\partial^4 r_2}(r_1, 0) = \frac{4(1 - r_1)(2 + r_1)^2}{\pi^2 (1 + r_1)^3}, \tag{4.28}$$

$$\frac{\partial^6 \Phi^{**}}{\partial^6 r_2}(r_1, 0) = \frac{16(1 - r_1)(7 + 6r_1 + 2r_1^2)^2}{\pi^2 (1 + r_1)^5}. \tag{4.29}$$

4.5 The variance of the empirical zero crossing rate

In this section, we apply the previous results to the analysis of the variance of empirical zero crossing rates. Recall formula (4.3),

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) = \frac{1}{n^2}\Big(n\, \gamma_{\boldsymbol{\vartheta}}(0) + 2 \sum_{k=1}^{n-1} (n - k)\, \gamma_{\boldsymbol{\vartheta}}(k) \Big).$$

In order to evaluate $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ numerically, we can use formulas (4.5) and (4.6) for the computation of $\gamma_{\boldsymbol{\vartheta}}(0)$ and $\gamma_{\boldsymbol{\vartheta}}(1)$. For $k > 1$, formula (4.20) yields

$$\gamma_{\boldsymbol{\vartheta}}(k) = \Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(k+1),\, \rho_{\boldsymbol{\vartheta}}(k-1)),$$

and the right hand side can be evaluated numerically using the integral representation of Ψ and Ψ^*, respectively, given in (4.14).

When n is large, an "exact" numerical evaluation of $\gamma_{\boldsymbol{\vartheta}}(k)$ for every $k = 0, 1, \ldots, n-1$ is time-consuming. A quick way for getting approximate values of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ is to use approximations of $\gamma_{\boldsymbol{\vartheta}}(k)$ in terms of the function Ψ^{**}. Due to a smaller number of numerical integrations to be performed, evaluating Ψ^{**} takes less than half of the time required for the evaluation of Ψ^*. In Corollaries 4.16 and 4.17 below we give sufficient conditions on $k \mapsto \rho_{\boldsymbol{\vartheta}}(k)$ such that

$$\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1)) < \gamma_{\boldsymbol{\vartheta}}(k) < \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k-1)).$$

In this case, approximating $\gamma_{\boldsymbol{\vartheta}}(k)$ by $\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1))$ and $\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k-1))$, respectively, yields lower and upper bounds for $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$. A further speed-up can be achieved by using the finite-order approximations of Ψ^{**} provided by Theorem 4.15. For instance, when the autocorrelations of $\mathbf{Y}$ are not too large, one can use the approximation

$$\gamma_{\boldsymbol{\vartheta}}(k) \approx \frac{2(1 - \rho_{\boldsymbol{\vartheta}}(1))}{\pi^2 (1 + \rho_{\boldsymbol{\vartheta}}(1))}\, (\rho_{\boldsymbol{\vartheta}}(k))^2$$

which only includes the partial derivative of Φ^{**} of second order (see (4.27)). Note that the fraction does not depend on k and thus only needs to be computed once.

An alternative method for computing approximate values of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ is to use the exact values of $\gamma_{\boldsymbol{\vartheta}}(k)$ for $k = 2, 3, \ldots$ until the relative error of the approximations falls below a given threshold $\epsilon > 0$. After that, we use the approximations of $\gamma_{\boldsymbol{\vartheta}}(k)$. If the relative error does not get larger than ϵ anymore, then also the relative error of the resulting approximation of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ is not larger than ϵ. For the calculations behind Figures 5.1-5.3 in Section 5.2 and for the computation of the confidence intervals in Section 5.4, we have used this method with the threshold $\epsilon = 0.001$.

Corollary 4.16. *Let $k_0 \in \mathbb{N}$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. We have $\gamma_{\boldsymbol{\vartheta}}(k) > \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1))$ for all $k > k_0$ if one of the following conditions is satisfied:*

(i) $\rho_{\boldsymbol{\vartheta}}(1) < 0$, $1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(k_0+2) > 0$ and $\rho_{\boldsymbol{\vartheta}}(k) < \rho_{\boldsymbol{\vartheta}}(k+1) < 0$ for all $k \geq k_0$.

(ii) There exists some $\boldsymbol{a} \in (-1,1) \setminus \{0\}$ with $\rho_{\boldsymbol{\vartheta}}(1) = \boldsymbol{a}$ and $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ for all $k \geq k_0$.

Proof. (i) Let $k > k_0$. According to formula (4.20) and Lemma 4.2 (i), we have

$$\begin{aligned} \gamma_{\boldsymbol{\vartheta}}(k) &= \Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k),\, \rho_{\boldsymbol{\vartheta}}(k+1),\, \rho_{\boldsymbol{\vartheta}}(k-1)) \\ &= \Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, -\rho_{\boldsymbol{\vartheta}}(k),\, -\rho_{\boldsymbol{\vartheta}}(k+1),\, -\rho_{\boldsymbol{\vartheta}}(k-1))\,. \end{aligned}$$

Furthermore, $-\rho_{\boldsymbol{\vartheta}}(k-1), -\rho_{\boldsymbol{\vartheta}}(k) > -\rho_{\boldsymbol{\vartheta}}(k+1) > 0$ and

$$1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(k+1) \;\geq\; 1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(k_0+2) \;>\; 0\,.$$

Therefore, according to Lemma 4.13, Theorem 4.11 (i) and Remark 4.12,

$$\Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, -\rho_{\boldsymbol{\vartheta}}(k),\, -\rho_{\boldsymbol{\vartheta}}(k+1),\, -\rho_{\boldsymbol{\vartheta}}(k-1)) \;>\; \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, -\rho_{\boldsymbol{\vartheta}}(k+1))\,.$$

(ii) Let $k > k_0$. According to Lemma 4.2 (i), we have

$$\gamma_{\boldsymbol{\vartheta}}(k) \;=\; \Psi^*(\boldsymbol{a},\, \boldsymbol{a}^k,\, \boldsymbol{a}^{k+1},\, \boldsymbol{a}^{k-1}) \;=\; \Psi^*(-\boldsymbol{a},\, -\boldsymbol{a}^k,\, \boldsymbol{a}^{k+1},\, \boldsymbol{a}^{k-1}) \;=\; \Psi^*(-\boldsymbol{a},\, \boldsymbol{a}^k,\, -\boldsymbol{a}^{k+1},\, -\boldsymbol{a}^{k-1})\,.$$

Thus, without loss of generality, we may assume that $\boldsymbol{a} > 0$. For $h \in [0,1]$, let

$$\mathbf{s}_h \;:=\; \big(\,\boldsymbol{a},\; (1-h)\cdot\boldsymbol{a}^k + h\cdot\boldsymbol{a}^{k+1},\; \boldsymbol{a}^{k+1},\; (1-h)\cdot\boldsymbol{a}^{k-1} + h\cdot\boldsymbol{a}^{k+1}\,\big)\,.$$

By elementary but tedious computations, we obtain

$$\begin{aligned}\sigma_{13}(\pi^*(\mathbf{s}_h)) &= (1-\boldsymbol{a})^2\,\boldsymbol{a}^k\,h\left(\boldsymbol{a}+\boldsymbol{a}^{2k}(1-h)(1-h(1-\boldsymbol{a}))\right),\\ \sigma_{14}(\pi^*(\mathbf{s}_h)) &= (1-\boldsymbol{a})^2\,\boldsymbol{a}^{k-1}\,h\left(\boldsymbol{a}^2-\boldsymbol{a}^{2k}(1-h)(1-h(1-\boldsymbol{a}^2))\right),\end{aligned}$$

so it is easy to see that $\sigma_{13}(\pi^*(\mathbf{s}_h)) \geq 0$ and $\sigma_{14}(\pi^*(\mathbf{s}_h)) \geq 0$ for all $h \in [0,1]$, with strict inequality for $h > 0$. Furthermore, $\boldsymbol{a}^{k-1}, \boldsymbol{a}^k > \boldsymbol{a}^{k+1} > 0$ and $1+\boldsymbol{a}-2\boldsymbol{a}^{k+1} > 0$. Therefore, according to Theorem 4.11 (i) and Remark 4.12,

$$\Psi^*(\boldsymbol{a},\,\boldsymbol{a}^k,\,\boldsymbol{a}^{k+1},\,\boldsymbol{a}^{k-1}) \;>\; \Psi^{**}(\boldsymbol{a},\,\boldsymbol{a}^{k+1}).$$

The proof is complete. □

Corollary 4.17. *Let $k_0 \in \mathbb{N}$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. We have $\gamma_{\boldsymbol{\vartheta}}(k) < \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\,\rho_{\boldsymbol{\vartheta}}(k-1))$ for all $k > k_0$ if one of the following conditions is satisfied:*

(i) *$\rho_{\boldsymbol{\vartheta}}(1) < 0$, $1+\rho_{\boldsymbol{\vartheta}}(1)+2\,\rho_{\boldsymbol{\vartheta}}(k_0) > 0$ and $\rho_{\boldsymbol{\vartheta}}(k) < \rho_{\boldsymbol{\vartheta}}(k+1) < 0$ for all $k \geq k_0$.*

(ii) *There exists some $\boldsymbol{a} \in (-1,1)\setminus\{0\}$ with $\rho_{\boldsymbol{\vartheta}}(1) = \boldsymbol{a}$ and $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ for all $k \geq k_0$.*

Proof. (i) The proof is similar to the proof of Corollary 4.16 (i).

(ii) By the same argument as in the proof of Corollary 4.16 (ii), we may assume that $\boldsymbol{a} > 0$. For $h \in [0,1]$, define

$$\mathbf{t}_h \;:=\; \left(\boldsymbol{a},\;(1-h)\cdot\boldsymbol{a}^k + h\cdot\boldsymbol{a}^{k-1},\;(1-h)\cdot\boldsymbol{a}^{k+1} + h\cdot\boldsymbol{a}^{k-1},\;\boldsymbol{a}^{k-1}\right).$$

By elementary but tedious computations, we obtain

$$\begin{aligned}\sigma_{13}(\pi^*(\mathbf{t}_h)) &= (1-\boldsymbol{a})^2\,\boldsymbol{a}^{k-3}\,h\left(\boldsymbol{a}^2+\boldsymbol{a}^{2k}(1-h)(\boldsymbol{a}(1-h)+h)\right),\\ \sigma_{23}(\pi^*(\mathbf{t}_h)) &= (1-\boldsymbol{a})^2\,\boldsymbol{a}^{k-1}\left(1+(1-h)(2\boldsymbol{a}-h(1-h)\boldsymbol{a}^{2k})\right)\\ &+ (1-\boldsymbol{a})^2\,\boldsymbol{a}^{k-3}\,(1-h)\left(\boldsymbol{a}^4-h^2\boldsymbol{a}^{2k}\right),\end{aligned}$$

which shows that $\sigma_{13}(\pi^*(\mathbf{t}_h)), \sigma_{23}(\pi^*(\mathbf{t}_h)) \geq 0$ for all $h \in [0,1]$, with strict inequality for $h > 0$. Clearly, $\boldsymbol{a}^{k-1}, \boldsymbol{a}^k > \boldsymbol{a}^{k+1} > 0$ and $1+\boldsymbol{a}-2\boldsymbol{a}^{k-1} > 0$. Thus, the statement follows from Theorem 4.11 (ii) and Remark 4.12. □

Note that numerical experiments suggest that the conclusions of Corollaries 4.16 and 4.17 are valid under more general conditions. However, we do not have a rigorous proof at this time. The difficulty is to verify the assumptions of Theorem 4.11. As the proofs of Corollaries 4.16 (ii) and 4.17 (ii) show, this can be hard even for a very simple structure of $k \mapsto \rho_{\boldsymbol{\vartheta}}(k)$.

Asymptotic properties. According to Corollary 4.9, we obtain the following statement on the asymptotics of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$.

Theorem 4.18. *Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Suppose there exists a function $f : \mathbb{N} \to \mathbb{R}$ such that $\rho_{\boldsymbol{\vartheta}}(k) \sim f(k)$.*

(i) If $|f(k)| = o(k^{-\beta})$ with $\beta > \frac{1}{2}$, then $\sigma^2_{\boldsymbol{\vartheta}} := \gamma_{\boldsymbol{\vartheta}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{\vartheta}}(k) < \infty$ and

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;\sim\; \sigma^2_{\boldsymbol{\vartheta}}\, n^{-1} \,.$$

(ii) If $f(k) = \alpha k^{-\frac{1}{2}}$ for some $\alpha \in (-1,1) \setminus \{0\}$, then

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;\sim\; \frac{4\,\alpha^2(1-\rho_{\boldsymbol{\vartheta}}(1))}{\pi^2(1+\rho_{\boldsymbol{\vartheta}}(1))}\,\frac{\ln n}{n} \,.$$

(iii) If $f(k) = \alpha k^{-\beta}$ for some $\alpha \in (-1,1) \setminus \{0\}$ and $\beta \in (0, \frac{1}{2})$, then

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;\sim\; \frac{4\,\alpha^2(1-\rho_{\boldsymbol{\vartheta}}(1))}{\pi^2(1+\rho_{\boldsymbol{\vartheta}}(1))(1-2\beta)}\, n^{-2\beta} \,.$$

Proof. (i) According to Corollary 4.9 (i), we have $\gamma_{\boldsymbol{\vartheta}}(k) = O((f(k))^2)$, which shows that $\sum_{k=1}^{\infty} |\gamma_{\boldsymbol{\vartheta}}(k)| < \infty$. By the Dominated Convergence Theorem, we obtain

$$\begin{aligned}\sum_{k=1}^{n-1} \frac{n-k}{n}\,\gamma_{\boldsymbol{\vartheta}}(k) \;&\sim\; \lim_{n\to\infty} \sum_{k=1}^{\infty} \max\Big\{\frac{n-k}{n},\, 0\Big\}\,\gamma_{\boldsymbol{\vartheta}}(k) \\ &=\; \sum_{k=1}^{\infty} \lim_{n\to\infty} \max\Big\{\frac{n-k}{n},\, 0\Big\}\,\gamma_{\boldsymbol{\vartheta}}(k) \;=\; \sum_{k=1}^{\infty} \gamma_{\boldsymbol{\vartheta}}(k)\,.\end{aligned}$$

Now, with formula (4.3), the result follows.

(ii) Note that $f(k) \sim f(k+1)$ and thus, according to Corollary 4.9 (iii),

$$\gamma_{\boldsymbol{\vartheta}}(k) \;\sim\; \frac{2\,\alpha^2\,(1-\rho_{\boldsymbol{\vartheta}}(1))}{\pi^2(1+\rho_{\boldsymbol{\vartheta}}(1))}\,k^{-1}\,.$$

Using the fact that $\sum_{k=1}^{n-1} k^{-1} \sim \ln n$, we obtain

$$\sum_{k=1}^{n-1} \gamma_{\boldsymbol{\vartheta}}(k) \;\sim\; \frac{2\,\alpha^2\,(1-\rho_{\boldsymbol{\vartheta}}(1))}{\pi^2(1+\rho_{\boldsymbol{\vartheta}}(1))}\,\ln n\,. \tag{4.30}$$

Furthermore,

$$\begin{aligned}\sum_{k=1}^{n-1}\gamma_{\boldsymbol{\vartheta}}(k)-\sum_{k=1}^{n-1}\frac{n-k}{n}\,\gamma_{\boldsymbol{\vartheta}}(k) &= \sum_{k=1}^{n-1}\frac{k}{n}\,\gamma_{\boldsymbol{\vartheta}}(k)\\ &\sim \frac{1}{n}\sum_{k=1}^{n-1}\frac{2\,\alpha^2\,(1-\rho_{\boldsymbol{\vartheta}}(1))}{\pi^2(1+\rho_{\boldsymbol{\vartheta}}(1))} \;=\; o(\ln n)\end{aligned}$$

which shows that

$$\sum_{k=1}^{n-1}\gamma_{\boldsymbol{\vartheta}}(k) \;\sim\; \sum_{k=1}^{n-1}\frac{n-k}{n}\,\gamma_{\boldsymbol{\vartheta}}(k)\,.$$

According to formula (4.3), we obtain

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;\sim\; \frac{2}{n}\sum_{k=1}^{n-1}\gamma_{\boldsymbol{\vartheta}}(k)\,,$$

and together with (4.30) the statement follows.

(*iii*) The proof is similar to (*ii*), using the fact that $\sum_{k=1}^{n-1}k^{-2\beta}\sim\frac{1}{1-2\beta}\,n^{1-2\beta}$. □

In the following section, we investigate further properties of the quantity $\sigma^2_{\boldsymbol{\vartheta}}$ in (*i*). As shown by numerical experiments, the expressions in (*ii*) and (*iii*) are "good" approximations of $\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ only if n is extremely large.

4.6 Asymptotic variance

Consider the subset $\boldsymbol{\Theta}_0$ of $\boldsymbol{\Theta}$ given by

$$\boldsymbol{\Theta}_0 \;:=\; \Big\{\,\boldsymbol{\vartheta}\in\boldsymbol{\Theta}\;\Big|\;\text{There exist a function } f:\mathbb{N}\to\mathbb{R} \text{ and some } \beta>\tfrac{1}{2} \text{ with } \rho_{\boldsymbol{\vartheta}}(k)\sim f(k) \text{ and } |f(k)|=o(k^{-\beta})\,\Big\}\,.$$

Let $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_0$. By Corollary 4.9 (*i*) we obtain that $|\gamma_{\boldsymbol{\vartheta}}(k)|=o(k^{-1})$. Therefore,

$$\sigma^2_{\boldsymbol{\vartheta}} \;:=\; \gamma_{\boldsymbol{\vartheta}}(0)+2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{\vartheta}}(k)$$

is well-defined and finite. In the remaining part of this chapter, we establish further properties of $\sigma^2_{\boldsymbol{\vartheta}}$. Theorem 4.22 gives sufficient conditions on $k\mapsto\rho_{\boldsymbol{\vartheta}}(k)$ such that $\sigma^2_{\boldsymbol{\vartheta}}$ is strictly positive. In the case where $\boldsymbol{\Theta}_0$ is a subset of $\mathbb{R}$ and I is a compact subset of $\boldsymbol{\Theta}_0$, Theorem 4.24 gives sufficient conditions on $k\mapsto\rho_{\boldsymbol{\vartheta}}(k)$ such that $\boldsymbol{\vartheta}\mapsto n\cdot\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ converges to $\boldsymbol{\vartheta}\mapsto\sigma^2_{\boldsymbol{\vartheta}}$ uniformly on I, and $\boldsymbol{\vartheta}\mapsto\sigma^2_{\boldsymbol{\vartheta}}$ is continuous on I. In Section 5.3, we will apply these results to derive confidence intervals for $\boldsymbol{\vartheta}$. In order to prove Theorem 4.22, we first show the following three lemmas.

Lemma 4.19. *If* $\mathbf{R} \in \mathbb{R}^{3\times 3}$ *is a symmetric Toeplitz matrix with all entries on the main diagonal equal to* 1*, that is,*

$$\mathbf{R} \;=\; \begin{pmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{pmatrix}$$

for some $r_1, r_2 \in \mathbb{R}$*, then* $\frac{1}{2}\big(2 + r_2 - \sqrt{8r_1^2 + r_2^2}\big)$ *is an eigenvalue of* $\mathbf{R}$*.*

Proof. The result is obtained by simple algebraic calculation. □

Lemma 4.20. *For* $(x, y) \in [-1, 1]^2$*, let*

$$f(x,y) \;:=\; \frac{1}{4} - \frac{3}{\pi^2}(\arcsin x)^2 + \frac{1}{\pi}\arcsin y\,.$$

If $x \in [-1, 0]$ *and* $\max\{x, 2x^2 - 1\} \le y \le 0$*, then* $f(x, y) \ge 0$*.*

Proof. Clearly, $y \mapsto f(x, y)$ is increasing on $[\max\{x, 2x^2 - 1\}, 0]$ for every $x \in [-1, 0]$, so we only need to show $f(x, \max\{x, 2x^2 - 1\}) \ge 0$ for $x \in [-1, 0]$, or, equivalently, $f_1(x) := f(x, 2x^2 - 1) \ge 0$ for all $x \in [-1, -\frac{1}{2}]$ and $f_2(x) := f(x, x) \ge 0$ for all $x \in [-\frac{1}{2}, 0]$. Clearly, f_1 and f_2 are continuous on $[-1, -\frac{1}{2}]$ and $[-\frac{1}{2}, 0]$, respectively, and $f_1(-1) = f_1(\frac{1}{2}) = f_2(\frac{1}{2}) = 0$. By elementary calculus we obtain that f_1 has a unique extremal point in $(\sin(-\pi/3), \frac{1}{12})$ and f_2 is increasing on $[-\frac{1}{2}, 0]$, which yields the statement. □

Lemma 4.21. *For every* $x \in [-1, 1]$*,*

$$g(x) \;:=\; \frac{1}{2\pi}\arcsin x^2 - \frac{1}{\pi^2}(\arcsin x)^2 \;\ge\; 0\,.$$

Proof. Since $g(x) = g(-x)$, we may assume $x \in [-1, 0]$. Clearly, $g(-1) = g(0) = 0$. Thus, it suffices to show that there exists an $x_0 \in [-1, 0]$ with $g'(x) \ge 0$ for all $x \in (-1, x_0)$ and $g'(x) \le 0$ for all $x \in (x_0, 0)$. Note that, for $x \in (-1, 0)$,

$$g'(x) \;=\; \frac{2\,h(x)}{\pi^2\sqrt{1 - x^4}}$$

where $h(x) := x\pi/2 - \sqrt{1 + x^2}\arcsin x$. By elementary calculus we obtain

$$h''(x) \;=\; \frac{-3x - 2x^3 + x^5 - (1 - x^2)^{\frac{3}{2}}\arcsin x}{(1 - x^4)^{\frac{3}{2}}} \;\ge\; 0$$

for $x \in (-1, 0)$, which shows that h is convex on $(-1, 0)$. Now, because $\lim_{x\to -1} h(x) > 0$ and $\lim_{x\to 0} h(x) = 0$, the statement follows. □

Theorem 4.22. *Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0$. We have $\sigma^2_{\boldsymbol{\vartheta}} > 0$ if one of the following conditions is satisfied:*

(i) $|\rho_{\boldsymbol{\vartheta}}(1)| < \sin(\pi/\sqrt{12}\,)$ *and* $\rho_{\boldsymbol{\vartheta}}(k) = 0$ *for all* $k > 1$.

(ii) $1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(3) > 0$ *and* $\rho_{\boldsymbol{\vartheta}}(k) < \rho_{\boldsymbol{\vartheta}}(k+1) < 0$ *for all* $k \in \mathbb{N}$.

(iii) *There exists some* $\boldsymbol{a} \in (-1, 1)$ *with* $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ *for all* $k \in \mathbb{N}$.

Proof. *(i)* By Lemma 4.2 *(ii)*, we obtain $\gamma_{\boldsymbol{\vartheta}}(k) = \Psi^*(\rho_{\boldsymbol{\vartheta}}(1),\, 0,\, 0,\, 0) = 0$ for all $k > 1$. Thus, according to formulas (4.5) and (4.6),

$$\sigma^2_{\boldsymbol{\vartheta}} \;=\; \gamma_{\boldsymbol{\vartheta}}(0) + 2\,\gamma_{\boldsymbol{\vartheta}}(1) \;=\; \frac{1}{4} - \frac{3}{\pi^2}(\arcsin \rho_{\boldsymbol{\vartheta}}(1))^2\,.$$

Under the assumptions, the latter expression is strictly positive.

(ii) Let $k > 1$. By Corollary 4.16, we obtain $\gamma_{\boldsymbol{\vartheta}}(k) > \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1))$. Thus, according to Corollary 4.14 *(i)*, $\gamma_{\boldsymbol{\vartheta}}(k) > 0$. It remains to show that $\gamma_{\boldsymbol{\vartheta}}(0) + 2\,\gamma_{\boldsymbol{\vartheta}}(1) \geq 0$. Note that, according to formulas (4.5) and (4.6),

$$\gamma_{\boldsymbol{\vartheta}}(0) \,+\, 2\,\gamma_{\boldsymbol{\vartheta}}(1) \;=\; f(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(2))$$

with f as defined in Lemma 4.20. Since $\rho_{\boldsymbol{\vartheta}}(1), \rho_{\boldsymbol{\vartheta}}(2) \in (-1, 0)$ and $\rho_{\boldsymbol{\vartheta}}(1) < \rho_{\boldsymbol{\vartheta}}(2)$, we only need to show that $2(\rho_{\boldsymbol{\vartheta}}(1))^2 - 1 \leq \rho_{\boldsymbol{\vartheta}}(2)$. By the model assumptions (M1)-(M3) on p. 42, $\mathbf{Y}$ is non-degenerate and Gaussian. Therefore, the matrix

$$\begin{pmatrix} 1 & \rho_{\boldsymbol{\vartheta}}(1) & \rho_{\boldsymbol{\vartheta}}(2) \\ \rho_{\boldsymbol{\vartheta}}(1) & 1 & \rho_{\boldsymbol{\vartheta}}(1) \\ \rho_{\boldsymbol{\vartheta}}(2) & \rho_{\boldsymbol{\vartheta}}(1) & 1 \end{pmatrix}$$

is strictly positive definite (see Theorem 2.4 *(i)*). Since the eigenvalues of a strictly positive definite matrix are greater than 0 (see Horn and Johnson [49], p. 398), Lemma 4.19 yields

$$\frac{1}{2}\big(\,2 + \rho_{\boldsymbol{\vartheta}}(2) - \sqrt{8(\rho_{\boldsymbol{\vartheta}}(1))^2 + (\rho_{\boldsymbol{\vartheta}}(2))^2}\,\big) \;>\; 0$$

and thus $\rho_{\boldsymbol{\vartheta}}(2) > 2(\rho_{\boldsymbol{\vartheta}}(1))^2 - 1$.

(iii) If $\boldsymbol{a} = 0$, then the statement follows according to *(i)*. Otherwise, Corollary 4.16 *(ii)* and Corollary 4.14 *(i)* yield $\gamma_{\boldsymbol{\vartheta}}(k) > 0$ for all $k > 1$. Furthermore, since $\gamma_{\boldsymbol{\vartheta}}(1) = g(\boldsymbol{a}, \boldsymbol{a}^2)$ with the function g as defined in Lemma 4.21 (compare to (4.6)), we obtain $\gamma_{\boldsymbol{\vartheta}}(1) \geq 0$. Clearly, $\gamma_{\boldsymbol{\vartheta}}(0) \geq 0$ (see (4.5)), and thus the proof is complete. □

For the proof of Theorem 4.24, we need the following lemma.

Lemma 4.23. *Assume $\mathcal{X}$ is a topological space and $I \subset \mathcal{X}$ with $I \neq \emptyset$ is compact. If $(g_k)_{k\in\mathbb{N}}$ is a sequence of continuous mappings from I onto $\mathbb{R}$ and there exists a $\beta > \frac{1}{2}$ with $|g_k(x_k)| = o(k^{-2\beta})$ for every sequence $(x_k)_{k\in\mathbb{N}}$ in I, then the sequence of mappings $(f_n)_{n\in\mathbb{N}}$ given by*

$$f_n(x) \;:=\; \frac{1}{n}\sum_{k=1}^{n-1}(n-k)\,g_k(x)$$

for $n \in \mathbb{N}$ and $x \in I$ uniformly converges to the mapping f given by

$$f(x) \;:=\; \sum_{k=1}^{\infty} g_k(x)$$

for $x \in I$, and f is continuous on I.

Proof. Note that $\sum_{k=1}^{\infty}|g_k(x)| < \infty$ for every $x \in I$, so f is well-defined. Since g_k is continuous on I for every $k \in \mathbb{N}$, there exists a sequence $(x_k)_{k\in\mathbb{N}}$ in I satisfying $|g_k(x)| \leq |g_k(x_k)|$ for all $k \in \mathbb{N}$ and $x \in I$. With this we obtain

$$\begin{aligned}
\big|f(x) - f_n(x)\big| &= \Big|\sum_{k=1}^{n-1}\frac{k}{n}\,g_k(x) + \sum_{k=n}^{\infty} g_k(x)\Big| \\
&= \Big|\sum_{k=1}^{\infty}\min\big\{k/n,\,1\big\}\,g_k(x)\Big| \\
&\leq \sum_{k=1}^{\infty}\min\big\{k/n,\,1\big\}\,\big|g_k(x)\big| \\
&\leq \sum_{k=1}^{\infty}\min\big\{k/n,\,1\big\}\,\big|g_k(x_k)\big|
\end{aligned}$$

for all $x \in I$, that is, the absolute difference between f and f_n on I is uniformly bounded by $\sum_{k=1}^{\infty}\min\big\{k/n,\,1\big\}\,\big|g_k(x_k)\big|$. Since $|g_k(x_k)| = o(k^{-2\beta})$, we obtain $\sum_{k=1}^{\infty}\big|g_k(x_k)\big| < \infty$ and thus, according to the Dominated Convergence Theorem,

$$\lim_{n\to\infty}\sum_{k=1}^{\infty}\min\big\{k/n,1\big\}\,\big|g_k(x_k)\big| \;=\; \sum_{k=1}^{\infty}\lim_{n\to\infty}\min\big\{k/n,1\big\}\,\big|g_k(x_k)\big| \;=\; 0\,,$$

which shows that f_n converges to f uniformly on I. Obviously, f_n is continuous on I for every $n \in \mathbb{N}$. Since the limit of a uniformly converging sequence of continuous mappings on a compact set is again a continuous mapping, it follows that f is continuous. □

Theorem 4.24. *Suppose $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}$. Furthermore, let $I \subset \boldsymbol{\Theta}_0$ with $I \neq \emptyset$ be compact, and suppose $\boldsymbol{\vartheta} \mapsto \rho_{\boldsymbol{\vartheta}}(k)$ is continuous on I for every $k \in \mathbb{Z}$.*

Then $\boldsymbol{\vartheta} \mapsto n \cdot \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ converges to $\boldsymbol{\vartheta} \mapsto \sigma^2_{\boldsymbol{\vartheta}}$ uniformly on I as $n \to \infty$ and $\boldsymbol{\vartheta} \mapsto \sigma^2_{\boldsymbol{\vartheta}}$ is continuous on I if one of the following conditions is satisfied:

(i) There exist a $k_0 \in \mathbb{N}$ and a function $\nu : \mathbb{N} \to \mathbb{R}$ with $\nu(1) > -1$ and $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ such that, for every $\boldsymbol{\vartheta} \in I$,

(a) $1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(k_0) > 0$,

(b) $\rho_{\boldsymbol{\vartheta}}(1) < 0$ and $\rho_{\boldsymbol{\vartheta}}(k) < \rho_{\boldsymbol{\vartheta}}(k+1) < 0$ for all $k \geq k_0$,

(c) $\nu(1) \leq \rho_{\boldsymbol{\vartheta}}(1)$ and $\nu(k) \leq \rho_{\boldsymbol{\vartheta}}(k)$ for all $k \geq k_0$.

(ii) There exists a $k_0 \in \mathbb{N}$ such that the following is valid: for every $\boldsymbol{\vartheta} \in I$, we find some $\boldsymbol{a} \in (-1, 1)$ with $\rho_{\boldsymbol{\vartheta}}(1) = \boldsymbol{a}$ and $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ for all $k \geq k_0$.

Proof. Define $g_k(\boldsymbol{\vartheta}) := \gamma_{\boldsymbol{\vartheta}}(k)$ for $k \in \mathbb{Z}$ and $\boldsymbol{\vartheta} \in I$. Furthermore, let

$$f_n(\boldsymbol{\vartheta}) \;:=\; \frac{1}{n} \sum_{k=1}^{n-1} (n-k)\, g_k(\boldsymbol{\vartheta}) \qquad \text{and} \qquad f(\boldsymbol{\vartheta}) \;:=\; \sum_{k=1}^{\infty} g_k(\boldsymbol{\vartheta})$$

for $n \in \mathbb{N}$ and $\boldsymbol{\vartheta} \in I$. Note that $n \cdot \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) \;=\; g_0(\boldsymbol{\vartheta}) + 2 f_n(\boldsymbol{\vartheta})$ and $\sigma^2_{\boldsymbol{\vartheta}} \;=\; g_0(\boldsymbol{\vartheta}) + 2 f(\boldsymbol{\vartheta})$. Thus, according to Lemma 4.23, it is sufficient to show that $\boldsymbol{\vartheta} \mapsto \gamma_{\boldsymbol{\vartheta}}(k)$ is continuous on I for all $k \geq 0$ and that there exists a $\beta > \frac{1}{2}$ such that

$$|\gamma_{\boldsymbol{\vartheta}_k}(k)| \;\;=\;\; o(k^{-2\beta}) \tag{4.31}$$

for every sequence $(\boldsymbol{\vartheta}_k)_{k \in \mathbb{N}}$ in I.

First, we establish continuity of $\boldsymbol{\vartheta} \mapsto \gamma_{\boldsymbol{\vartheta}}(k)$. For $k = 0$ and $k = 1$, this is an immediate consequence of formulas (4.5), (4.6) and the assumption that $\boldsymbol{\vartheta} \mapsto \rho_{\boldsymbol{\vartheta}}(k)$ is continuous on I for every $k \in \mathbb{Z}$. In order to establish continuity for $k > 1$, let $m \in \mathbb{N}$ and $t_1, t_2, \ldots, t_m \in \mathbb{Z}$. Suppose $(\boldsymbol{\vartheta}_k)_{k \in \mathbb{N}}$ is a convergent sequence in I, and let $\boldsymbol{\vartheta}^* := \lim_{k \to \infty} \boldsymbol{\vartheta}_k$. Under the assumptions, $\boldsymbol{\vartheta} \mapsto \rho_{\boldsymbol{\vartheta}}(t_i - t_j)$ is continuous on I and thus

$$\lim_{k \to \infty} \mathrm{Corr}_{\boldsymbol{\vartheta}_k}(Y_{t_i}, Y_{t_j}) \;\;=\;\; \mathrm{Corr}_{\boldsymbol{\vartheta}^*}(Y_{t_i}, Y_{t_j})$$

for all $i, j = 1, 2, \ldots, m$. Since $\mathbf{Y}$ is zero-mean Gaussian and has unit variance for every $\boldsymbol{\vartheta} \in I$ (see model assumption (M3), p. 42), Theorem 2.4 (iv) shows that the distribution of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}_k}$ converges to that of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}^*}$. Now, let $B_i = [0, \infty)$ or $B_i = (-\infty, 0]$ for $i = 1, 2, \ldots, m$. Since $\partial(B_1 \times B_2 \times \ldots \times B_m) = \{\mathbf{0}\}$ has Lebesgue-measure 0, we obtain

$$\mathbb{P}_{\boldsymbol{\vartheta}_k}(Y_{t_1} \in B_1,\, Y_{t_2} \in B_2,\, \ldots,\, Y_{t_m} \in B_m) \;\;\longrightarrow\;\; \mathbb{P}_{\boldsymbol{\vartheta}^*}(Y_{t_1} \in B_1,\, Y_{t_2} \in B_2,\, \ldots,\, Y_{t_m} \in B_m)$$

as $k \to \infty$, which shows that

$$\boldsymbol{\vartheta} \;\mapsto\; \mathbb{P}_{\boldsymbol{\vartheta}}(Y_{t_1} \in B_1, Y_{t_2} \in B_2, \ldots, Y_{t_m} \in B_m) \tag{4.32}$$

is continuous on I. According to equation (4.7), $\boldsymbol{\vartheta} \mapsto \gamma_{\boldsymbol{\vartheta}}(k)$ with $k > 1$ can be written as the sum and product, respectively, of mappings as given in (4.32). Therefore, $\boldsymbol{\vartheta} \mapsto \gamma_{\boldsymbol{\vartheta}}(k)$ is continuous on I.

In order to establish (4.31), let $(\boldsymbol{\vartheta}_k)_{k\in\mathbb{N}}$ be a sequence in I. First, we assume that (i) holds. By Corollaries 4.16 (i) and 4.17 (i) we obtain

$$\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1)) \;<\; \gamma_{\boldsymbol{\vartheta}}(k) \;<\; \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k-1))$$

for all $k > k_0$ and $\boldsymbol{\vartheta} \in I$. Since $\Psi^{**}(\mathbf{r}) \geq 0$ for every $\mathbf{r} \in \mathcal{R}^{**}$ (see Corollary 4.14 (i)), it follows that

$$|\gamma_{\boldsymbol{\vartheta}_k}(k)| \;\leq\; \Psi^{**}(\rho_{\boldsymbol{\vartheta}_k}(1),\, \rho_{\boldsymbol{\vartheta}_k}(k-1))$$

for all $k > k_0$. Under the assumptions, $\nu(k) \to 0$ as $k \to \infty$. Therefore, we find a $k_1 \geq k_0$ such that $2|\nu(k)| < 1 + \nu(1)$ and thus $(\nu(1), \nu(k)) \in \mathcal{R}^{**}$ for all $k \geq k_1$ (see (4.21)). Since $\nu(1) \leq \rho_{\boldsymbol{\vartheta}_k}(1)$ and $\nu(k-1) \leq \rho_{\boldsymbol{\vartheta}_k}(k-1) < 0$ for all $k > k_1$, Corollary 4.14 (ii) yields

$$\Psi^{**}(\rho_{\boldsymbol{\vartheta}_k}(1),\, \rho_{\boldsymbol{\vartheta}_k}(k-1)) \;\leq\; \Psi^{**}(\nu(1),\, \nu(k-1))$$

for all $k > k_1$. According to (4.22), we have $\Psi^{**}(\nu(1),\, \nu(k-1)) = O((\nu(k))^2)$. Therefore,

$$\Psi^{**}(\nu(1),\, \nu(k-1)) \;=\; o(k^{-2\beta})$$

and hence the statement follows.

Now, we assume that (ii) holds. Let $\boldsymbol{\vartheta} \in I$ and $\boldsymbol{a} \in (-1, 1)$ be such that $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ for $k = 1$ and $k \geq k_0$. If $\boldsymbol{a} = 0$, then Lemma 4.2 (ii) shows that $\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1)) = \gamma_{\boldsymbol{\vartheta}}(k) = \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k-1)) = 0$ for all $k > k_0$. Otherwise, Corollaries 4.16 (ii) and 4.17 (ii) yield $\Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k+1)) < \gamma_{\boldsymbol{\vartheta}}(k) < \Psi^{**}(\rho_{\boldsymbol{\vartheta}}(1),\, \rho_{\boldsymbol{\vartheta}}(k-1))$ for all $k > k_0$. By the same argument as above, we obtain

$$|\gamma_{\boldsymbol{\vartheta}_k}(k)| \;\leq\; \Psi^{**}(\rho_{\boldsymbol{\vartheta}_k}(1),\, \rho_{\boldsymbol{\vartheta}_k}(k-1))$$

for all $k > k_0$. Since I is compact and $\boldsymbol{\vartheta} \mapsto \rho_{\boldsymbol{\vartheta}}(1)$ is continuous, $\boldsymbol{a}_{\max} := \max_{\boldsymbol{\vartheta}\in I} |\rho_{\boldsymbol{\vartheta}}(1)|$ is well-defined and lies in $[0, 1)$. Note that $-\boldsymbol{a}_{\max} \leq \rho_{\boldsymbol{\vartheta}_k}(1)$ and $-\boldsymbol{a}_{\max}^{k-1} \leq -|\rho_{\boldsymbol{\vartheta}_k}(k-1)| < 0$ for all $k > k_0$. Therefore, according to Corollary 4.14 (ii),

$$\Psi^{**}(\rho_{\boldsymbol{\vartheta}_k}(1),\, \rho_{\boldsymbol{\vartheta}_k}(k-1)) \;\leq\; \Psi^{**}(-\boldsymbol{a}_{\max},\, -\boldsymbol{a}_{\max}^{k-1})$$

for all $k > k_0$. By (4.22) we obtain $\Psi^{**}(-\boldsymbol{a}_{\max},\, -\boldsymbol{a}_{\max}^{k-1}) = O(\boldsymbol{a}_{\max}^{2k})$, and hence

$$\Psi^{**}(-\boldsymbol{a}_{\max},\, -\boldsymbol{a}_{\max}^{k-1}) \;=\; o(k^{-2\beta})$$

for every $\beta > \frac{1}{2}$. The proof is complete. □

Chapter 5

The case d=2

In this chapter, we focus on ordinal patterns of order $d = 2$. The framework of analysis is the same as in the previous chapters: $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a family of real-valued measurable mappings defined on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a non-empty family of probability measures $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}$. By $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ we denote the process of increments given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$. **We always assume that $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) on p. 42.** For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z}$, let $\rho_{\boldsymbol{\vartheta}}(k) = \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0, Y_k)$ denote the autocorrelations of $\mathbf{Y}$.

In Section 5.1, we establish a simple relation between the ordinal patterns of order $d = 2$ and changes between "upwards" and "downwards". Using the results of Chapter 4, we evaluate the variance of estimators of ordinal pattern probabilities. Examples are given in Section 5.2.

In Section 5.3, we consider the estimation of the parameter $\boldsymbol{\vartheta}$ in the case where $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}$ and the probability of a change is strictly monotone in $\boldsymbol{\vartheta}$. Under additional assumptions on the autocorrelations of $\mathbf{Y}$, we derive asymptotic confidence intervals for $\boldsymbol{\vartheta}$. We apply the results to the estimation of the Hurst parameter in FBM, of the fractional differencing parameter in ARFIMA(0,d,0) processes and of the autoregressive coefficient in AR(1) processes.

In Section 5.4, we evaluate the performance of the parameter estimates in a simulation study. Furthermore, we consider the distribution of the number of changes between "upwards" and "downwards" in equidistant discretizations of FBM. It turns out that the distribution is very irregular when the Hurst parameter is large.

Note that in the case $d = 1$, the estimation of ordinal pattern probabilities is "trivial". In particular, $p_{\mathbf{r}}(\cdot) = \frac{1}{2}$ for $\mathbf{r} = (0, 1)$ and $\mathbf{r} = (1, 0)$. Furthermore, $\hat{p}_{\mathbf{r},n}$ is consistently equal to $\frac{1}{2}$ for all $n \in \mathbb{N}$. According to this argument, we consider the simplest non-trivial case in this chapter.

5.1 Changes between "upwards" and "downwards"

For $t \in \mathbb{Z}$, define

$$C(t) \;:=\; \mathbf{1}_{\{X_t \geq X_{t+1} < X_{t+2}\}} + \mathbf{1}_{\{X_t < X_{t+1} \geq X_{t+2}\}} \,.$$

We may regard $C(t)$ as the indicator for a *change*, either from "downwards" to "upwards" (when $X_t \geq X_{t+1}$ and $X_{t+1} < X_{t+2}$), or from "upwards" to "downwards" (when $X_t < X_{t+1}$ and $X_{t+1} \geq X_{t+2}$). In terms of the increments, $C(t)$ can be written as

$$C(t) \;=\; \mathbf{1}_{\{Y_{t+1} \leq 0,\, Y_{t+2} > 0\}} + \mathbf{1}_{\{Y_{t+1} > 0,\, Y_{t+2} \leq 0\}} \,. \tag{5.1}$$

Thus, changes between "upwards" and "downwards" in $\mathbf{X}$ are equivalent to zero-crossings in $\mathbf{Y}$.

As illustrated by Table 5.1, a change from "downwards" to "upwards" is equivalent to an occurrence of the ordinal patterns $(2,0,1)$ or $(0,2,1)$, that is,

$$X_t \geq X_{t+1} < X_{t+2} \quad \text{if and only if} \quad \Pi(t) = (2,0,1) \text{ or } \Pi(t) = (0,2,1) \,.$$

A change from "upwards" to "downwards" is equivalent to an occurrence of $(1,2,0)$ or $(1,0,2)$, that is,

$$X_t < X_{t+1} \geq X_{t+2} \quad \text{if and only if} \quad \Pi(t) = (1,2,0) \text{ or } \Pi(t) = (1,0,2) \,.$$

The event "no change" is equivalent to an occurrence of the patterns $(2,1,0)$ or $(0,1,2)$. Consequently, the indicator for a change can be written as

$$C(t) \;=\; \mathbf{1}_{\{\Pi(t)=(2,0,1)\}} + \mathbf{1}_{\{\Pi(t)=(0,2,1)\}} + \mathbf{1}_{\{\Pi(t)=(1,2,0)\}} + \mathbf{1}_{\{\Pi(t)=(1,0,2)\}} \,, \tag{5.2}$$

and the indicator for "no change" as

$$1 - C(t) \;=\; \mathbf{1}_{\{\Pi(t)=(2,1,0)\}} + \mathbf{1}_{\{\Pi(t)=(0,1,2)\}} \,. \tag{5.3}$$

With the set $\bar{\mathbf{r}}$ defined on p. 47, these relations can be represented in a more compact form, namely,

$$C(t) \;=\; \begin{cases} \mathbf{1}_{\{\Pi(t) \in \bar{\mathbf{r}}\}} & \text{if } \mathbf{r} \in \big\{(2,0,1),\, (0,2,1),\, (1,2,0),\, (1,0,2)\big\} \\ 1 - \mathbf{1}_{\{\Pi(t) \in \bar{\mathbf{r}}\}} & \text{if } \mathbf{r} \in \big\{(2,1,0),\, (0,1,2)\big\} \end{cases} \,. \tag{5.4}$$

The probability of a change. Since $(Y_t)_{t \in \mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ (see model assumption (M2)) and $C(t)$ is a measurable function of Y_{t+1} and Y_{t+2} for every $t \in \mathbb{Z}$ (see (5.1)), we immediately obtain the following statement.

					
$(2,1,0)$	$(2,0,1)$	$(0,2,1)$	$(1,2,0)$	$(1,0,2)$	$(0,1,2)$

Table 5.1: Ordinal patterns of order $d = 2$.

Corollary 5.1. *The process $(C(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.*

As in Chapter 4, let $c(\cdot)$ denote the probability of a change in $\mathbf{X}$ or, equivalently, of a zero-crossing in $\mathbf{Y}$, that is,

$$c(\boldsymbol{\vartheta}) \;:=\; \mathbb{P}_{\boldsymbol{\vartheta}}(C(t) = 1) \tag{5.5}$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Since the process $(C(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, $c(\boldsymbol{\vartheta})$ does not depend on the specific time point t on the right hand side of (5.5). According to equation (4.4) on p. 69, we have

$$c(\boldsymbol{\vartheta}) \;=\; \frac{1}{2} - \frac{1}{\pi} \arcsin \rho_{\boldsymbol{\vartheta}}(1) \tag{5.6}$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Thus, the higher the first-order autocorrelation of $\mathbf{Y}$, the lower the probability of a change. In particular, when the first-order autocorrelation tends to -1, the probability of a change tends to 1, and when the first-order autocorrelation tends to 1, the probability of a change tends to 0.

Next, we show how the probability of a change is related to ordinal pattern probabilities. Note that, for any $\mathbf{r} \in S_d$ with $d \in \mathbb{N}$, the ordinal patterns in $\bar{\mathbf{r}}$ have the same probability (see the discussion in Section 3.3). In particular,

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (2,0,1)) \;&=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (0,2,1)) \\ &=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (1,2,0)) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (1,0,2)) \end{aligned}$$

and

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (2,1,0)) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (0,1,2))$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Furthermore, according to (5.2) and (5.3),

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\vartheta}}(C(t) = 1) \;=\; &\mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (2,0,1)) + \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (0,2,1)) \\ &+ \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (1,2,0)) + \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (1,0,2)) \end{aligned}$$

and

$$1 - \mathbb{P}_{\boldsymbol{\vartheta}}(C(t) = 1) \;=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (2,1,0)) + \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi(t) = (0,1,2))\,,$$

which shows that

$$(5.7) \qquad p_{\mathbf{r}}(\boldsymbol{\vartheta}) \;=\; \begin{cases} \frac{1}{4}\, c(\boldsymbol{\vartheta}) & \text{if } \mathbf{r} \in \big\{(1,0,2),\, (1,2,0),\, (0,2,1),\, (2,0,1)\big\} \\ \frac{1}{2}\,(1 - c(\boldsymbol{\vartheta})) & \text{if } \mathbf{r} \in \big\{(2,1,0),\, (0,1,2)\big\} \end{cases}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

According to (5.7), any ordinal pattern probability in the case $d = 2$ is an affine function of the probability of a change. By inserting the expression for $c(\cdot)$ given in (5.6) into the right hand side of (5.7), we obtain closed-form expressions for $p_{\mathbf{r}}(\cdot)$. Note that these expressions have been first derived in Bandt and Shiha [12].

Estimating the probability of a change. As in Chapter 4, let $\hat{c}_n$ denote the estimator for the probability of a change given by

$$\hat{c}_n \;:=\; \frac{1}{n} \sum_{t=0}^{n-1} C(t)\,.$$

According to (5.4), we have

$$\begin{aligned} \hat{c}_n \;&=\; \frac{1}{n} \sum_{t=0}^{n-1} \Big(\mathbf{1}_{\{\Pi(t)=(2,0,1)\}} + \mathbf{1}_{\{\Pi(t)=(0,2,1)\}} + \mathbf{1}_{\{\Pi(t)=(1,2,0)\}} + \mathbf{1}_{\{\Pi(t)=(1,0,2)\}} \Big) \\ &=\; \frac{1}{n} \sum_{t=0}^{n-1} \Big(1 - \mathbf{1}_{\{\Pi(t)=(2,1,0)\}} - \mathbf{1}_{\{\Pi(t)=(0,1,2)\}} \Big)\,. \end{aligned}$$

Therefore, by the definition of $\hat{p}_{\mathbf{r},n}$ (see (3.13) on p. 50),

$$(5.8) \qquad \hat{p}_{\mathbf{r},n} \;=\; \begin{cases} \frac{1}{4}\, \hat{c}_n & \text{if } \mathbf{r} \in \big\{(1,0,2),\, (1,2,0),\, (0,2,1),\, (2,0,1)\big\} \\ \frac{1}{2}\,(1 - \hat{c}_n) & \text{if } \mathbf{r} \in \big\{(2,1,0),\, (0,1,2)\big\} \end{cases}\,.$$

This shows that, in the case $d = 2$, any reasonable estimator of ordinal pattern probabilities ("reasonable" in the sense of Theorem 3.5) is an affine function of $\hat{c}_n$. In particular,

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{p}_{\mathbf{r},n}) \;=\; \begin{cases} \frac{1}{16}\, \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) & \text{if } \mathbf{r} \in \big\{(1,0,2),\, (1,2,0),\, (0,2,1),\, (2,0,1)\big\} \\ \frac{1}{4}\, \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n) & \text{if } \mathbf{r} \in \big\{(2,1,0),\, (0,1,2)\big\} \end{cases}$$

and thus the results on the variance of $\hat{c}_n$ established in Chapter 4 can be used to evaluate the variance of $\hat{p}_{\mathbf{r},n}$. On the other hand, $\hat{c}_n$ has essentially the same statistical properties as $\hat{p}_{\mathbf{r},n}$. Thus, according to the results of Chapter 3, we obtain the following statement.

Corollary 5.2.

(*i*) *$\hat{c}_n$ is an unbiased estimator of $c(\cdot)$.*

(*ii*) *If $\rho_{\boldsymbol{\vartheta}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $h : [0,1] \to \mathbb{R}$ is continuous on an open set containing $c(\boldsymbol{\Theta})$, then $h(\hat{c}_n)$ is a strongly consistent estimator of $h(c(\cdot))$. If, additionally, h is bounded on $[0,1]$, then $h(\hat{c}_n)$ is an asymptotically unbiased estimator of $h(c(\cdot))$.*

(*iii*) *If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ and $h : [0,1] \to \mathbb{R}$ has a non-vanishing first derivative at $c(\boldsymbol{\vartheta})$, then*

$$\sqrt{n}\,\big(\,h(\hat{c}_n) - h(c(\boldsymbol{\vartheta}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{\vartheta}}[h'(c(\boldsymbol{\vartheta}))]^2\,)\,,$$

where

$$\sigma^2_{\boldsymbol{\vartheta}} \;:=\; \gamma_{\boldsymbol{\vartheta}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{\vartheta}}(k)$$

and $\gamma_{\boldsymbol{\vartheta}}(k) := \mathrm{Cov}_{\boldsymbol{\vartheta}}(C(0), C(k))$ for $k \in \mathbb{Z}$.

Proof. (*i*) is an immediate consequence of Corollary 5.1.

(*ii*) Let $\mathbf{r} = (2,0,1)$. (For any other choice of $\mathbf{r} \in S_d$ with $d = 2$, the proof is similar.) According to (5.7) and (5.8), we have

$$\hat{c}_n \;=\; 4\,\hat{p}_{\mathbf{r},n} \quad \text{ and } \quad c(\boldsymbol{\vartheta}) \;=\; 4\,p_{\mathbf{r}}(\boldsymbol{\vartheta}) \text{ for every } \boldsymbol{\vartheta} \in \boldsymbol{\Theta}.$$

Now, let $\tilde{h}(x) := h(4x)$ for $x \in [0,1]$. Under the assumptions, Theorem 3.8 shows that $\tilde{h}(\hat{p}_{\mathbf{r},n})$ is a strongly consistent (asymptotically unbiased) estimator of $\tilde{h}(p_{\mathbf{r}}(\cdot))$, and thus $h(\hat{c}_n)$ is a strongly consistent (asymptotically unbiased) estimator of $h((\cdot))$.

(*iii*) Under the assumptions, the mapping $\tilde{h}$ defined in the proof of (*ii*) has a non-vanishing first derivative at $p_{\mathbf{r}}(\boldsymbol{\vartheta})$. Therefore, according to Corollary 3.17,

$$\sqrt{n}\,\big(\,\tilde{h}(\hat{p}_{\mathbf{r},n}) - \tilde{h}(p_{\mathbf{r}}(\boldsymbol{\vartheta}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}} \quad \mathrm{N}(0,\, \tilde{\sigma}^2_{\boldsymbol{\vartheta}}[\tilde{h}'(p_{\mathbf{r}}(\boldsymbol{\vartheta}))]^2\,)\,,$$

where

$$\tilde{\sigma}^2_{\boldsymbol{\vartheta}} \;:=\; \tilde{\gamma}_{\boldsymbol{\vartheta}}(0) + 2\sum_{k=1}^{\infty} \tilde{\gamma}_{\boldsymbol{\vartheta}}(k)$$

and $\tilde{\gamma}_{\boldsymbol{\vartheta}}(k) := \frac{1}{(\sharp\,\bar{\mathbf{r}})^2}\,\mathrm{Cov}_{\boldsymbol{\vartheta}}(\mathbf{1}_{\{\Pi(0)\in\bar{\mathbf{r}}\}},\, \mathbf{1}_{\{\Pi(k)\in\bar{\mathbf{r}}\}})$ for $k \in \mathbb{Z}$. Now, note that $\tilde{h}(\hat{p}_{\mathbf{r},n}) = h(\hat{c}_n)$, $\tilde{h}(p_{\mathbf{r}}(\boldsymbol{\vartheta})) = h(c(\boldsymbol{\vartheta}))$ and $[\tilde{h}'(p_{\mathbf{r}}(\boldsymbol{\vartheta}))]^2 = 16\,[h'(c(\boldsymbol{\vartheta}))]^2$. Since $\sharp\,\bar{\mathbf{r}} = 4$ and $\mathbf{1}_{\{\Pi(t)\in\bar{\mathbf{r}}\}} = C(t)$ for every $t \in \mathbb{Z}$, we obtain $\tilde{\sigma}^2_{\boldsymbol{\vartheta}} = \frac{1}{16}\,\sigma^2_{\boldsymbol{\vartheta}}$, and thus the proof is complete. □

5.2 Examples

In this section, we consider the estimation of the probability of a change in equidistant discretizations of FBM and in processes where the increments are ARFIMA(0,d,0) and AR(1). In particular, we use the results of Chapter 4 to evaluate the variance of $\hat{c}_n$.

Equidistant discretizations of FBM. Suppose that $\boldsymbol{\Theta} = (0,1)$ and, for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ is standard FGN with the Hurst parameter $\boldsymbol{\vartheta}$. As in Section 3.5, we write $\boldsymbol{H}$ instead of $\boldsymbol{\vartheta}$.

According to the definition of FGN, the first-order autocorrelation of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is given by

$$\rho_{\boldsymbol{H}}(1) \;=\; 2^{2\boldsymbol{H}-1} - 1 \tag{5.9}$$

for $\boldsymbol{H} \in (0,1)$ (see Section 2.2.3). By formula (5.6) and the fact that

$$\arcsin x \;=\; 2\arcsin\sqrt{(1+x)/2} - \frac{\pi}{2}$$

for $x \in [-1,1]$, we obtain

$$c(\boldsymbol{H}) \;=\; \frac{1}{2} - \frac{1}{\pi}\arcsin\rho_{\boldsymbol{H}}(1) \;=\; 1 - \frac{2}{\pi}\arcsin 2^{\boldsymbol{H}-1} \tag{5.10}$$

for $\boldsymbol{H} \in (0,1)$. Equation (5.10) shows that the higher the Hurst parameter, the lower the probability of a change. Since $\arcsin\frac{1}{2} = \frac{\pi}{6}$ and $\arcsin 1 = 0$, we obtain $c(\boldsymbol{H}) \to \frac{2}{3}$ as $\boldsymbol{H} \to 0$ and $c(\boldsymbol{H}) \to 0$ as $\boldsymbol{H} \to 1$. In Section 5.3, we use the monotonic relation between $\boldsymbol{H}$ and $c(\boldsymbol{H})$ to derive an estimator for the Hurst parameter.

Note that the sample paths of FBM with the Hurst parameter $\boldsymbol{H}$ almost surely have Hausdorff dimension $2 - \boldsymbol{H}$ (see Proposition 2.5 in Taqqu [96]). Therefore, the Hurst parameter is often regarded as a measure for the "roughness" of FBM. The monotonic relation between $\boldsymbol{H}$ and $c(\boldsymbol{H})$ shows that the Hurst parameter is also a measure for the roughness of equidistant discretizations of FBM: the higher the Hurst parameter, the lower the number of changes in the sample paths.

Next, let us investigate the variance of $\hat{c}_n$. According to equation (4.3) in Chapter 4, we have

$$\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n) \;=\; \frac{1}{n^2}\Big(n\,\gamma_{\boldsymbol{H}}(0) + 2\sum_{k=1}^{n-1}(n-k)\,\gamma_{\boldsymbol{H}}(k)\Big)\,,$$

where

$$\gamma_{\boldsymbol{H}}(k) \;:=\; \mathrm{Cov}_{\boldsymbol{H}}(C(0), C(k))$$

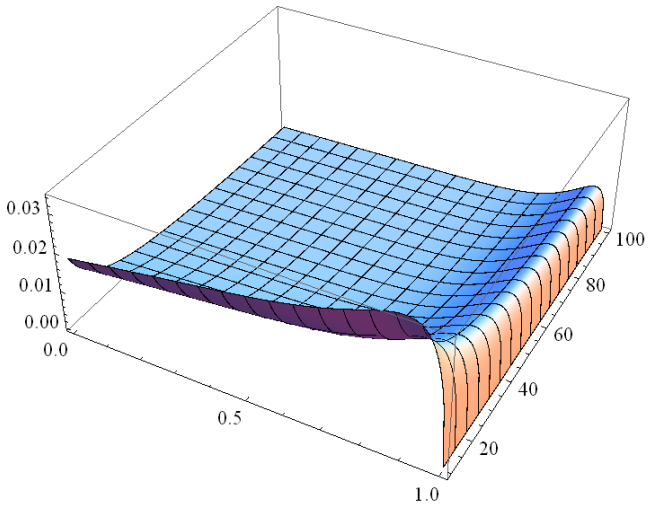

Figure 5.1: $\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n)$ for $\boldsymbol{H} \in (0,1)$ and $n = 10, 11, \ldots, 100$.

for $k \in \mathbb{Z}$. Using formulas (4.5) and (4.6), we can compute $\gamma_{\boldsymbol{H}}(0)$ and $\gamma_{\boldsymbol{H}}(1)$. For $k > 1$, $\gamma_{\boldsymbol{H}}(k)$ can be represented in terms of the mapping Ψ^* defined in (4.19), namely,

$$\gamma_{\boldsymbol{H}}(k) \quad = \quad \Psi^*(\rho_{\boldsymbol{H}}(1),\, \rho_{\boldsymbol{H}}(k),\, \rho_{\boldsymbol{H}}(k+1),\, \rho_{\boldsymbol{H}}(k-1)) \tag{5.11}$$

(see (4.20)). Using the integral representation of Ψ^* given in Theorem 4.6, we can evaluate the right hand side of (5.11) numerically. As explained in Section 4.5, when n is large and thus an "exact" numerical evaluation of $\gamma_{\boldsymbol{\vartheta}}(k)$ for $k = 0, 1, \ldots, n-1$ is time-consuming, approximate methods can be used.

For $\boldsymbol{H} = \frac{1}{2}$, we obtain a simple closed-form expression for $\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n)$. Since $\rho_{\boldsymbol{H}}(k) = 0$ for every $k \in \mathbb{N}$ (see Lemma 2.7 (i)), formulas (4.5) and (4.6) yield $\gamma_{\boldsymbol{H}}(0) = \frac{1}{4}$ and $\gamma_{\boldsymbol{H}}(1) = 0$. Furthermore, according to Lemma 4.2 (ii), $\Psi^*(0,\, 0,\, 0,\, 0) = 0$ and hence $\gamma_{\boldsymbol{H}}(k) = 0$ for $k > 1$. (This can also be seen from the fact that $Y_1, Y_2, \ldots$ are uncorrelated and thus independent, and $C(0)$, $C(k)$ depend on the disjoint blocks (Y_1, Y_2) and (Y_{k+1}, Y_{k+2}), respectively.) Thus, altogether, we have $\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n) = \frac{1}{n}\,\gamma_{\boldsymbol{H}}(0) = \frac{1}{4n}$.

Figure 5.1 displays $\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n)$ for $\boldsymbol{H} \in (0,1)$ and $n = 10, 11, \ldots, 100$. As can be seen, the variance of $\hat{c}_n$ is particularly high for large values of the Hurst parameter, but decreases to 0 as $\boldsymbol{H}$ tends to 1. The reason for the latter is that, according to (5.10), the probability of a change tends to 0 as $\boldsymbol{H}$ tends to 1. Therefore, the probability that $\hat{c}_n$ is equal to 0 tends to 1, and hence the variance of $\hat{c}_n$ decreases to 0.

Next, we investigate asymptotic properties of $\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n)$. Let $\boldsymbol{H} \in (0,1)$. For $k \in \mathbb{N}$, define $f(k) := k^{-(2-2\boldsymbol{H})}$. According to Lemma 2.7 (ii), we have $\rho_{\boldsymbol{H}}(k) \sim \boldsymbol{H}(2\boldsymbol{H}-1)f(k)$. We distinguish three cases:

(*i*) If $\boldsymbol{H} < \frac{3}{4}$, then $2 - 2\boldsymbol{H} > \frac{1}{2}$ and thus, according to Theorem 4.18 (*i*),

$$\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n) \;\sim\; \sigma_{\boldsymbol{H}}^2\, n^{-1}\,,$$

where $\sigma_{\boldsymbol{H}}^2 := \gamma_{\boldsymbol{H}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{H}}(k) < \infty$.

(*ii*) If $\boldsymbol{H} = \frac{3}{4}$, then $2-2\boldsymbol{H} = \frac{1}{2}$. Thus, according to Theorem 4.18 (*ii*) and the expression for $\rho_{\boldsymbol{H}}(1)$ given in (5.9),

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n) &\sim \frac{4\,(\boldsymbol{H}(2\boldsymbol{H}-1))^2(1-\rho_{\boldsymbol{H}}(1))}{\pi^2(1+\rho_{\boldsymbol{H}}(1))}\,\frac{\ln n}{n}\\ &= \frac{9\,(\sqrt{2}-1)}{16\,\pi^2}\,\frac{\ln n}{n}\,.\end{aligned}$$

(*iii*) If $\boldsymbol{H} > \frac{3}{4}$, then $2 - 2\boldsymbol{H} < \frac{1}{2}$ and thus, according to Theorem 4.18 (*iii*),

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{H}}(\hat{c}_n) &\sim \frac{4\,(\boldsymbol{H}(2\boldsymbol{H}-1))^2(1-\rho_{\boldsymbol{H}}(1))}{\pi^2(1+\rho_{\boldsymbol{H}}(1))(1-2(2-2\boldsymbol{H}))}\,n^{-2(2-2\boldsymbol{H})}\\ &= \frac{4\,(\boldsymbol{H}(2\boldsymbol{H}-1))^2(2^{2-2\boldsymbol{H}}-1)}{\pi^2\,(4\boldsymbol{H}-3)}\,n^{4\boldsymbol{H}-4}\,.\end{aligned}$$

ARFIMA(0,d,0) processes. Let $\boldsymbol{\Theta} = (-\frac{1}{2}, \frac{1}{2})$. As in Section 3.5, we write $\boldsymbol{d}$ instead of $\boldsymbol{\vartheta}$ for $\boldsymbol{\vartheta} \in (-\frac{1}{2}, \frac{1}{2})$. Suppose $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ is an ARFIMA(0,d,0) process with the fractional differencing parameter $\boldsymbol{d}$. According to the definition of ARFIMA(0,d,0) processes, the first-order autocorrelation of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{d}}$ is given by

$$\rho_{\boldsymbol{d}}(1) \;=\; \frac{\boldsymbol{d}}{1-\boldsymbol{d}}$$

for $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ (see Section 2.2.4). By formula (5.6), we obtain

$$c(\boldsymbol{d}) \;=\; \frac{1}{2} - \frac{1}{\pi}\arcsin\frac{\boldsymbol{d}}{1-\boldsymbol{d}}\,, \tag{5.12}$$

which shows that the higher the fractional differencing parameter, the lower the probability of a change.

In order to evaluate the variance of $\hat{c}_n$, we proceed as for equidistant discretizations of FBM in the previous paragraph. By the same argument as in the case $\boldsymbol{H} = \frac{1}{2}$, we obtain that $\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n) = \frac{1}{4n}$ if $\boldsymbol{d} = 0$. Figure 5.2 displays $\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n)$ for $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ and $n = 10, 11, \ldots, 100$. The picture is very similar to Figure 5.1. In particular, the variance

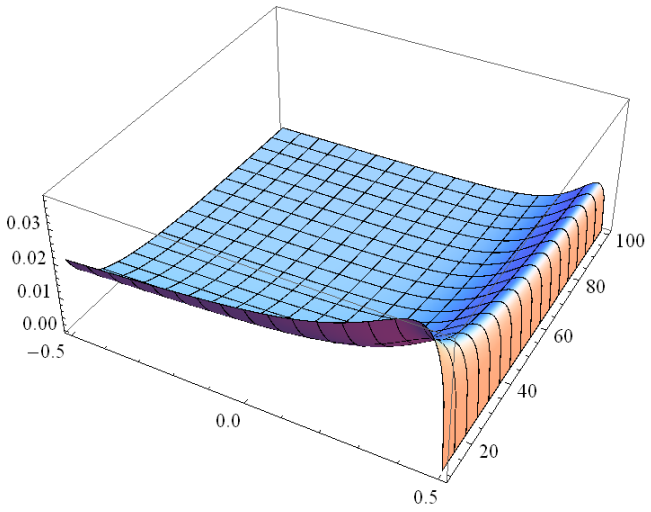

Figure 5.2: $\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n)$ for $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ and $n = 10, 11, \ldots, 100$.

of $\hat{c}_n$ is high when $\boldsymbol{d}$ is large and tends to 0 as $\boldsymbol{d}$ tends to $\frac{1}{2}$. Again, the reason for the latter is that the probability of a change tends to 0.

Let $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$. Similarly as in the previous paragraph, we derive asymptotic properties of $\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n)$. Define $f(k) := k^{-(1-2\boldsymbol{d})}$ for $k \in \mathbb{N}$, and note that $\rho_{\boldsymbol{d}}(k) \sim \frac{\Gamma(1-\boldsymbol{d})}{\Gamma(\boldsymbol{d})} f(k)$ (see Lemma 2.8 (i)). According to Theorem 4.18, we distinguish the following three cases:

(i) If $\boldsymbol{d} < \frac{1}{4}$, then $\sigma_{\boldsymbol{d}}^2 := \gamma_{\boldsymbol{d}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{d}}(k) < \infty$ and

$$\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n) \quad \sim \quad \sigma_{\boldsymbol{d}}^2 \, n^{-1} \, .$$

(ii) If $\boldsymbol{d} = \frac{1}{4}$, then

$$\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n) \quad \sim \quad \frac{2\left(\Gamma(\frac{3}{4})\right)^2}{\pi^2 \left(\Gamma(\frac{1}{4})\right)^2} \frac{\ln n}{n} \, .$$

(iii) If $\boldsymbol{d} > \frac{1}{4}$, then

$$\mathrm{Var}_{\boldsymbol{d}}(\hat{c}_n) \quad \sim \quad \frac{4\left(\Gamma(1-\boldsymbol{d})\right)^2 (1-2\boldsymbol{d})}{\pi^2 \left(\Gamma(\boldsymbol{d})\right)^2 (4\boldsymbol{d}-1)} \, n^{4\boldsymbol{d}-2} \, .$$

AR(1) processes. Now, suppose that $\boldsymbol{\Theta} = (-1, 1)$ and $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ with $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ is an AR(1) process with the autoregressive coefficient $\boldsymbol{\vartheta}$. As in Section 3.5, we

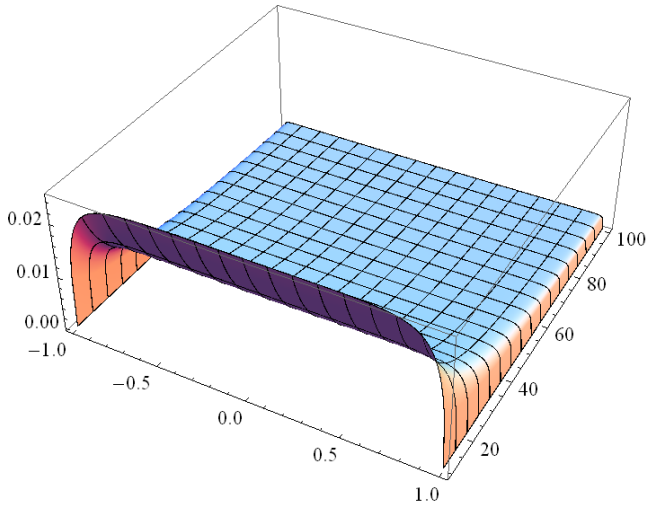

Figure 5.3: $\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n)$ for $\boldsymbol{a} \in (-1,1)$ and $n = 10, 11, \ldots, 100$.

write $\boldsymbol{a}$ instead of $\boldsymbol{\vartheta}$. By the definition of AR(1) processes, the first order autocorrelation of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ is given by

$$\rho_{\boldsymbol{a}}(1) \;=\; \boldsymbol{a}$$

for $\boldsymbol{a} \in (-1,1)$ (see Section 2.2.5). Thus, according to formula (5.6),

$$c(\boldsymbol{a}) \;=\; \frac{1}{2} - \frac{1}{\pi} \arcsin \boldsymbol{a} \tag{5.13}$$

for $\boldsymbol{a} \in (-1,1)$, which shows that the larger the autoregressive coefficient $\boldsymbol{a}$, the lower the probability of a change $c(\boldsymbol{a})$.

Similarly as in the previous paragraphs, we can evaluate the variance of $\hat{c}_n$. In the case $\boldsymbol{a} = 0$ where $\rho_{\boldsymbol{a}}(k) = 0$ for every $k \in \mathbb{N}$, we obtain $\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n) = \frac{1}{4n}$ by the same argument as for $\boldsymbol{d} = 0$ and $\boldsymbol{H} = \frac{1}{2}$.

Figure 5.3 displays $\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n)$ for $\boldsymbol{a} \in (-1,1)$ and $n = 10, 11, \ldots, 100$. Remarkably, the variance of $\hat{c}_n$ measured with respect to $\mathbb{P}_{\boldsymbol{a}}$ is equal to the variance of $\hat{c}_n$ measured with respect to $\mathbb{P}_{-\boldsymbol{a}}$ for every $n \in \mathbb{N}$. In particular, one can show that $\gamma_{\boldsymbol{a}}(k) = \gamma_{-\boldsymbol{a}}(k)$ for all $\boldsymbol{a} \in (-1,1)$ and $k \in \mathbb{Z}$. For $k = 0$ and $k = 1$, this is a direct consequence of formulas (4.5) and (4.6). For $k > 1$, Lemma 4.2 (i) yields

$$\begin{aligned}
\gamma_{\boldsymbol{a}}(k) &= \Psi^*(\rho_{\boldsymbol{a}}(1),\, \rho_{\boldsymbol{a}}(k),\, \rho_{\boldsymbol{a}}(k+1),\, \rho_{\boldsymbol{a}}(k-1)) \\
&= \Psi^*(\boldsymbol{a},\, \boldsymbol{a}^k,\, \boldsymbol{a}^{k+1},\, \boldsymbol{a}^{k-1}) \\
&= \Psi^*(-\boldsymbol{a},\, (-\boldsymbol{a})^k,\, (-\boldsymbol{a})^{k+1},\, (-\boldsymbol{a})^{k-1}) \\
&= \Psi^*(\rho_{-\boldsymbol{a}}(1),\, \rho_{-\boldsymbol{a}}(k),\, \rho_{-\boldsymbol{a}}(k+1),\, \rho_{-\boldsymbol{a}}(k-1)) \\
&= \gamma_{-\boldsymbol{a}}(k)\,.
\end{aligned}$$

Since $\gamma_{\boldsymbol{a}}(k) = \gamma_{\boldsymbol{a}}(-k)$, we obtain $\gamma_{\boldsymbol{a}}(k) = \gamma_{-\boldsymbol{a}}(k)$ for all $k \in \mathbb{Z}$.

Note that $\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n)$ decreases to 0 as $\boldsymbol{a}$ tends to -1 and 1, respectively. The reason for this is that the probability of a change tends to 1 as $\boldsymbol{a} \to -1$, and to 0 as $\boldsymbol{a} \to 0$ (see (5.13)), and thus the probability that $\hat{c}_n$ is constant tends to 1.

Finally, let us derive asymptotic properties of $\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n)$. Let $\boldsymbol{a} \in (-1, 1)$. For $k \in \mathbb{N}$, define $f(k) := \boldsymbol{a}^k$. Clearly, $\rho_{\boldsymbol{a}}(k) \sim f(k)$ and $|f(k)| = o(k^{-\beta})$ for any $\beta > \frac{1}{2}$. Thus, according to Theorem 4.18 (i),

$$\mathrm{Var}_{\boldsymbol{a}}(\hat{c}_n) \quad \sim \quad \sigma_{\boldsymbol{a}}^2 \, n^{-1} \,,$$

where $\sigma_{\boldsymbol{a}}^2 := \gamma_{\boldsymbol{a}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{a}}(k) < \infty$. This shows that, in contrast to the previous examples, the variance of $\hat{c}_n$ in AR(1) processes decreases with a rate of n^{-1} for the entire range of parameters.

5.3 Parameter estimates

Estimation of the first-order autocorrelation. Let $\rho(\cdot)$ denote the mapping which assigns to $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ the first-order autocorrelation of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$, that is,

$$\rho(\boldsymbol{\vartheta}) \quad := \quad \rho_{\boldsymbol{\vartheta}}(1) \tag{5.14}$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. According to equation (5.6), we have

$$c(\cdot) \quad = \quad \frac{1}{2} - \frac{1}{\pi} \arcsin \rho(\cdot) \,. \tag{5.15}$$

Thus, $\rho(\cdot)$ and $c(\cdot)$ are monotonically related: the larger the first-order autocorrelation of $\mathbf{Y}$, the lower the probability of a change. Using the fact that $\sin(\pi(\frac{1}{2} - x)) = \cos(\pi x)$ for all $x \in \mathbb{R}$, we can write equation (5.15) as

$$\rho(\cdot) \quad = \quad \cos(\pi c(\cdot)) \,. \tag{5.16}$$

Plugging the estimate $\hat{c}_n$ of $c(\cdot)$ into the right hand side of (5.16), we obtain

$$\hat{\rho}_n \quad := \quad \cos(\pi \hat{c}_n) \tag{5.17}$$

as an estimate of $\rho(\cdot)$. By the model assumption that $\mathbf{Y}$ is non-degenerate for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, we have $c(\Theta) \subset (0, 1)$. Furthermore, the mapping $x \mapsto \cos(\pi x)$ is bounded, continuous and has a non-vanishing first derivative at x for every $x \in (0, 1)$. Therefore, according to Corollary 5.2, if $\lim_{k\to\infty} \rho_{\boldsymbol{\vartheta}}(k) = 0$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, then $\hat{\rho}_n$ is a strongly consistent and asymptotically unbiased estimator of $\rho(\cdot)$. If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$, then $\hat{\rho}_n$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ is asymptotically normally distributed.

Note that the estimator $\hat{\rho}_n$ of $\rho(\cdot)$ has been known for more than half a century. The idea to use empirical zero crossing rates for estimating second-order properties of stationary processes (such as autocorrelations, dominant frequencies in the spectral domain, etc.) has attracted considerable attention in engineering, e.g., for speech recognition or the analysis of vibrations. See Kedem [52, 53] and Section 1.1 in the introduction.

Let us compare $\hat{\rho}_n$ to the sample autocorrelation, given by

$$\tilde{\rho}_n \quad := \quad \frac{\sum_{t=1}^{n-1}(Y_t - \bar{Y}_n)(Y_{t+1} - \bar{Y}_n)}{\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2}\,, \tag{5.18}$$

where $\bar{Y}_n := \frac{1}{n}\sum_{t=1}^{n} Y_t$ is the sample mean. Under the assumption that $\mathbf{Y}$ is stationary Gaussian, a sufficient condition for $\tilde{\rho}_n$ to be asymptotically normally distributed is that $\rho_{\boldsymbol{\vartheta}}(k) = O(k^{-\beta})$ for some $\beta \geq \frac{1}{2}$ (see Hosking [50]). Note that in the case $\beta = \frac{1}{2}$, the rate of convergence is $\sqrt{n}/\sqrt{\ln n}$. If there exist $\alpha > 0$ and $0 < \beta < \frac{1}{2}$ such that $\rho_{\boldsymbol{\vartheta}}(k) \sim \alpha k^{-\beta}$, then the limit distribution is a Modified Rosenblatt distribution.

Note that $\tilde{\rho}_n$ does not require $\mathbf{Y}$ to be Gaussian, whereas $\hat{\rho}_n$ is invariant with respect to monotonic transformations of $\mathbf{X}$. In Section 5.4, we compare the performance of $\tilde{\rho}_n$ and $\hat{\rho}_n$ for the estimation of the first-order autocorrelation in Fractional Gaussian Noise. It turns out that $\tilde{\rho}_n$ has a smaller variance but larger bias than $\hat{\rho}_n$.

Estimation of real-valued $\boldsymbol{\vartheta}$. Suppose that $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}$, and there exists a function $h : [0,1] \to \mathbb{R}$ with

$$h(c(\boldsymbol{\vartheta})) \quad = \quad \boldsymbol{\vartheta} \tag{5.19}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Clearly, a sufficient condition for the existence of such a function h is that $\boldsymbol{\vartheta} \mapsto c(\boldsymbol{\vartheta})$ is strictly monotone on $\boldsymbol{\Theta}$ which, according to (5.15), is equivalent to $\boldsymbol{\vartheta} \mapsto \rho(\boldsymbol{\vartheta})$ being strictly monotone. Plugging the estimate $\hat{c}_n$ of $c(\cdot)$ into the left hand side of (5.19), we obtain

$$\hat{\boldsymbol{\vartheta}}_n \quad := \quad h(\hat{c}_n)$$

as an estimate of $\boldsymbol{\vartheta}$. Properties of this estimator can be derived by Corollary 5.2.

Corollary 5.3.

(*i*) *If $\rho_{\boldsymbol{\vartheta}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and h is continuous on an open set containing $c(\boldsymbol{\Theta})$, then $\hat{\boldsymbol{\vartheta}}_n$ is a strongly consistent estimator of $\boldsymbol{\vartheta}$. If, additionally, h is bounded, then $\hat{\boldsymbol{\vartheta}}_n$ is an asymptotically unbiased estimator of $\boldsymbol{\vartheta}$.*

(*ii*) *If $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ and h has a non-vanishing first derivative at $c(\boldsymbol{\vartheta})$, then*

$$\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{\vartheta}}[h'(c(\boldsymbol{\vartheta}))]^2\,)\,,$$

with $\sigma^2_{\boldsymbol{\vartheta}}$ as given in Corollary 5.2 (iii).

Confidence intervals. If $\hat{\boldsymbol{\vartheta}}_n$ is asymptotically normal and the limit distribution is non-degenerate, that is, $\sigma^2_{\boldsymbol{\vartheta}}$ given in Corollary 5.2 (iii) is strictly positive, then we can derive confidence intervals for $\boldsymbol{\vartheta}$. As in Section 4.6, let $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ be given by

$$\boldsymbol{\Theta}_0 \;:=\; \Big\{\, \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \;\Big|\; \text{There exist a function } f : \mathbb{N} \to \mathbb{R} \text{ and some } \beta > \tfrac{1}{2} \text{ with } |\rho_{\boldsymbol{\vartheta}}(k)| \sim f(k) \text{ and } f(k) = o(k^{-\beta}) \,\Big\}.$$

Note that $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0$ implies that $\sigma^2_{\boldsymbol{\vartheta}}$ is well-defined and finite (see Section 4.6). Now, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0$ be such that $\sigma^2_{\boldsymbol{\vartheta}} > 0$, and suppose h has a non-vanishing first derivative at $c(\boldsymbol{\vartheta})$. Clearly,

$$s^2_{\boldsymbol{\vartheta}} \;:=\; \sigma^2_{\boldsymbol{\vartheta}}\,[h'(c(\boldsymbol{\vartheta}))]^2$$

is strictly greater than zero in this case. Thus, according to Corollary 5.3 (ii),

$$\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta})\,/\,s_{\boldsymbol{\vartheta}} \;\xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}}\; \mathrm{N}(0,1)\,. \tag{5.20}$$

In the following, let Φ^{-1} denote the quantile function of the standard normal distribution. It is well-known that if Z is a standard normal random variable on some probability space $(\Omega', \mathcal{A}', \mathbb{P})$, then $\mathbb{P}(Z \in (-\infty,\, \Phi^{-1}(\alpha)\,]\,) = \alpha$ for every $\alpha \in (0,1)$. Furthermore, $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$ for every $\alpha \in (0,1)$.

For $n \in \mathbb{N}$ and $\alpha \in (0,1)$, let the random interval $K_n(\alpha, \boldsymbol{\vartheta}) \subset \mathbb{R}$ be defined by

$$K_n(\alpha,\boldsymbol{\vartheta}) \;:=\; \big[\; \hat{\boldsymbol{\vartheta}}_n \,-\, \Phi^{-1}(1-\alpha/2)\; s_{\boldsymbol{\vartheta}}/\sqrt{n}\,,\;\; \hat{\boldsymbol{\vartheta}}_n \,+\, \Phi^{-1}(1-\alpha/2)\; s_{\boldsymbol{\vartheta}}/\sqrt{n}\;\big]\,.$$

Note that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\vartheta}}(\,\boldsymbol{\vartheta} \in K_n(\alpha,\boldsymbol{\vartheta})\,) \;&=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\,\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta})/s_{\boldsymbol{\vartheta}} \;\in\; [\,-\Phi^{-1}(1-\alpha/2),\; \Phi^{-1}(1-\alpha/2)\,]\,) \\
&=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\,\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta})/s_{\boldsymbol{\vartheta}} \;\in\; (-\infty,\; \Phi^{-1}(1-\alpha/2)\,]\,) \\
&\qquad - \mathbb{P}_{\boldsymbol{\vartheta}}(\,\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta})/s_{\boldsymbol{\vartheta}} \;\in\; (-\infty,\; \Phi^{-1}(\alpha/2)\,]\,)\,.
\end{aligned} \tag{5.21}$$

Since $\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta})/s_{\boldsymbol{\vartheta}}$ converges in distribution to $\mathrm{N}(0,1)$ (see (5.20)) and the boundaries of $(-\infty,\, \Phi^{-1}(1-\alpha/2)\,]$ and $(-\infty,\, \Phi^{-1}(\alpha/2)\,]$ in $\mathbb{R}$ both have Lebesgue-measure zero, the probabilities on the right hand side of (5.21) converge to $1-\alpha/2$ and $\alpha/2$, respectively, and thus

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\vartheta}}(\,\boldsymbol{\vartheta} \in K_n(\alpha,\boldsymbol{\vartheta})\,) \;=\; 1-\alpha\,. \tag{5.22}$$

This shows that $K_n(\alpha,\boldsymbol{\vartheta})$ is an asymptotic $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\vartheta}$.

In practice, the confidence interval $K_n(\alpha,\boldsymbol{\vartheta})$ is not very useful because its computation requires the knowledge of $s^2_{\boldsymbol{\vartheta}}$ and hence of $\boldsymbol{\vartheta}$. However, if we know the true value of $\boldsymbol{\vartheta}$, then we do not need a confidence interval for it.

Next, we study sufficient conditions under which $s^2_{\boldsymbol{\vartheta}}$ can be replaced by an estimate, the computation of which does not require the knowledge of $\boldsymbol{\vartheta}$. Define

$$\hat{\sigma}^2_n \;:=\; n \cdot \mathrm{Var}_{\hat{\boldsymbol{\vartheta}}_n}(\hat{c}_n) \tag{5.23}$$

and

$$\hat{s}^2_n \;:=\; (h'(\hat{c}_n))^2\, \hat{\sigma}^2_n\,. \tag{5.24}$$

The following theorem establishes sufficient conditions for $s^2_{\boldsymbol{\vartheta}}$ to be strictly positive and for $\hat{s}^2_n$ to be a consistent estimator of $s^2_{\boldsymbol{\vartheta}}$.

Theorem 5.4. *Let $\boldsymbol{\Theta}_0^+$ be a subset of $\boldsymbol{\Theta}_0$ satisfying one of the following conditions:*

(*i*) *For every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$,*

(*a*) $1 + \rho_{\boldsymbol{\vartheta}}(1) + 2\,\rho_{\boldsymbol{\vartheta}}(3) > 0$,

(*b*) $\rho_{\boldsymbol{\vartheta}}(k) < \rho_{\boldsymbol{\vartheta}}(k+1) < 0$ *for all* $k \in \mathbb{N}$.

Furthermore, if $I \subset \boldsymbol{\Theta}_0^+$ is non-empty and compact, then there exist a $k_0 \in \mathbb{N}$ and a function $\nu : \mathbb{N} \to \mathbb{R}$ with $\nu(1) > -1$ and $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ such that, for all $\boldsymbol{\vartheta} \in I$ and $k \geq k_0$, $\nu(1) \leq \rho_{\boldsymbol{\vartheta}}(1)$ and $\nu(k) \leq \rho_{\boldsymbol{\vartheta}}(k)$.

(*ii*) *For every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$, there exists some $\boldsymbol{a} \in (-1, 1)$ such that $\rho_{\boldsymbol{\vartheta}}(k) = \boldsymbol{a}^k$ for all $k \in \mathbb{N}$.*

Then $\sigma^2_{\boldsymbol{\vartheta}} > 0$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$. If, additionally, $\boldsymbol{\Theta}_0^+$ is an open set in $\mathbb{R}$, the mapping $\boldsymbol{\vartheta} \mapsto \rho_{\boldsymbol{\vartheta}}(k)$ is continuous on $\boldsymbol{\Theta}_0^+$ for every $k \in \mathbb{N}$, and h' exists and is continuous on an open set containing $c(\boldsymbol{\Theta}_0^+)$, then

$$\lim_{n\to\infty} \hat{s}^2_n \;=\; s^2_{\boldsymbol{\vartheta}}$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$.

Proof. The statement $\sigma^2_{\boldsymbol{\vartheta}} > 0$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$ is an immediate consequence of Theorem 4.22. Furthermore, Theorem 4.24 shows that if $I \subset \boldsymbol{\Theta}_0^+$ is compact, then $\boldsymbol{\vartheta} \mapsto n \cdot \mathrm{Var}_{\boldsymbol{\vartheta}}(\hat{c}_n)$ converges to $\boldsymbol{\vartheta} \mapsto \sigma^2_{\boldsymbol{\vartheta}}$ uniformly on I as $n \to \infty$, and $\boldsymbol{\vartheta} \mapsto \sigma^2_{\boldsymbol{\vartheta}}$ is continuous on I. Now, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^+$. The existence of h' implies that h is continuous on an open set containing $c(\boldsymbol{\Theta}_0^+)$, and thus $\lim_{n\to\infty} \hat{\boldsymbol{\vartheta}}_n = \boldsymbol{\vartheta}$ $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely (see Corollary 5.3 (*i*)). Consequently, there exist $n_0 \in \mathbb{N}$ and $\delta > 0$ such that $[\boldsymbol{\vartheta} - \delta, \boldsymbol{\vartheta} + \delta] \subset \boldsymbol{\Theta}_0^+$ and $\hat{\boldsymbol{\vartheta}}_n \in [\boldsymbol{\vartheta} - \delta, \boldsymbol{\vartheta} + \delta]$ for all $n \geq n_0$ $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Since $[\boldsymbol{\vartheta} - \delta, \boldsymbol{\vartheta} + \delta]$ is compact, Theorem 4.24 yields

$$\lim_{n\to\infty} \hat{\sigma}^2_n \;=\; \sigma^2_{\boldsymbol{\vartheta}}$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Furthermore, according to Corollary 5.2 (ii),

$$\lim_{n\to\infty}[h'(\hat{c}_n)]^2 \;=\; [h'(c(\boldsymbol{\vartheta}))]^2$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely. Altogether, $\lim_{n\to\infty}\hat{s}_n^2 = s_{\boldsymbol{\vartheta}}^2$ $\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_0^+$, and hence the proof is complete. □

Suppose there exists a non-empty subset $\boldsymbol{\Theta}_0^+$ of $\boldsymbol{\Theta}_0$ such that the assumptions of Theorem 5.4 are satisfied. Let $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_0^+$. According to Theorem 2.3.3 in Lehmann [68], replacing $s_{\boldsymbol{\vartheta}}^2$ in (5.20) by the consistent estimate $\hat{s}_n^2$ has no effect on the limiting distribution. Thus,

$$\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta})\,/\,\hat{s}_n^2 \;\xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}}\; \mathrm{N}(0,1)\,.$$

Now, for $n\in\mathbb{N}$ and $\alpha\in(0,1)$, define

$$K_n(\alpha) \;:=\; \left[\,\hat{\boldsymbol{\vartheta}}_n \,-\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\,,\;\; \hat{\boldsymbol{\vartheta}}_n \,+\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\,\right].$$

Analogously to (5.22), we obtain that $K_n(\alpha)$ is an asymptotic $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\vartheta}$. We formally state this fact in the following corollary.

Corollary 5.5. *Suppose $\boldsymbol{\Theta}_0^+$ is a subset of $\boldsymbol{\Theta}_0$ satisfying condition (i) or (ii) in Theorem 5.4. Furthermore, suppose $\boldsymbol{\Theta}_0^+$ is an open set in $\mathbb{R}$, the mapping $\boldsymbol{\vartheta}\mapsto\rho_{\boldsymbol{\vartheta}}(k)$ is continuous on $\boldsymbol{\Theta}_0^+$ for every $k\in\mathbb{N}$, and h' exists and is continuous on an open set containing $c(\boldsymbol{\Theta}_0^+)$. Then, for all $\alpha\in(0,1)$ and $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_0^+$,*

$$\lim_{n\to\infty}\mathbb{P}_{\boldsymbol{\vartheta}}(\,\boldsymbol{\vartheta}\in K_n(\alpha)\,) \;=\; 1-\alpha\,.$$

The main difficulty in determining the confidence interval $K_n(\alpha)$ is the evaluation of $\mathrm{Var}_{\hat{\boldsymbol{\vartheta}}_n}(\hat{c}_n)$, which is required for the computation of $\hat{s}_n^2$ (see (5.23) and (5.24)). When n is large, an exact numerical evaluation can be time-consuming (see the discussion in Section 4.5). In order to speed-up computations, the variance of $\hat{c}_n$ for certain parameter values can be stored in a look-up table. If the outcome of $\hat{\boldsymbol{\vartheta}}_n$ is not among the parameter values in the table, the corresponding value of the variance can be approximated by interpolation.

Clearly, in practice, the confidence intervals obtained according to Corollary 5.5 should be interpreted with caution. Firstly, even when $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_0^+$ and n is large, the probability that $K_n(\alpha)$ covers $\boldsymbol{\vartheta}$ can be much lower than $1-\alpha$. Secondly, if we cannot exclude a priori that $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^+$, then the realization of $K_n(\alpha)$ basically does not allow to draw any conclusion on the true value of $\boldsymbol{\vartheta}$.

In Section 5.4, we provide simulations of the accuracy of confidence intervals for the Hurst parameter in FBM, for the fractional differencing parameter in ARFIMA(0,d,0) processes and for the autoregressive coefficient in AR(1) processes. It turns out that the coverage of these parameters is about 95% also when n is small and not all assumptions of Corollary 5.5 are satisfied.

Estimating the Hurst parameter. Let $h : [0,1] \to \mathbb{R}$ be defined by

$$h(x) := \max\left\{0,\ \log_2(\cos(\pi x/2)) + 1\right\} \tag{5.25}$$

for $x \in [0,1]$. Note that h is the maximum of two continuous functions and thus continuous on $[0,1]$. Furthermore, $h(x) = \log_2(\cos(\pi x/2)) + 1$ for $x \in [0, \frac{2}{3}]$, and the first derivative of h on $(0, \frac{2}{3})$ is given by

$$h'(x) = -\frac{\pi}{2\ln 2}\tan(\pi x/2) \tag{5.26}$$

for $x \in (0, \frac{2}{3})$. According to the relation between the probability of a change $c(\boldsymbol{H})$ and the Hurst parameter $\boldsymbol{H}$ established in (5.10), we have

$$h(c(\boldsymbol{H})) = \boldsymbol{H} \tag{5.27}$$

for $\boldsymbol{H} \in (0,1)$. Plugging the estimate $\hat{c}_n$ of $c(\cdot)$ into the left hand side of (5.27), we obtain

$$\hat{\boldsymbol{H}}_n := h(\hat{c}_n)$$

as an estimate of $\boldsymbol{H}$. Properties of $\hat{\boldsymbol{H}}_n$ and confidence intervals for $\boldsymbol{H}$ are established by the following corollary.

Note that the asymptotic properties of $\hat{\boldsymbol{H}}_n$ are the same if we replace h in the definition of $\hat{\boldsymbol{H}}_n$ with the mapping $\tilde{h} : [0,1] \to [-\infty, \infty]$ given by $\tilde{h}(x) := \log_2(\cos(\pi x/2)) + 1$ for $x \in [0,1]$. For finite sample sizes, however, the estimates of $\boldsymbol{H}$ might be negative or even equal to negative infinity.

Corollary 5.6. *The estimator $\hat{\boldsymbol{H}}_n$ has the following properties:*

(*i*) *$\hat{\boldsymbol{H}}_n$ is a strongly consistent and asymptotically unbiased estimator of $\boldsymbol{H}$.*

(*ii*) *If $\boldsymbol{H} < \frac{3}{4}$, then*

$$\sqrt{n}\,(\hat{\boldsymbol{H}}_n - \boldsymbol{H}) \xrightarrow{\mathbb{P}_{\boldsymbol{H}}} \mathrm{N}(0,\, \sigma^2_{\boldsymbol{H}}\,[h'(c(\boldsymbol{H}))]^2)\,,$$

where h' is given by (5.26) and

$$\sigma^2_{\boldsymbol{H}} := \gamma_{\boldsymbol{H}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{H}}(k)$$

with $\gamma_{\boldsymbol{H}}(k) := \mathrm{Cov}_{\boldsymbol{H}}(C(0), C(k))$ for $k \in \mathbb{Z}$.

(*iii*) *Let* $\alpha \in (0,1)$. *If* $\boldsymbol{H} < \frac{1}{2}$, *then*

$$K_n(\alpha) \;:=\; \left[\, \hat{\boldsymbol{H}}_n \,-\, \Phi^{-1}(1-\alpha/2)\; \hat{s}_n/\sqrt{n}\,,\quad \hat{\boldsymbol{H}}_n \,+\, \Phi^{-1}(1-\alpha/2)\; \hat{s}_n/\sqrt{n}\,\right]$$

with $\hat{s}_n$ *given by*

$$\hat{s}_n^2 \;:=\; (h'(\hat{c}_n))^2 \,\cdot n \cdot \mathrm{Var}_{\hat{\boldsymbol{H}}_n}(\hat{c}_n)$$

satisfies

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{H}}(\, \boldsymbol{H} \in K_n(\alpha)\,) \;=\; 1-\alpha\,.$$

Proof. (*i*) According to (5.10), the image of $(0,1)$ under $c(\cdot)$ is given by $(0,\frac{2}{3})$. Furthermore, h is continuous and hence bounded on $[0,1]$ (see (5.25)). Thus, the statement is an immediate consequence of Corollary 5.3 (*i*) and the fact that $\rho_{\boldsymbol{H}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{H} \in (0,1)$ (see Lemma 2.7 (*ii*)).

(*ii*) If $\boldsymbol{H} < \frac{3}{4}$, then there exists a $\beta > \frac{1}{2}$ such that $|\rho_{\boldsymbol{H}}(k)| = o(k^{-\beta})$ (for instance, we can choose $\beta = \frac{5}{4} - \boldsymbol{H}$, see Lemma 2.7 (*ii*)). Since h' is non-vanishing on $(0,\frac{2}{3})$, the statement follows by Corollary 5.3 (*ii*).

(*iii*) We show that the conditions of Corollary 5.5 are satisfied. Clearly, $(0,\frac{1}{2})$ is an open set in $\mathbb{R}$ with the image $(0,\frac{1}{2})$ under $c(\cdot)$, and h' is continuous on $(0,\frac{1}{2})$. According to the definition of FGN, the mapping $\boldsymbol{H} \mapsto \rho_{\boldsymbol{H}}(k)$ is continuous on $(0,\frac{1}{2})$ for every $k \in \mathbb{N}$ (see Section 2.2.3). Moreover, $\rho_{\boldsymbol{H}}(k) < \rho_{\boldsymbol{H}}(k+1) < 0$ for all $k \in \mathbb{N}$ and $\boldsymbol{H} \in (0,\frac{1}{2})$ (see Lemma 2.7 (*iii*)). By Lemma 2.7 (*iv*), we obtain that $\rho_{\boldsymbol{H}}(k) > -\frac{1}{2k}$ for all $\boldsymbol{H} \in (0,\frac{1}{2})$ and $k \in \mathbb{N}$, which shows that

$$1 + \rho_{\boldsymbol{H}}(1) + 2\,\rho_{\boldsymbol{H}}(3) \;>\; 2^{2\boldsymbol{H}-1} - \frac{1}{3} \;>\; \frac{1}{2} - \frac{1}{3} \;>\; 0$$

for every $\boldsymbol{H} \in (0,\frac{1}{2})$. Now, let $I \subset (0,\frac{1}{2})$ be compact. According to Lemma 2.7 (*iv*), there exists a function $\nu : \mathbb{N} \to \mathbb{R}$ with $\nu(1) > -1$ and $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ such that $\nu(k) \le \rho_{\boldsymbol{H}}(k)$ for all $k \in \mathbb{N}$ and $\boldsymbol{H} \in I$. The proof is complete. □

Let $\alpha \in (0,1)$. Since $\boldsymbol{H}$ is ranging in $(0,1)$, we have

$$\mathbb{P}_{\boldsymbol{H}}(\, \boldsymbol{H} \in K_n(\alpha)\,) \;=\; \mathbb{P}_{\boldsymbol{H}}(\, \boldsymbol{H} \in K_n(\alpha) \cap (0,1)\,)$$

for all $\boldsymbol{H} \in (0,1)$ and $n \in \mathbb{N}$. Therefore, $K_n(\alpha)$ in Corollary 5.6 (*iii*) can be replaced with $K_n(\alpha) \cap (0,1)$. Note that we may replace $K_n(\alpha)$ also with $K_n(\alpha) \cap (0,\frac{1}{2})$. In practice, however, when we cannot assert a priori that $\boldsymbol{H} \in (0,\frac{1}{2})$, it is preferable to take the confidence interval $K_n(\alpha) \cap (0,1)$ because it may cover $\boldsymbol{H}$ with a probability close to $1-\alpha$ also in the case $\boldsymbol{H} \notin (0,\frac{1}{2})$.

Indeed, the simulations in Section 5.4 indicate that the conclusion of Corollary 5.6 (*iii*) is valid also for $\boldsymbol{H} \in [\frac{1}{2}, \frac{3}{4})$ and thus for the whole range of parameters where $\hat{\boldsymbol{H}}_n$ is asymptotically normally distributed (see Corollary 5.6 (*ii*)). For $\boldsymbol{H} \geq \frac{3}{4}$, the simulations suggest that $\mathbb{P}_{\boldsymbol{H}}(\, \boldsymbol{H} \in K_n(\alpha) \cap (0,1)\,)$ still converges to $1-\alpha$ (or even to larger values), however, with a lower speed of convergence.

Note that $\hat{\boldsymbol{H}}_n$ has been known for some time as the *Zero Crossings* (*ZC*) *estimator* of the Hurst parameter (according to the equivalence of zero crossings in $\mathbf{Y}$ and changes in $\mathbf{X}$ discussed in Section 5.1). Coeurjolly [28] resumes properties of $\hat{\boldsymbol{H}}_n$ such as strong consistency and asymptotic normality for $\boldsymbol{H} < \frac{3}{4}$. Marković and Koch [73] investigates the robustness of $\hat{\boldsymbol{H}}_n$ in a simulation study and shows the application to the analysis of hydrological time series. Shi et al. [90] applies $\hat{\boldsymbol{H}}_n$ to the analysis of atmospheric turbulence data.

Alternative estimation methods. There are various alternative methods for estimating the Hurst parameter. Some of these methods can be more generally applied, for example, to the estimation of the index of selfsimilarity in selfsimilar process, or to the estimation of parameters in arbitrary parametric Gaussian models. For an overview, we refer to Beran [15] and Doukhan et al. [35]. Taqqu et al. [97] compares estimation methods in a simulation study.

Here, we briefly mention some of the available methods. An important class of estimators is given by Maximum Likelihood (ML) estimates and approximations thereof, such as (approximate) Whittle estimates. These methods are well understood theoretically and known to have asymptotic optimality properties. In practice, a drawback of ML methods is that they are computationally intensive and thus time-consuming.

Another class of estimators is given by semi-parametric methods. These include, for instance, estimates based on the rescaled range (R/S) statistic, on the periodogram (such as the Local Whittle Estimators, see Robinson [85]), or on coefficients of wavelet decompositions (see Abry and Veitch [1]). Typically, semi-parametric methods do not require a complete specification of the model distribution, which makes the methods more robust and generally applicable, but usually also less efficient. A disadvantage of semi-parametric methods is that the estimates depend on certain tuning parameters which are difficult to select automatically.

A simple estimator for the Hurst parameter has been proposed in Kettani and Gubner [63]. For $x \in [0,1]$, define

$$g(x) \;:=\; \max\left\{0,\; \frac{1}{2}(\log_2(1+x)+1)\right\}. \tag{5.28}$$

According to (5.9), we have

$$g(\rho(\boldsymbol{H})) \;=\; \boldsymbol{H} \tag{5.29}$$

for every $\boldsymbol{H} \in (0,1)$. The idea of Kettani and Gubner [63] is to plug the sample autocorrelation $\tilde{\rho}_n$ (see (5.18)) into the left hand side of (5.29), which yields the estimator

$$\tilde{\boldsymbol{H}}_n \;:=\; g(\tilde{\rho}_n)\,.$$

Note that the ZC estimator $\hat{\boldsymbol{H}}_n$ can be written as

$$\hat{\boldsymbol{H}}_n \;=\; g(\hat{\rho}_n)$$

where $\hat{\rho}_n$ is the estimator of $\rho(\cdot)$ given in (5.17). Thus, $\tilde{\boldsymbol{H}}_n$ differs from $\hat{\boldsymbol{H}}_n$ only by using a different estimator for $\rho(\cdot)$.

In Section 5.4, we compare the performance of $\hat{\boldsymbol{H}}_n$ and $\tilde{\boldsymbol{H}}_n$ in a simulation study. Similar to the results for $\hat{\rho}_n$ and $\tilde{\rho}_n$, the estimator $\hat{\boldsymbol{H}}_n$ has a larger variance but smaller bias than $\tilde{\boldsymbol{H}}_n$. Note that the estimator $\tilde{\boldsymbol{H}}_n$ can be more generally applied to the estimation of the index of self-similarity in (not necessarily Gaussian) selfsimilar processes with stationary increments. In contrast to $\hat{\boldsymbol{H}}_n$, however, $\tilde{\boldsymbol{H}}_n$ is not invariant with respect to monotone transformations of the process.

Estimating the fractional differencing parameter. Let $h : [0,1] \to \mathbb{R}$ be defined by

$$h(x) \;:=\; \max\big\{0,\, \frac{\cos(\pi x)}{1+\cos(\pi x)}\big\} \tag{5.30}$$

for $x \in [0,1]$. Clearly, h is the maximum of two continuous mappings and thus continuous on $[0,1]$. Note that $h(x) = \frac{\cos(\pi x)}{1+\cos(\pi x)}$ if $x \le \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{1}{3}$, and the first derivative of h on $(0, \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{1}{3})$ is given by

$$h'(x) \;=\; -\frac{\pi \sin(\pi x)}{(1+\cos(\pi x))^2} \tag{5.31}$$

for $x \in (0, \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{1}{3})$. According to the relation between the fractional differencing parameter $\boldsymbol{d}$ and the probability of a change $c(\boldsymbol{d})$ established in (5.12), we have

$$h(c(\boldsymbol{d})) \;=\; \boldsymbol{d} \tag{5.32}$$

for $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$. Plugging the estimate $\hat{c}_n$ of $c(\cdot)$ into the left hand side of (5.32) yields

$$\hat{\boldsymbol{d}}_n \;:=\; h(\hat{c}_n)$$

as an estimate of $\boldsymbol{d}$. Properties of $\hat{\boldsymbol{d}}_n$ and confidence intervals for $\boldsymbol{d}$ are given in the following corollary.

Corollary 5.7. *The estimator $\hat{\boldsymbol{d}}_n$ has the following properties:*

(*i*) *$\hat{\boldsymbol{d}}_n$ is a strongly consistent and asymptotically unbiased estimator of $\boldsymbol{d}$.*

(*ii*) *If $\boldsymbol{d} < \frac{1}{4}$, then*

$$\sqrt{n}\,(\hat{\boldsymbol{d}}_n - \boldsymbol{d}) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{d}}} \quad \mathrm{N}(0,\, \sigma_{\boldsymbol{d}}^2\,[h'(c(\boldsymbol{d}))]^2)\,,$$

where h' is given by (5.31) and

$$\sigma_{\boldsymbol{d}}^2 \;:=\; \gamma_{\boldsymbol{d}}(0) + 2\sum_{k=1}^{\infty}\gamma_{\boldsymbol{d}}(k)$$

with $\gamma_{\boldsymbol{d}}(k) := \mathrm{Cov}_{\boldsymbol{d}}(C(0), C(k))$ for $k \in \mathbb{Z}$.

(*iii*) *Let $\alpha \in (0,1)$. If $\boldsymbol{d} < 0$, then*

$$K_n(\alpha) \;:=\; \left[\; \hat{\boldsymbol{d}}_n \,-\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\,,\quad \hat{\boldsymbol{d}}_n \,+\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\;\right]$$

with $\hat{s}_n$ given by

$$\hat{s}_n^2 \;:=\; (h'(\hat{c}_n))^2 \cdot n \cdot \mathrm{Var}_{\hat{\boldsymbol{d}}_n}(\hat{c}_n)$$

satisfies

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{d}}(\,\boldsymbol{d} \in K_n(\alpha)\,) \;=\; 1-\alpha\,.$$

Proof. (*i*) According to (5.12), the image of $(0,1)$ under $c(\cdot)$ is given by $(0, \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{1}{3})$. Furthermore, h is continuous and thus bounded on $[0,1]$. Therefore, the statement is an immediate consequence of Corollary 5.3 (*i*) and the fact that $\rho_{\boldsymbol{d}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$ (see Lemma 2.8 (*i*)).

(*ii*) If $\boldsymbol{d} < \frac{1}{4}$, then there exists a $\beta > \frac{1}{2}$ with $|\rho_{\boldsymbol{d}}(k)| = o(k^{-\beta})$ (for instance, we may choose $\beta = \frac{3}{4} - \boldsymbol{d}$, see Lemma 2.8 (*i*)). Since h' is non-vanishing on $(0, \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{1}{3})$, the statement follows by Corollary 5.3 (*ii*).

(*iii*) We show that the conditions of Corollary 5.5 are satisfied. Clearly, $(-\frac{1}{2}, 0)$ is an open set in $\mathbb{R}$ with the image $(0, \frac{1}{2})$ under $c(\cdot)$, and h' is continuous on $(0, \frac{1}{2})$. According to the definition of ARFIMA(0,d,0) processes, the mapping $\boldsymbol{d} \mapsto \rho_{\boldsymbol{d}}(k)$ is continuous on $(-\frac{1}{2}, 0)$ for every $k \in \mathbb{N}$ (see Section 2.2.4). Moreover, $\rho_{\boldsymbol{d}}(k) < \rho_{\boldsymbol{d}}(k+1) < 0$ for all $k \in \mathbb{N}$ and $\boldsymbol{d} \in (-\frac{1}{2}, 0)$ (see Lemma 2.8 (*ii*)). In particular,

$$1 + \rho_{\boldsymbol{d}}(1) + 2\,\rho_{\boldsymbol{d}}(3) \;>\; 1 + 3\,\rho_{\boldsymbol{d}}(1) \;=\; \frac{1+2\,\boldsymbol{d}}{1-\boldsymbol{d}} \;>\; 0$$

for every $\boldsymbol{d} \in (-\frac{1}{2}, 0)$. Now, let $I \subset (-\frac{1}{2}, 0)$ be compact. According to Lemma 2.8 (*iii*), there exists a function $\nu : \mathbb{N} \to \mathbb{R}$ with $\nu(1) > -1$ and $|\nu(k)| = o(k^{-\beta})$ for some $\beta > \frac{1}{2}$ such that $\nu(k) \le \rho_{\boldsymbol{d}}(k)$ for all $k \in \mathbb{N}$ and $\boldsymbol{d} \in I$. The proof is complete. □

Similar to the remark after Corollary 5.6, we may replace $K_n(\alpha)$ in Corollary 5.7 (*iii*) by the confidence interval $K_n(\alpha) \cap (-\frac{1}{2}, \frac{1}{2})$. Note that the simulations in Section 5.4 suggest that the conclusion of Corollary 5.7 (*iii*) is true for all $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{4})$, which corresponds to the whole range of parameters where $\hat{\boldsymbol{d}}_n$ is asymptotically normally distributed (see Corollary 5.7 (*ii*)).

Estimating the autoregressive coefficient. Consider the function $h : [0,1] \to \mathbb{R}$ defined by

$$h(x) \;:=\; \cos(\pi x) \tag{5.33}$$

for $x \in [0,1]$. Clearly, h is continuous on $[0,1]$ and has the first derivative

$$h'(x) \;=\; -\pi \sin(\pi x) \tag{5.34}$$

for $x \in (0,1)$. According to the relation between the autoregressive coefficient $\boldsymbol{a}$ and the probability of a change $c(\boldsymbol{a})$ established in (5.13), we have

$$h(c(\boldsymbol{a})) \;=\; \boldsymbol{a} \tag{5.35}$$

for $\boldsymbol{a} \in (-1,1)$. By plugging the estimate $\hat{c}_n$ of $c(\cdot)$ into the left hand side of (5.35), we obtain

$$\hat{\boldsymbol{a}}_n \;:=\; h(\hat{c}_n)$$

as an estimate of $\boldsymbol{a}$. Note that $\hat{\boldsymbol{a}}_n$ is equal to the estimator $\hat{\rho}_n$ of the first-order autocorrelation $\rho(\boldsymbol{a}) = \boldsymbol{a}$. The following corollary establishes properties of $\hat{\boldsymbol{a}}_n$ and confidence intervals for the autoregressive coefficient.

Corollary 5.8. *The estimator $\hat{\boldsymbol{a}}_n$ has the following properties:*

(*i*) *$\hat{\boldsymbol{a}}_n$ is a strongly consistent and asymptotically unbiased estimator of $\boldsymbol{a}$.*

(*ii*) *For every $\boldsymbol{a} \in (-1,1)$,*

$$\sqrt{n}\,(\hat{\boldsymbol{a}}_n - \boldsymbol{a}) \;\xrightarrow{\mathbb{P}_{\boldsymbol{a}}}\; \mathrm{N}(0,\, \sigma_{\boldsymbol{a}}^2\,[h'(c(\boldsymbol{a}))]^2)\,,$$

where h' is given by (5.34) and

$$\sigma_{\boldsymbol{a}}^2 \;:=\; \gamma_{\boldsymbol{a}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{a}}(k)$$

with $\gamma_{\boldsymbol{a}}(k) := \mathrm{Cov}_{\boldsymbol{a}}(C(0), C(k))$ for $k \in \mathbb{Z}$.

(*iii*) *Let* $\alpha \in (0,1)$. *For every* $\boldsymbol{a} \in (-1,1)$,

$$K_n(\alpha) \;:=\; \left[\, \hat{\boldsymbol{a}}_n \,-\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\,,\;\; \hat{\boldsymbol{a}}_n \,+\, \Phi^{-1}(1-\alpha/2)\;\hat{s}_n/\sqrt{n}\,\right]$$

with $\hat{s}_n$ *given by*

$$\hat{s}_n^2 \;:=\; (h'(\hat{c}_n))^2 \,\cdot n \cdot \mathrm{Var}_{\hat{\boldsymbol{a}}_n}(\hat{c}_n)$$

satisfies

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{a}}(\, \boldsymbol{a} \in K_n(\alpha)\,) \;=\; 1-\alpha\,.$$

Proof. (*i*) According to (5.13), the image of $(-1,1)$ under $c(\cdot)$ is given by $(0,1)$. Furthermore, h is continuous and thus bounded on $[0,1]$. Therefore, the statement is an immediate consequence of Corollary 5.3 (*i*) and the fact that $\rho_{\boldsymbol{a}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{a} \in (-1,1)$ (see Section 2.2.5).

(*ii*) Note that $|\rho_{\boldsymbol{a}}(k)| = o(k^{-\beta})$ for all $\boldsymbol{a} \in (-1,1)$ and any $\beta > \frac{1}{2}$ (see Section 2.2.5). Since h' is non-vanishing on $(0,1)$, the statement follows by Corollary 5.3 (*ii*).

(*iii*) We show that the conditions of Corollary 5.5 are satisfied. Clearly, $(-1,1)$ is an open set in $\mathbb{R}$ with the image $(0,1)$ under $c(\cdot)$, and h' is continuous on $(0,1)$. Moreover, according to the definition of AR(1) processes, $\rho_{\boldsymbol{a}}(k) = \boldsymbol{a}^k$ for $k \in \mathbb{N}$ and $\boldsymbol{a} \in (-1,1)$ (see Section 2.2.5) which in particular shows that $\boldsymbol{a} \mapsto \rho_{\boldsymbol{a}}(k)$ is continuous on $(-1,1)$ for every $k \in \mathbb{N}$. The proof is complete. □

5.4 Simulation studies

We use the pseudo random number generator of Matlab 7.6.0 and the algorithm of Davies and Harte [33] for the simulation of stationary zero-mean Gaussian processes. For more information on the algorithm and a comparison to other simulation methods, we refer to Bardet et al. [13].

Estimation of the first-order autocorrelation in FGN. For different sample sizes and values of the Hurst parameter, we generate each 100 000 sample paths of FGN and compute the resulting estimates $\hat{\rho}_n$ and $\tilde{\rho}_n$ of the first-order autocorrelation $\rho(\boldsymbol{H})$ (see (5.17) and (5.18)). The sample mean and the sample standard deviation of the 100 000 estimates are taken as estimates of the mean and of the standard deviation of $\hat{\rho}_n$ and $\tilde{\rho}_n$.

Mean and standard deviation of $\hat{\rho}_n$:

		$n = 100$		$n = 1000$		$n = 10\,000$	
$\boldsymbol{H}$	$\rho(\boldsymbol{H})$	μ	σ	μ	σ	μ	σ
0.05	−0.464	−0.459	0.120	−0.464	0.038	−0.464	0.012
0.10	−0.426	−0.422	0.124	−0.426	0.039	−0.426	0.012
0.15	−0.384	−0.381	0.128	−0.384	0.041	−0.384	0.013
0.20	−0.340	−0.337	0.133	−0.340	0.043	−0.340	0.013
0.25	−0.293	−0.289	0.138	−0.293	0.044	−0.293	0.014
0.30	−0.242	−0.240	0.141	−0.242	0.045	−0.242	0.014
0.35	−0.188	−0.185	0.146	−0.187	0.047	−0.188	0.015
0.40	−0.129	−0.128	0.149	−0.129	0.048	−0.129	0.015
0.45	−0.067	−0.066	0.153	−0.067	0.049	−0.067	0.015
0.50	0.000	−0.000	0.155	0.000	0.050	0.000	0.016
0.55	0.072	0.070	0.158	0.072	0.050	0.072	0.016
0.60	0.149	0.147	0.160	0.148	0.051	0.149	0.016
0.65	0.231	0.227	0.161	0.231	0.052	0.231	0.016
0.70	0.320	0.315	0.164	0.319	0.054	0.320	0.017
0.75	0.414	0.407	0.168	0.413	0.058	0.414	0.020
0.80	0.516	0.504	0.173	0.513	0.066	0.515	0.025
0.85	0.625	0.606	0.176	0.621	0.079	0.624	0.038
0.90	0.741	0.712	0.175	0.732	0.093	0.738	0.057
0.95	0.866	0.822	0.161	0.844	0.099	0.854	0.071

Mean and standard deviation of $\tilde{\rho}_n$:

		$n = 100$		$n = 1000$		$n = 10\,000$	
$\boldsymbol{H}$	$\rho(\boldsymbol{H})$	μ	σ	μ	σ	μ	σ
0.05	−0.464	−0.459	0.074	−0.463	0.023	−0.464	0.007
0.10	−0.426	−0.421	0.076	−0.425	0.024	−0.426	0.008
0.15	−0.384	−0.380	0.080	−0.384	0.025	−0.384	0.008
0.20	−0.340	−0.337	0.083	−0.340	0.026	−0.340	0.008
0.25	−0.293	−0.290	0.086	−0.293	0.027	−0.293	0.009
0.30	−0.242	−0.240	0.088	−0.242	0.028	−0.242	0.009
0.35	−0.188	−0.188	0.091	−0.188	0.029	−0.188	0.009
0.40	−0.129	−0.132	0.094	−0.129	0.030	−0.129	0.009
0.45	−0.067	−0.073	0.096	−0.067	0.031	−0.067	0.010
0.50	0.000	−0.010	0.099	−0.001	0.032	−0.000	0.010
0.55	0.072	0.056	0.101	0.070	0.032	0.071	0.010
0.60	0.149	0.124	0.103	0.145	0.033	0.148	0.011
0.65	0.231	0.194	0.105	0.224	0.035	0.230	0.011
0.70	0.320	0.266	0.106	0.307	0.036	0.317	0.012
0.75	0.414	0.339	0.107	0.393	0.038	0.408	0.013
0.80	0.516	0.411	0.108	0.480	0.040	0.503	0.016
0.85	0.625	0.482	0.107	0.567	0.043	0.598	0.019
0.90	0.741	0.549	0.106	0.649	0.044	0.691	0.022
0.95	0.866	0.613	0.102	0.724	0.044	0.775	0.024

Table 5.2: Mean (μ) and standard deviation (σ) of the estimators $\hat{\rho}_n$ and $\tilde{\rho}_n$ of the first-order autocorrelation in FGN.

Mean and standard deviation of $\hat{\boldsymbol{H}}_{\boldsymbol{n}}$, and coverage of $\boldsymbol{H}$ by the asymptotic 95% confidence interval:

	$n = 100$			$n = 1000$			$n = 10\,000$		
$\boldsymbol{H}$	μ	σ	cov.	μ	σ	cov.	μ	σ	cov.
0.05	0.087	0.101	0.965	0.054	0.043	0.973	0.050	0.016	0.949
0.10	0.119	0.113	0.973	0.099	0.048	0.971	0.100	0.016	0.951
0.15	0.156	0.124	0.964	0.149	0.048	0.948	0.150	0.015	0.951
0.20	0.199	0.131	0.969	0.199	0.047	0.952	0.200	0.015	0.949
0.25	0.245	0.134	0.973	0.249	0.045	0.949	0.250	0.014	0.948
0.30	0.291	0.133	0.962	0.299	0.043	0.949	0.300	0.014	0.950
0.35	0.341	0.131	0.945	0.349	0.042	0.948	0.350	0.013	0.950
0.40	0.390	0.127	0.946	0.399	0.040	0.949	0.400	0.013	0.949
0.45	0.441	0.121	0.943	0.449	0.038	0.951	0.450	0.012	0.950
0.50	0.491	0.116	0.954	0.499	0.036	0.949	0.500	0.011	0.949
0.55	0.541	0.110	0.949	0.549	0.034	0.950	0.550	0.011	0.949
0.60	0.592	0.103	0.954	0.599	0.032	0.951	0.600	0.010	0.950
0.65	0.641	0.097	0.961	0.649	0.031	0.954	0.650	0.010	0.951
0.70	0.692	0.093	0.967	0.699	0.030	0.955	0.700	0.010	0.950
0.75	0.741	0.089	0.962	0.749	0.029	0.961	0.750	0.010	0.954
0.80	0.790	0.085	0.955	0.798	0.031	0.959	0.800	0.012	0.958
0.85	0.837	0.081	0.923	0.848	0.035	0.955	0.849	0.017	0.972
0.90	0.884	0.076	0.874	0.895	0.039	0.927	0.898	0.023	0.971
0.95	0.930	0.067	0.750	0.941	0.039	0.834	0.944	0.027	0.899

Mean and standard deviation of $\tilde{\boldsymbol{H}}_{\boldsymbol{n}}$:

	$n = 100$		$n = 1000$		$n = 10\,000$	
$\boldsymbol{H}$	μ	σ	μ	σ	μ	σ
0.05	0.070	0.072	0.051	0.030	0.050	0.010
0.10	0.108	0.082	0.100	0.031	0.100	0.010
0.15	0.152	0.088	0.150	0.030	0.150	0.009
0.20	0.199	0.089	0.200	0.029	0.200	0.009
0.25	0.248	0.087	0.250	0.028	0.250	0.009
0.30	0.297	0.085	0.300	0.027	0.300	0.008
0.35	0.345	0.082	0.350	0.026	0.350	0.008
0.40	0.394	0.079	0.400	0.025	0.400	0.008
0.45	0.442	0.076	0.449	0.024	0.450	0.008
0.50	0.489	0.073	0.499	0.023	0.500	0.007
0.55	0.536	0.070	0.548	0.022	0.550	0.007
0.60	0.581	0.067	0.597	0.021	0.600	0.007
0.65	0.625	0.064	0.646	0.020	0.649	0.007
0.70	0.668	0.061	0.693	0.020	0.698	0.007
0.75	0.708	0.059	0.739	0.020	0.747	0.007
0.80	0.746	0.056	0.783	0.020	0.794	0.008
0.85	0.782	0.053	0.824	0.020	0.838	0.009
0.90	0.814	0.050	0.860	0.019	0.879	0.009
0.95	0.843	0.046	0.893	0.018	0.914	0.010

Table 5.3: Mean (μ) and standard deviation (σ) of the estimators $\hat{\boldsymbol{H}}_n$ and $\tilde{\boldsymbol{H}}_n$ of the Hurst parameter, and the coverage (cov.) of $\boldsymbol{H}$ by the asymptotic 95% confidence interval $K_n(0.05)$.

The results are shown in Table 5.2. We find that the estimator $\hat{\rho}_n$ has a larger standard deviation than $\tilde{\rho}_n$. For instance, when $n = 10\,000$ and the Hurst parameter is smaller than 0.85, the standard deviation is about 1.5 times as large. This loss of efficiency of $\hat{\rho}_n$ compared to $\tilde{\rho}_n$ is not surprising, because $\hat{\rho}_n$ is based only on the number of changes between upwards and downwards whereas $\tilde{\rho}_n$ uses the whole metric information.

When n is small, both estimators overestimate the first-order autocorrelation for $\boldsymbol{H} < 0.5$, and underestimate it when $\boldsymbol{H} > 0.5$. Particularly for large values of the Hurst parameter, $\hat{\rho}_n$ has smaller bias than $\tilde{\rho}_n$. Note that it is well-known that the sample autocorrelation is positively biased in processes with negative correlations and negatively biased in processes with positive correlations (see Beran [15]).

Estimation of the Hurst parameter. The mean and the standard deviation of the estimators $\hat{\boldsymbol{H}}_n$ and $\tilde{\boldsymbol{H}}_n$ of the Hurst parameter are shown in Table 5.3. As mentioned before, with the mapping g given in (5.28), these estimators can be written as $\hat{\boldsymbol{H}}_n = g(\hat{\rho}_n)$ and $\tilde{\boldsymbol{H}}_n = g(\tilde{\rho}_n)$, respectively. Similarly as for $\hat{\rho}_n$ and $\tilde{\rho}_n$, the estimator $\hat{\boldsymbol{H}}_n$ has a larger standard deviation but smaller bias than $\tilde{\boldsymbol{H}}_n$.

For $n = 100$ and $n = 1000$, the standard deviation of $\hat{\boldsymbol{H}}_n$ is particularly large for small values of $\boldsymbol{H}$. A possible explanation is the following: With the mapping h given in (5.25) we have $\hat{\boldsymbol{H}}_n = h(\hat{c}_n)$, and (5.26) shows that the first derivative of h at $c(\boldsymbol{H})$ is particularly large when $\boldsymbol{H}$ is small. Thus, when $\boldsymbol{H}$ is small, the variance of $\hat{c}_n$ is relatively small (see Figure 5.1), but small deviations of $\hat{c}_n$ result in relatively large deviations of $\hat{\boldsymbol{H}}_n$.

Table 5.3 also provides estimates of the probability that the asymptotic confidence interval $K_n(0.05)$ covers the true Hurst parameter $\boldsymbol{H}$. These values are obtained by the relative frequency of realizations of $K_n(0.05)$ which cover $\boldsymbol{H}$. The results suggest that $K_n(0.05)$ covers $\boldsymbol{H}$ with a probability of 95% or higher except when $\boldsymbol{H}$ is close to 1 and n is small.

As the following argument shows, it is very likely that the coverage rates provided by Table 5.3 have a precision of at least two decimal places. Suppose we have a binomial experiment with 100 000 trials and $p = \mathbb{P}_{\boldsymbol{H}}(\boldsymbol{H} \in K_n(0.05))$ being the probability of a success. The standard deviation of the sample mean is equal to $\sqrt{p(1-p)/100\,000}$, and thus ≈ 0.00069 when $p \approx 0.95$. By a normal approximation of the binomial distribution we obtain that, with a probability of $\approx 95\%$, the deviation of the sample mean from p is not larger than two times 0.00069.

The distribution of the number of changes. Figure 5.4 shows the distribution of the number of changes in samples of size $n = 100$. For instance, the upper left plot is obtained by determining the number of changes in the 100 000 sample paths generated for the Hurst parameter $\boldsymbol{H} = 0.70$ and displaying the relative frequencies of the outcomes in a histogram. It can be seen that the number of changes is concentrated about 40, corresponding to the probability of a change $c(0.70) = 0.396$ (see (5.10)).

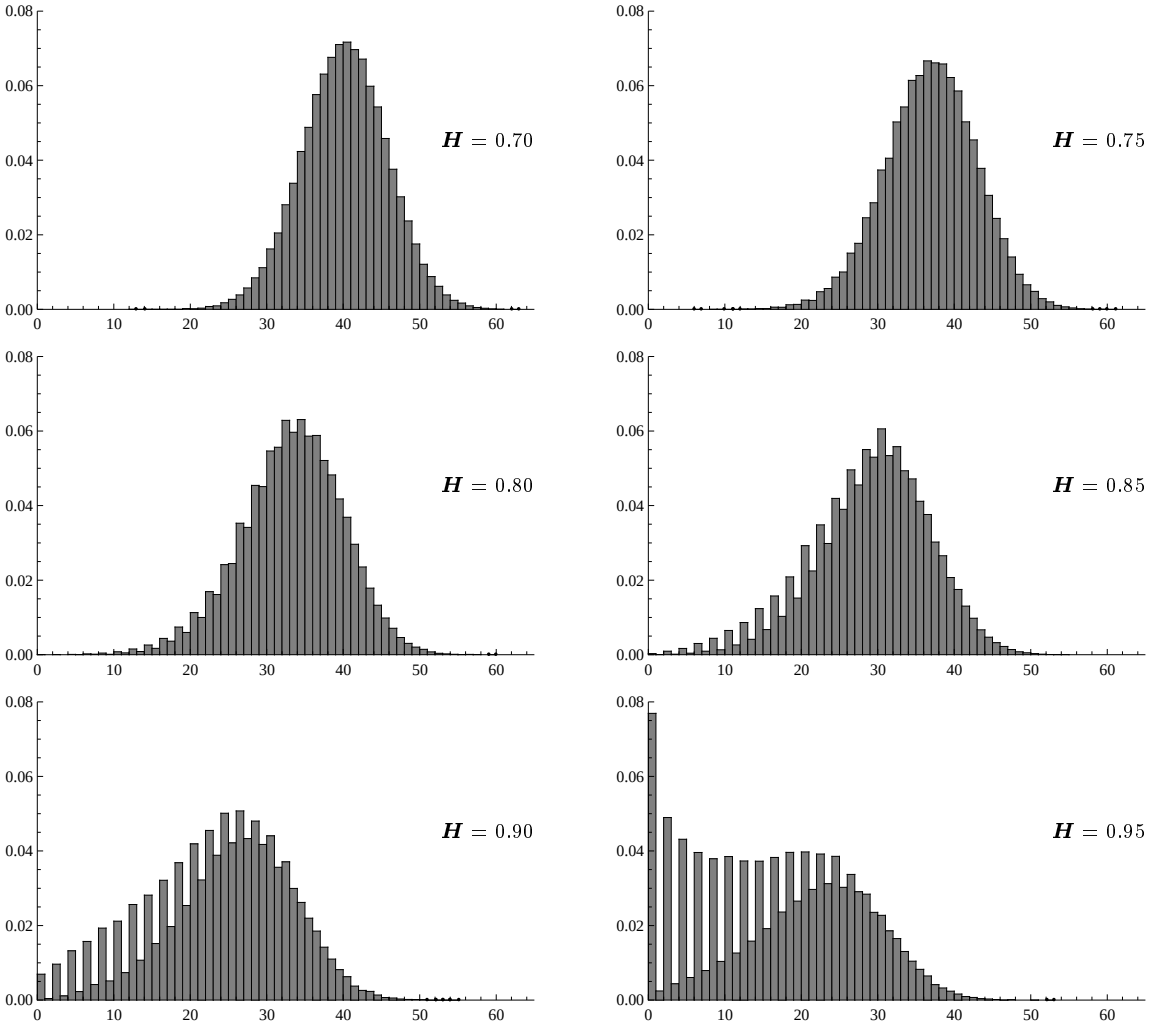

Figure 5.4: Distribution of the number of changes in samples of size $n = 100$.

For $\boldsymbol{H} = 0.70$, the distribution of the number of changes is approximately normal. Note that we find similar distributions also for smaller values of the Hurst parameter. For larger values of $\boldsymbol{H}$, the distributions become more and more irregular. Remarkably, the frequency of even numbers is larger than the frequency of odd numbers, and the distributions conditioned on an odd and an even number, respectively, look entirely different. For instance, in the case $\boldsymbol{H} = 0.95$, the frequencies of odd and even numbers are 0.321 and 0.679, respectively. The distribution conditioned on an odd number is slightly left-skewed and has the mean 21.5 and the mode 23. The distribution conditioned on an even number has the mean 14.4, the mode 0 and roughly looks like the mixture of a geometric and a binomial distribution. As a consequence of the contrast between the conditional distributions, the probability of a change (which is given by $c(0.95) = 0.167$) is overestimated by

the relative frequency of changes in a sample given that the number of changes is odd, and underestimated given that the number of changes is even.

An intuitive explanation for the high frequency of even numbers is the following: When $\boldsymbol{H}$ is large, there is a high probability to observe sample path segments which roughly look like a straight line. Typically for such segments, there are only local changes in direction. Globally, there is one prevailing trend, either "upwards" or "downwards", and thus the overall number of changes between "upwards" and "downwards" is even. In other words: a sample path segment with an even number of changes is more similar to a straight line than a sample path segment with an odd number of changes.

A related finding is that, when $\boldsymbol{H}$ is large, the occurence of a change at time $t-1$ increases the probability of a change at time t. In order to calculate the conditional probability $\mathbb{P}_{\boldsymbol{H}}(C(t) = 1 \,|\, C(t-1) = 1)$, note that

$$\begin{aligned}\mathbb{P}_{\boldsymbol{H}}(C(t) = 1,\, C(t-1) = 1) \;&=\; \mathbb{P}_{\boldsymbol{H}}(Y_t \le 0,\, Y_{t+1} > 0,\, Y_{t+2} \le 0) \\ &\qquad + \mathbb{P}_{\boldsymbol{H}}(Y_t > 0,\, Y_{t+1} \le 0,\, Y_{t+2} > 0)\end{aligned}$$

and

$$\mathbb{P}_{\boldsymbol{H}}(C(t-1) = 1) \;=\; \mathbb{P}_{\boldsymbol{H}}(Y_t \le 0,\, Y_{t+1} > 0) + \mathbb{P}_{\boldsymbol{H}}(Y_t > 0,\, Y_{t+1} \le 0)$$

(compare to (5.1)). Since $\mathbf{Y}$ is zero-mean Gaussian and non-degenerate, we have

$$\mathbb{P}_{\boldsymbol{H}}(C(t) = 1,\, C(t-1) = 1) \;=\; 2\,\mathbb{P}_{\boldsymbol{H}}(Y_t \ge 0,\, -Y_{t+2} \ge 0,\, Y_{t+2} \ge 0)$$

and

$$\mathbb{P}_{\boldsymbol{H}}(C(t-1) = 1) \;=\; 2\,\mathbb{P}_{\boldsymbol{H}}(Y_t \ge 0,\, -Y_{t+1} \ge 0)\,.$$

Thus, according to Lemma 4.1, we obtain

$$\begin{aligned}\mathbb{P}_{\boldsymbol{H}}(C(t) = 1 \mid C(t-1) = 1) \;&=\; \frac{\mathbb{P}_{\boldsymbol{H}}(Y_t \ge 0,\, -Y_{t+2} \ge 0,\, Y_{t+2} \ge 0)}{\mathbb{P}_{\boldsymbol{H}}(Y_t \ge 0,\, -Y_{t+1} \ge 0)} \\ &=\; \frac{\frac{1}{8} - \frac{1}{2\pi}\arcsin\rho_{\boldsymbol{H}}(1) + \frac{1}{4\pi}\arcsin\rho_{\boldsymbol{H}}(2)}{\frac{1}{4} - \frac{1}{2\pi}\arcsin\rho_{\boldsymbol{H}}(1)}\,.\end{aligned} \tag{5.36}$$

For $\boldsymbol{H} = 0.95$, formula (5.36) yields $\mathbb{P}_H(C(t) = 1 \,|\, C(t-1) = 1) = 0.385$, compared to the unconditional probability $\mathbb{P}_{\boldsymbol{H}}(C(t) = 1) = 0.167$ obtained by formula (5.10). Thus, the probability of a change at time t given a change at time $t-1$ is more than two times larger than the unconditional probability of a change at time t. The occurence of change "clusters" with the paths "keeping track" of the modulo of the number of changes shows how complicated the dependency structure is for large values of the Hurst parameter.

Figure 5.5 shows the cumulative standardized distribution of the number of changes in samples of size $n = 10\,000$, compared to the cumulative standard normal distribution. For values of $\boldsymbol{H}$ greater than 0.75, there is a clear difference between both curves, thus indicating that the number of changes is not asymptotically normally distributed. The difference between the frequencies of odd and even numbers of changes is not as large as in the case $n = 100$. For instance, when $\boldsymbol{H} = 0.95$, the frequencies of odd and even numbers are given by 0.391 and 0.609, respectively. Note that the distributions in Figure 5.5 are similar to those obtained for ARFIMA(0,d,0) processes with the fractional differencing parameter $\boldsymbol{d} = \boldsymbol{H} - \frac{1}{2}$.

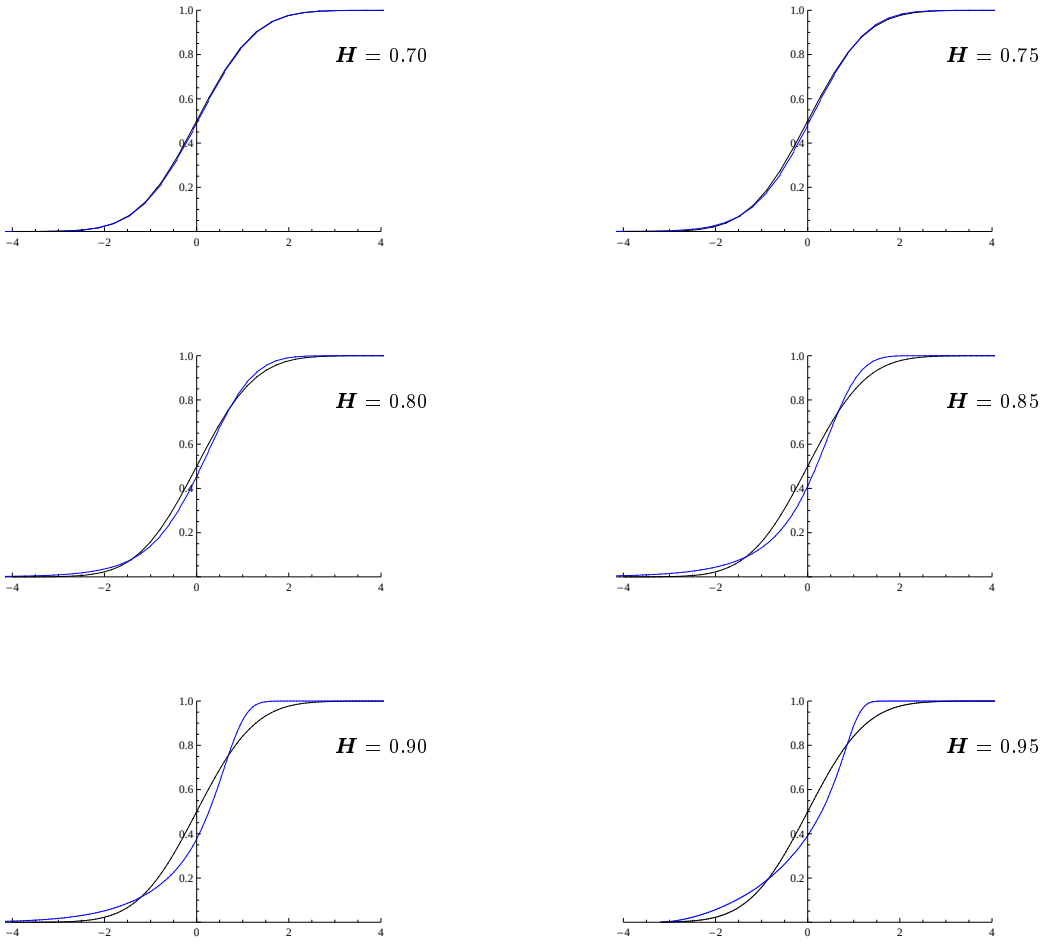

Figure 5.5: Cumulative distribution of the number of changes in samples of size $n = 10\,000$ (blue line), compared to the standard normal distribution (black line).

Mean and standard deviation of $\hat{d}_n$,
and coverage of d by the asymptotic 95% confidence interval:

	$n = 100$			$n = 1000$			$n = 10\,000$		
$\boldsymbol{d}$	μ	σ	cov.	μ	σ	cov.	μ	σ	cov.
−0.45	−0.379	0.147	0.950	−0.436	0.065	0.963	−0.450	0.027	0.972
−0.40	−0.354	0.158	0.954	−0.398	0.074	0.954	−0.401	0.027	0.950
−0.35	−0.325	0.168	0.954	−0.352	0.077	0.967	−0.350	0.026	0.949
−0.30	−0.292	0.175	0.955	−0.304	0.076	0.949	−0.300	0.024	0.950
−0.25	−0.254	0.181	0.954	−0.253	0.072	0.951	−0.250	0.023	0.950
−0.20	−0.214	0.184	0.951	−0.204	0.068	0.951	−0.200	0.021	0.950
−0.15	−0.169	0.184	0.944	−0.153	0.063	0.953	−0.150	0.020	0.950
−0.10	−0.123	0.180	0.959	−0.103	0.059	0.951	−0.100	0.018	0.951
−0.05	−0.075	0.174	0.949	−0.054	0.054	0.951	−0.050	0.017	0.950
0.00	−0.024	0.165	0.956	−0.002	0.050	0.951	−0.000	0.016	0.951
0.05	0.027	0.154	0.958	0.048	0.046	0.950	0.050	0.014	0.950
0.10	0.079	0.142	0.971	0.098	0.042	0.951	0.010	0.013	0.951
0.15	0.131	0.130	0.975	0.148	0.039	0.956	0.150	0.012	0.950
0.20	0.182	0.119	0.974	0.198	0.036	0.957	0.200	0.011	0.950
0.25	0.233	0.108	0.970	0.248	0.034	0.964	0.250	0.011	0.954
0.30	0.282	0.099	0.967	0.297	0.034	0.964	0.300	0.013	0.957
0.35	0.331	0.091	0.943	0.346	0.037	0.954	0.349	0.018	0.972
0.40	0.379	0.083	0.890	0.394	0.040	0.927	0.398	0.024	0.971
0.45	0.425	0.072	0.790	0.439	0.040	0.832	0.444	0.028	0.900

Mean and standard deviation of $\hat{a}_n$,
and coverage of a by the asymptotic 95% confidence interval:

	$n = 100$			$n = 1000$			$n = 10\,000$		
$\boldsymbol{a}$	μ	σ	cov.	μ	σ	cov.	μ	σ	cov.
−0.90	−0.889	0.073	0.905	−0.899	0.023	0.948	−0.900	0.007	0.951
−0.80	−0.790	0.097	0.930	−0.799	0.030	0.947	−0.800	0.010	0.949
−0.70	−0.691	0.113	0.927	−0.699	0.036	0.945	−0.700	0.011	0.951
−0.60	−0.593	0.126	0.949	−0.599	0.040	0.947	−0.600	0.013	0.949
−0.50	−0.494	0.136	0.940	−0.500	0.043	0.948	−0.500	0.014	0.949
−0.40	−0.395	0.143	0.954	−0.400	0.046	0.948	−0.400	0.014	0.951
−0.30	−0.296	0.149	0.940	−0.300	0.048	0.949	−0.300	0.015	0.949
−0.20	−0.197	0.152	0.943	−0.200	0.049	0.950	−0.200	0.015	0.950
−0.10	−0.100	0.155	0.941	−0.100	0.049	0.951	−0.100	0.016	0.950
0.00	−0.000	0.155	0.943	0.000	0.050	0.945	0.000	0.016	0.948
0.10	0.099	0.154	0.944	0.100	0.049	0.950	0.100	0.016	0.950
0.20	0.197	0.152	0.943	0.200	0.049	0.949	0.200	0.015	0.950
0.30	0.296	0.149	0.940	0.299	0.047	0.949	0.300	0.015	0.950
0.40	0.395	0.143	0.952	0.400	0.046	0.948	0.400	0.014	0.950
0.50	0.494	0.136	0.941	0.499	0.043	0.946	0.500	0.014	0.950
0.60	0.592	0.126	0.950	0.599	0.040	0.947	0.600	0.013	0.950
0.70	0.690	0.114	0.927	0.699	0.036	0.947	0.700	0.011	0.951
0.80	0.790	0.098	0.928	0.799	0.030	0.948	0.800	0.010	0.951
0.90	0.889	0.074	0.906	0.899	0.022	0.948	0.900	0.007	0.949

Table 5.4: Mean (μ) and standard deviation (σ) of the estimators $\hat{\boldsymbol{d}}_n$ and $\hat{\boldsymbol{a}}_n$, and the coverage (cov.) of $\boldsymbol{d}$ and $\boldsymbol{a}$ by the 95% asymptotic confidence intervals.

Estimation of the fractional differencing parameter and of the autoregressive coefficient. Table 5.4 shows the mean and the standard deviation of the estimators $\hat{\boldsymbol{d}}_n$ and $\hat{\boldsymbol{a}}_n$. Furthermore, Table 5.4 provides the coverage of $\boldsymbol{d}$ and $\boldsymbol{a}$ by the 95%-confidence intervals.

With $\boldsymbol{d} = \boldsymbol{H} - \frac{1}{2}$, the results for $\hat{\boldsymbol{d}}_n$ are similar to those for $\hat{\boldsymbol{H}}_n$ provided by Table 5.3. Except for very large values of the fractional differencing parameter, the probability that the confidence interval $K_n(0.05)$ covers $\boldsymbol{d}$ is 95%, also when the sample size is small. For $\boldsymbol{d} = -0.45$, we obtain coverage rates greater than 95% which may be due to numerical problems of the simulation method when the autocorrelations of the simulated process are negative (see Bardet et al. [13]).

Most remarkably for the estimator $\hat{\boldsymbol{a}}_n$, the results for the mean and the standard deviation are symmetric about $\boldsymbol{a} = 0$. While the bias is particularly large when $\boldsymbol{a}$ is close to 1 or -1, the standard deviation reaches its maximum when $\boldsymbol{a} = 0$. For $n = 100$, the coverage of $\boldsymbol{a}$ by the confidence interval $K_n(0.05)$ is below 95%, except for some outliers (e.g., $\boldsymbol{a} = 0.40$ and $\boldsymbol{a} = 0.60$). For $n = 1000$ and $n = 10\,000$, the coverage rates are close to 95%.

Chapter 6

Ordinal patterns on different scales

In this chapter, we introduce *delays* of ordinal patterns as an additional parameter besides the order. The framework of our analysis is the same as in the previous chapters, namely, we consider a real-valued stochastic process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ on a measurable space $(\Omega, \mathcal{A})$ which is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}$ where $\boldsymbol{\Theta} \neq \emptyset$. The process of increments $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ is given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$. **We always assume that $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) on p. 42.** For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z}$, let $\rho_{\boldsymbol{\vartheta}}(k) = \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0, Y_k)$ denote the autocorrelations of $\mathbf{Y}$.

Introducing the delays allows to consider the order relations among values at arbitrary times instead of immediately subsequent values. As we will see in Sections 6.1 and 6.2, the results from Chapters 3 - 5 are also valid in this more general context. In Section 6.3, we demonstrate how ordinal patterns with increasing delays can be used to estimate the Hurst parameter in equidistant discretizations of Fractional Brownian Motion superimposed by short range dependent "noise". More generally, we consider the estimation of the index of asymptotically self-similar processes in Section 6.4. The application to practical time series is illustrated in Section 6.5.

6.1 Ordinal pattern delays

Let $d \in \mathbb{N}$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_d) \in \mathbb{N}^d$. Define $\boldsymbol{\tau}(0) := 0$ and $\boldsymbol{\tau}(k) := \tau_1 + \tau_2 + \ldots + \tau_k$ for $k = 1, 2, \ldots, d$. The *ordinal pattern* of order d with the *delays* $\boldsymbol{\tau}$ at time t is given by

$$\Pi^{\boldsymbol{\tau}}(t) \;:=\; \pi\big(X_t,\, X_{t+\boldsymbol{\tau}(1)},\, X_{t+\boldsymbol{\tau}(2)},\, \ldots,\, X_{t+\boldsymbol{\tau}(d)}\big)\,,$$

with the mapping $\pi : \mathbb{R}^{d+1} \to S_d$ as defined in Section 3.2. Ordinal patterns as defined in Chapter 3 are included here as the special case where $\tau_k = 1$ for $k = 1, 2, \ldots, d$. Next, we generalize the results of Chapter 3 to the estimation of ordinal pattern probabilities with arbitrary delays.

Stationarity. Let $\mathbf{x} = (x_0, x_1, \ldots, x_{\boldsymbol{\tau}(d)}) \in \mathbb{R}^{\boldsymbol{\tau}(d)+1}$. Clearly, shifting a vector by a constant does not change any order relation among its components, so we have

$$\pi(x_0,\, x_{\boldsymbol{\tau}(1)},\, x_{\boldsymbol{\tau}(2)},\, \ldots,\, x_{\boldsymbol{\tau}(d)}) \;=\; \pi(0,\, x_{\boldsymbol{\tau}(1)} - x_0,\, x_{\boldsymbol{\tau}(2)} - x_0,\, \ldots,\, x_{\boldsymbol{\tau}(d)} - x_0)\,.$$

Note that, for $k = 1, 2, \ldots, \boldsymbol{\tau}(d)$, we can write the difference $x_{\boldsymbol{\tau}(k)} - x_0$ as the telescoping sum $(x_1 - x_0) + (x_2 - x_1) + \ldots + (x_{\boldsymbol{\tau}(k)} - x_{\boldsymbol{\tau}(k)-1})$. Thus, with the mapping $\tilde{\pi}^{\boldsymbol{\tau}}$ from $\mathbb{R}^{\boldsymbol{\tau}(d)}$ onto S_d given by

$$\tilde{\pi}^{\boldsymbol{\tau}}(\mathbf{y}) \;:=\; \pi(\,0,\, \sum_{k=1}^{\boldsymbol{\tau}(1)} y_k,\, \sum_{k=1}^{\boldsymbol{\tau}(2)} y_k,\, \ldots,\, \sum_{k=1}^{\boldsymbol{\tau}(d)} y_k\,)$$

for $\mathbf{y} = (y_1, y_2, \ldots, y_{\boldsymbol{\tau}(d)}) \in \mathbb{R}^{\boldsymbol{\tau}(d)}$, we obtain

$$\pi(x_0,\, x_{\boldsymbol{\tau}(1)},\, x_{\boldsymbol{\tau}(2)},\, \ldots,\, x_{\boldsymbol{\tau}(d)}) \;=\; \tilde{\pi}^{\boldsymbol{\tau}}((x_1 - x_0),\, (x_2 - x_1),\, \ldots,\, (x_{\boldsymbol{\tau}(k)} - x_{\boldsymbol{\tau}(k)-1}))\,.$$

Consequently,

$$\Pi^{\boldsymbol{\tau}}(t) \;=\; \tilde{\pi}^{\boldsymbol{\tau}}(Y_{t+1},\, Y_{t+2},\, \ldots,\, Y_{t+\boldsymbol{\tau}(d)})$$

for every $t \in \mathbb{Z}$. Similar to Corollary 3.1 on p. 45, we obtain the following statement.

Corollary 6.1. *$(\Pi^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.*

Let $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$. For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, define

$$p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) \;:=\; \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi^{\boldsymbol{\tau}}(t) = \mathbf{r})\,.$$

According to Corollary 6.1, the function $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot)$ does not depend on the specific value of $t \in \mathbb{Z}$ on the right hand side of the definition. We call $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot)$ the *probability of the ordinal pattern* $\mathbf{r}$ *with the delays* $\boldsymbol{\tau}$. The following corollary shows that, for any delays, ordinal pattern probabilities are strictly positive.

Corollary 6.2. *For every $\mathbf{r} = (r_0, r_1, \ldots, r_d) \in S_d$ with $d \in \mathbb{N}$ and every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$,*

$$0 \;<\; p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) \;<\; 1\,.$$

Proof. Let $\mathbf{s} = (s_0, s_1, \ldots, s_{\boldsymbol{\tau}(d)}) \in S_{\boldsymbol{\tau}(d)}$ be such that $s_k = \boldsymbol{\tau}(r_k)$ for $k = 0, 1, \ldots, d$. Clearly, $X_{s_0} > X_{s_1} > \ldots > X_{s_{\boldsymbol{\tau}(d)}}$ implies $X_{\boldsymbol{\tau}(r_0)} > X_{\boldsymbol{\tau}(r_1)} > \ldots > X_{\boldsymbol{\tau}(r_d)}$ and thus

$$\begin{aligned} p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) &= \mathbb{P}_{\boldsymbol{\vartheta}}(X_{\boldsymbol{\tau}(r_0)} > X_{\boldsymbol{\tau}(r_1)} > \ldots > X_{\boldsymbol{\tau}(r_d)}) \\ &\geq \mathbb{P}_{\boldsymbol{\vartheta}}(X_{s_0} > X_{s_1} > \ldots > X_{s_{\boldsymbol{\tau}(d)}}) \\ &= p_{\mathbf{s}}(\boldsymbol{\vartheta}) \end{aligned}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. By Corollary 3.2 we obtain $p_{\mathbf{s}}(\boldsymbol{\vartheta}) > 0$ and thus $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) > 0$. Furthermore, since there exists an $\tilde{\mathbf{r}} \in S_d$ with $\tilde{\mathbf{r}} \neq \mathbf{r}$, we have $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) \leq 1 - p_{\tilde{\mathbf{r}}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) < 1$. The proof is complete. □

Let $n \in \mathbb{N}$ and suppose we are given an observation of the ordinal pattern sample

$$\mathbf{\Pi}_n^{\boldsymbol{\tau}} \;:=\; (\Pi^{\boldsymbol{\tau}}(0), \Pi^{\boldsymbol{\tau}}(1), \ldots, \Pi^{\boldsymbol{\tau}}(n-1))$$

governed by $\mathbb{P}_{\boldsymbol{\vartheta}}$ with $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ unknown. A natural estimator of $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta})$ is given by the relative frequency of observations of $\mathbf{r}$ in $\mathbf{\Pi}_n^{\boldsymbol{\tau}}$, namely,

$$\hat{q}_{\mathbf{r},n}^{\boldsymbol{\tau}} \;:=\; \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)=\mathbf{r}\}} \,.$$

According to Corollary 6.1, $\hat{q}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ is an unbiased estimator of $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot)$.

Similarly as in Section 3.3, we obtain better estimators by averaging the number of observations of $\mathbf{r}$ and of its spatial and time reversals. With the mappings α and β given in (3.8) on p. 47, let the subset $\bar{\mathbf{r}}(\boldsymbol{\tau})$ of S_d be defined by

$$\bar{\mathbf{r}}(\boldsymbol{\tau}) \;:=\; \begin{cases} \{\mathbf{r},\, \alpha(\mathbf{r}),\, \beta(\mathbf{r}),\, \beta\circ\alpha(\mathbf{r})\} & \text{if } \tau_k = \tau_{d-k+1} \text{ for } k = 1,2,\ldots,d \\ \{\mathbf{r},\, \alpha(\mathbf{r})\} & \text{otherwise} \end{cases} \,.$$

Now, let

$$\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}} \;:=\; \frac{1}{n} \sum_{t=0}^{n-1} \frac{1}{\sharp\, \bar{\mathbf{r}}(\boldsymbol{\tau})}\, \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)\in\bar{\mathbf{r}}(\boldsymbol{\tau})\}} \,.$$

If $\tau_k = \tau_{d-k+1}$ for $k = 1,2,\ldots,d$, then $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ is the average of the relative frequencies of $\mathbf{r}$, $\alpha(\mathbf{r})$, $\beta(\mathbf{r})$ and $\beta\circ\alpha(\mathbf{r})$. Otherwise, $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ is the average of the relative frequencies of $\mathbf{r}$ and $\alpha(\mathbf{r})$.

Statement (i) of the following theorem shows that $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ is a better estimator of $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot)$ than $\hat{q}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ ("better" in terms of the risk with respect to convex loss functions, see Theorem 3.5 on p. 51). Statements (ii) and (iii) establish sufficient conditions for strong consistency, asymptotic unbiasedness and asymptotic normality of $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ (and of certain functions of $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$).

Theorem 6.3.

(*i*) *The estimator $\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ of $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot)$ is unbiased and has (strictly) lower risk than $\hat{q}_{\mathbf{r},n}^{\boldsymbol{\tau}}$ with respect to any (strictly) convex loss function.*

(*ii*) *If $\rho_{\boldsymbol{\vartheta}}(k) \to 0$ as $k \to \infty$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $h : [0,1] \to \mathbb{R}$ is continuous on an open set containing $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\Theta})$, then $h(\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}})$ is a strongly consistent estimator of $h(p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot))$. If, additionally, h is bounded on $[0,1]$, then $h(\hat{p}_{\mathbf{r},n}^{\boldsymbol{\tau}})$ is an asymptotically unbiased estimator of $h(p_{\mathbf{r}}^{\boldsymbol{\tau}}(\cdot))$.*

(*iii*) *If* $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ *for some* $\beta > \frac{1}{2}$ *and* $h : [0,1] \to \mathbb{R}$ *has a non-vanishing first derivative at* $p^{\boldsymbol{\tau}}_{\mathbf{r}}(\boldsymbol{\vartheta})$, *then*

$$\sqrt{n}\left(h(\hat{p}^{\boldsymbol{\tau}}_n) - h(p^{\boldsymbol{\tau}}_{\mathbf{r}}(\boldsymbol{\vartheta}))\right) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}} \quad \mathrm{N}(0,\, \sigma^2_{\boldsymbol{\vartheta}}[h'(p^{\boldsymbol{\tau}}_{\mathbf{r}}(\boldsymbol{\vartheta}))]^2\,)\,,$$

where

$$\sigma^2_{\boldsymbol{\vartheta}} \;:=\; \gamma^{\boldsymbol{\tau}}_{\boldsymbol{\vartheta}}(0) \,+\, 2\sum_{k=1}^{\infty} \gamma^{\boldsymbol{\tau}}_{\boldsymbol{\vartheta}}(k)$$

and $\gamma^{\boldsymbol{\tau}}_{\boldsymbol{\vartheta}}(k) := \frac{1}{(\sharp\, \bar{\mathbf{r}}(\boldsymbol{\tau}))^2}\,\mathrm{Cov}(\mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(0)\in\bar{\mathbf{r}}(\boldsymbol{\tau})\}},\, \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(k)\in\bar{\mathbf{r}}(\boldsymbol{\tau})\}})$ *for* $k \in \mathbb{Z}$.

Proof. (*i*) Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $\mathbf{x} = (x_0, x_1, \ldots, x_{\boldsymbol{\tau}(d)}) \in \mathbb{R}^{\boldsymbol{\tau}(d)+1}$. If $\mathbf{x}$ has pairwise different components, then

$$-x_{\boldsymbol{\tau}(r_0)} > -x_{\boldsymbol{\tau}(r_1)} > \ldots > -x_{\boldsymbol{\tau}(r_d)} \quad \Leftrightarrow \quad x_{\boldsymbol{\tau}(r_d)} > x_{\boldsymbol{\tau}(r_{d-1})} > \ldots > x_{\boldsymbol{\tau}(r_0)}\,.$$

By the same argument as in the proof of Lemma 3.3, we obtain that

$$\alpha(\tilde{\pi}^{\boldsymbol{\tau}}(Y_{t+1}, Y_{t+2}, \ldots, Y_{t+\boldsymbol{\tau}(d)})) \;=\; \tilde{\pi}^{\boldsymbol{\tau}}(-Y_{t+1}, -Y_{t+2}, \ldots, -Y_{t+\boldsymbol{\tau}(d)})$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $t \in \mathbb{Z}$ and thus, with the mapping A defined on p. 48,

$$\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n \;\overset{\text{dist}}{=}\; A(\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n)\,.$$

Now, suppose $\tau_k = \tau_{d-k+1}$ for $k = 1, 2, \ldots, d$. Clearly,

$$\begin{aligned}\boldsymbol{\tau}(d) - \boldsymbol{\tau}(k) &= \tau_{k+1} + \tau_{k+2} + \ldots + \tau_d \\ &= \tau_1 + \tau_2 + \ldots + \tau_{d-k} \;=\; \boldsymbol{\tau}(d-k)\end{aligned}$$

for $k = 0, 1, \ldots, d$ and hence

$$x_{\boldsymbol{\tau}(d)-\boldsymbol{\tau}(r_0)} > x_{\boldsymbol{\tau}(d)-\boldsymbol{\tau}(r_1)} > \ldots > x_{\boldsymbol{\tau}(d)-\boldsymbol{\tau}(r_d)} \quad \Leftrightarrow \quad x_{\boldsymbol{\tau}(d-r_0)} > x_{\boldsymbol{\tau}(d-r_1)} > \ldots > x_{\boldsymbol{\tau}(d-r_d)}\,.$$

Similar to the proof of Lemma 3.3, we obtain

$$\beta(\tilde{\pi}^{\boldsymbol{\tau}}(Y_{t+1}, Y_{t+2}, \ldots, Y_{t+\boldsymbol{\tau}(d)})) \;=\; \tilde{\pi}^{\boldsymbol{\tau}}(-Y_{t+\boldsymbol{\tau}(d)}, -Y_{t+\boldsymbol{\tau}(d)-1}, \ldots, -Y_{t+1})$$

$\mathbb{P}_{\boldsymbol{\vartheta}}$-almost surely for every $t \in \mathbb{Z}$ and thus

$$\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n \;\overset{\text{dist}}{=}\; A(\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n) \;\overset{\text{dist}}{=}\; B(\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n) \;\overset{\text{dist}}{=}\; B \circ A(\boldsymbol{\Pi}^{\boldsymbol{\tau}}_n)\,.$$

Now, by the same arguments as in the proof of Theorem 3.5 and in Remark 3.6, it follows that $\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n}$ has lower risk than $\hat{q}^{\boldsymbol{\tau}}_{\mathbf{r},n}$ with respect to any convex loss function.

Let $\mathbf{s} = (s_0, s_1, \ldots, s_{n+\boldsymbol{\tau}(d)-1}) \in S_{n+\boldsymbol{\tau}(d)-1}$ be such that $X_{s_0} > X_{s_1} > \ldots > X_{s_{n+\boldsymbol{\tau}(d)-1}}$ implies $X_{\boldsymbol{\tau}(r_d)} > X_{\boldsymbol{\tau}(r_{d-1})} > \ldots > X_{\boldsymbol{\tau}(r_0)}$ and $X_{t+\boldsymbol{\tau}(r_d)} > X_{t+\boldsymbol{\tau}(r_{d-1})}$ for $t = 1, 2, \ldots, n-1$. (The existence of such a permutation is obvious.) Then

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n} \neq \hat{q}^{\boldsymbol{\tau}}_{\mathbf{r},n}) &\geq \mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n} > 0,\, \hat{q}^{\boldsymbol{\tau}}_{\mathbf{r},n} = 0) \\ &\geq \mathbb{P}_{\boldsymbol{\vartheta}}(\Pi^{\boldsymbol{\tau}}(0) = \alpha(\mathbf{r}),\, \Pi^{\boldsymbol{\tau}}(1) \neq \mathbf{r},\, \Pi^{\boldsymbol{\tau}}(2) \neq \mathbf{r},\, \ldots,\, \Pi^{\boldsymbol{\tau}}(n-1) \neq \mathbf{r}) \\ &\geq \mathbb{P}_{\boldsymbol{\vartheta}}(X_{s_0} > X_{s_1} > \ldots > X_{s_{n+\boldsymbol{\tau}(d)-1}})\end{aligned}$$

for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. According to Corollary 6.2, the latter probability is strictly positive, so $\mathbb{P}_{\boldsymbol{\vartheta}}(\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n} \neq \hat{q}^{\boldsymbol{\tau}}_{\mathbf{r},n}) > 0$ for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Now, by the same argument as in the proof of Theorem 3.5, it follows that $\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n}$ has strictly lower risk than $\hat{q}^{\boldsymbol{\tau}}_{\mathbf{r},n}$ with respect to strictly convex loss functions.

(*ii*) For $\mathbf{y} = (y_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$, let $f(\mathbf{y}) := 1$ if $\tilde{\pi}^{\boldsymbol{\tau}}(y_1, y_2, \ldots, y_{\boldsymbol{\tau}(d)}) \in \bar{\mathbf{r}}(\boldsymbol{\tau})$, and $f(\mathbf{y}) := 0$, otherwise. Then the result is obtained analogously to Theorem 3.8.

(*iii*) Let $g : \mathbb{R}^{\boldsymbol{\tau}(d)} \to \mathbb{R}$ be defined by

$$g(\mathbf{z}) := \begin{cases} \frac{1}{\sharp\,\bar{\mathbf{r}}(\boldsymbol{\tau})} & \text{if } \tilde{\pi}^{\boldsymbol{\tau}}(\mathbf{z}) \in \bar{\mathbf{r}}(\boldsymbol{\tau}) \\ 0 & \text{otherwise} \end{cases}$$

for $\mathbf{z} \in \mathbb{R}^{\boldsymbol{\tau}(d)}$. By the same argument as in the proof of Theorem 3.10, we obtain that g has Hermite rank $\kappa \geq 2$ with respect to any zero-mean non-degenerate Gaussian random vector. Therefore, similar to Theorem 3.10 and Corollary 3.17, the statement follows. □

Delays in equidistant discretizations of FBM. As we show next, the distribution of ordinal patterns in equidistant discretizations of FBM is invariant with respect to a simultaneous scaling of time and delays. Same as in Section 3.5, suppose $(\Omega, \mathcal{A})$ is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{H}})_{\boldsymbol{H}\in(0,1)}$ such that $\mathbf{X}$ measured with respect to $\mathbb{P}_{\boldsymbol{H}}$ is an equidistant discretization of FBM with the Hurst parameter $\boldsymbol{H}$. As shown in Section 3.5, the distribution of ordinal patterns in $\mathbf{X}$ does not depend on the sampling interval length δ, so we may assume $\delta = 1$ without loss of generality.

Now, let $\boldsymbol{H} \in (0,1)$ and $N \in \mathbb{N}$. By $N\boldsymbol{\tau}$ we denote the vector $(N\tau_1, N\tau_2, \ldots, N\tau_d)$. Note that $(\Pi^{N\boldsymbol{\tau}}(Nt))_{t\in\mathbb{Z}}$ can be regarded as the process of ordinal patterns with delays $\boldsymbol{\tau}$ in an equidistant discretization of FBM with the sampling interval length N. Thus, $(\Pi^{N\boldsymbol{\tau}}(Nt))_{t\in\mathbb{Z}}$ has the same distribution as the process of ordinal patterns with delays $\boldsymbol{\tau}$ in an equidistant discretization of FBM with the sampling interval length 1, that is,

$$(\Pi^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}} \overset{\text{dist}}{=} (\Pi^{N\boldsymbol{\tau}}(Nt))_{t\in\mathbb{Z}}\,.$$

As a consequence, we obtain

$$p^{\boldsymbol{\tau}}_{\mathbf{r}}(\cdot) = p^{N\boldsymbol{\tau}}_{\mathbf{r}}(\cdot) \tag{6.1}$$

for every $N \in \mathbb{N}$. Note that (6.1) is valid for the ordinal patterns in any self-similar process with stationary increments. Therefore, as a simple non-rigorous method for testing whether a time series is generated by a self-similar process, we propose to check whether the estimates $\hat{p}^{N\boldsymbol{\tau}}_{\mathbf{r},n}$ are "similar" for different values of N. We consider some examples at the end of this chapter.

Analogous to the definition of $\boldsymbol{\Pi_H}$ given in Section 3.5, let $\boldsymbol{\Pi}^{\boldsymbol{\tau}}_{\boldsymbol{H}}$ denote the distribution of the ordinal pattern process $(\Pi^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ in an equidistant discretization of FBM with the Hurst parameter $\boldsymbol{H}$.

6.2 The case d=2

Let us consider ordinal patterns of order $d = 2$. We assume that the delays are given by $\boldsymbol{\tau} = (\tau, \tau)$ for some $\tau \in \mathbb{N}$. For $t \in \mathbb{Z}$, define

$$C^{\boldsymbol{\tau}}(t) \;:=\; \mathbf{1}_{\{X_t \geq X_{t+\tau} < X_{t+2\tau}\}} + \mathbf{1}_{\{X_t < X_{t+\tau} \geq X_{t+2\tau}\}}\,. \tag{6.2}$$

If we consider τ-distant values of $\mathbf{X}$, we may regard $C^{\boldsymbol{\tau}}(t)$ as the indicator for a change between "upwards" and "downwards". Similarly as in the case $\boldsymbol{\tau} = (1, 1)$ discussed in Chapter 5, we have

$$C^{\boldsymbol{\tau}}(t) \;=\; \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)=(2,0,1)\}} + \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)=(0,2,1)\}} + \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)=(1,2,0)\}} + \mathbf{1}_{\{\Pi^{\boldsymbol{\tau}}(t)=(1,0,2)\}}$$

(compare to (5.2) on p. 98). Since $(\Pi^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ is stationary, it follows that $(C^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ is stationary for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Consequently, the relative frequency of changes, given by

$$\hat{c}^{\boldsymbol{\tau}}_n \;:=\; \frac{1}{n}\sum_{t=0}^{n-1} C^{\boldsymbol{\tau}}(t) \tag{6.3}$$

for $n \in \mathbb{N}$, is an unbiased estimator of the probability of a change, given by

$$c^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) \;:=\; \mathbb{P}_{\boldsymbol{\vartheta}}(C^{\boldsymbol{\tau}}(t) = 1)$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Similar to the case $\boldsymbol{\tau} = (1, 1)$, we obtain

$$\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n} \;=\; \begin{cases} \frac{1}{4}\,\hat{c}^{\boldsymbol{\tau}}_n & \text{if } \mathbf{r} \in \big\{(1,0,2),\,(1,2,0),\,(0,2,1),\,(2,0,1)\big\} \\ \frac{1}{2}\,(1-\hat{c}^{\boldsymbol{\tau}}_n) & \text{if } \mathbf{r} \in \big\{(2,1,0),\,(0,1,2)\big\} \end{cases}.$$

Analogous to Corollary 5.2, we obtain statistical properties of $\hat{c}^{\boldsymbol{\tau}}_n$ by the corresponding properties of $\hat{p}^{\boldsymbol{\tau}}_{\mathbf{r},n}$.

Corollary 6.4.

(*i*) $\hat{c}_n^{\boldsymbol{\tau}}$ *is an unbiased estimator of* $c^{\boldsymbol{\tau}}(\cdot)$.

(*ii*) *If* $\lim_{k\to\infty} \rho_{\boldsymbol{\vartheta}}(k) = 0$ *for every* $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ *and* $h : [0,1] \to \mathbb{R}$ *is continuous on an open set containing* $c^{\boldsymbol{\tau}}(\boldsymbol{\Theta})$, *then* $h(\hat{c}_n^{\boldsymbol{\tau}})$ *is a strongly consistent estimator of* $h(c^{\boldsymbol{\tau}}(\cdot))$. *If, additionally,* h *is bounded on* $[0,1]$, *then* $h(\hat{c}_n^{\boldsymbol{\tau}})$ *is an asymptotically unbiased estimator of* $h(c^{\boldsymbol{\tau}}(\cdot))$.

(*iii*) *If* $|\rho_{\boldsymbol{\vartheta}}(k)| = o(k^{-\beta})$ *for some* $\beta > \frac{1}{2}$ *and* $h : [0,1] \to \mathbb{R}$ *has a non-vanishing first derivative at* $c^{\boldsymbol{\tau}}(\boldsymbol{\vartheta})$, *then*

$$\sqrt{n}\,\big(\, h(\hat{c}_n^{\boldsymbol{\tau}}) - h(c^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}))\,\big) \quad \xrightarrow{\mathbb{P}_{\boldsymbol{\vartheta}}} \quad \mathrm{N}(0,\, \sigma_{\boldsymbol{\vartheta}}^2 [h'(c^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}))]^2\,)\,,$$

where

$$\sigma_{\boldsymbol{\vartheta}}^2 \;:=\; \gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(0) + 2\sum_{k=1}^{\infty} \gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k)$$

and $\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) := \mathrm{Cov}_{\boldsymbol{\vartheta}}(C^{\boldsymbol{\tau}}(0), C^{\boldsymbol{\tau}}(k))$ *for* $k \in \mathbb{Z}$.

Computation of moments. Next, we investigate the evaluation of the first and second moments of the process $(C^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$. This allows us, e.g., to compute the variance of $\hat{c}_n^{\boldsymbol{\tau}}$. The key ingredient is the evaluation of correlations between sums of increments. For $t \in \mathbb{Z}$, define

$$Z_t \;:=\; Y_{t+1} + Y_{t+2} + \ldots + Y_{t+\tau}\,.$$

Furthermore, let

$$\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) \;:=\; \mathrm{Corr}_{\boldsymbol{\vartheta}}(Z_0, Z_k)$$

for $k \in \mathbb{Z}$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Since $\mathbf{Y}$ is stationary, non-degenerate and Gaussian (see the model assumptions (M1)-(M3) on p. 42), we obtain that $(Z_t)_{t\in\mathbb{Z}}$ is stationary, non-degenerate and Gaussian for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Moreover,

$$\begin{aligned} \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) &= \frac{\mathrm{Cov}_{\boldsymbol{\vartheta}}(Z_0, Z_k)}{\mathrm{Var}_{\boldsymbol{\vartheta}}(Z_0)} \\ &= \frac{\sum_{l=-(\tau-1)}^{\tau-1} (\tau - |l|)\, \rho_{\boldsymbol{\vartheta}}(k+l)}{\tau + 2\sum_{l=1}^{\tau-1} (\tau - l)\, \rho_{\boldsymbol{\vartheta}}(l)} \end{aligned} \tag{6.4}$$

for $k \in \mathbb{Z}$ and $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, where $\rho_{\boldsymbol{\vartheta}}$ is the autocorrelation function of $\mathbf{Y}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$. According to (6.2), we can express $C^{\boldsymbol{\tau}}(t)$ in terms of sums of increments, namely,

$$C^{\boldsymbol{\tau}}(t) \;=\; \mathbf{1}_{\{Z_t \le 0,\, Z_{t+\tau} > 0\}} + \mathbf{1}_{\{Z_t > 0,\, Z_{t+\tau} \le 0\}}$$

for $t \in \mathbb{Z}$. Thus, analogous to formula (4.4) on p. 69, we obtain

$$c^{\boldsymbol{\tau}}(\boldsymbol{\vartheta}) \;=\; \frac{1}{2} - \frac{1}{\pi} \arcsin \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau)$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. The autocovariances $\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) = \mathrm{Cov}_{\boldsymbol{\vartheta}}(C^{\boldsymbol{\tau}}(0), C^{\boldsymbol{\tau}}(k))$ for $k = 0, 1, \ldots$ can be evaluated similarly as in the case $\boldsymbol{\tau} = (1,1)$ discussed in Chapter 4. In particular, according to formula (4.5) on p. 69, we obtain

$$\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(0) \;=\; \frac{1}{4} - \frac{1}{\pi^2} (\arcsin \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau))^2 \,,$$

and the analogue of formula (4.6) is given by

$$\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau) \;=\; \frac{1}{2\pi} \arcsin \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(2\tau) - \frac{1}{\pi^2} (\arcsin \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau))^2 \,.$$

For $k = 1, 2, \ldots, \tau - 1$ and $k = \tau + 1, \tau + 2, \ldots$, respectively, we have

$$\begin{aligned} \gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) \;&=\; 2\,\mathbb{P}_{\boldsymbol{\vartheta}}(Z_0 > 0,\, Z_\tau > 0,\, Z_k > 0,\, Z_{k+\tau} > 0) \\ &+\; 2\,\mathbb{P}_{\boldsymbol{\vartheta}}(Z_0 > 0,\, Z_\tau > 0,\, -Z_k > 0,\, -Z_{k+\tau} > 0) \\ &-\; 4\,\mathbb{P}_{\boldsymbol{\vartheta}}(Z_0 > 0,\, Z_\tau > 0)\,\mathbb{P}_{\boldsymbol{\vartheta}}(Z_k > 0,\, Z_{k+\tau} > 0) \end{aligned}$$

(compare to (4.7)). Thus, analogous to formula (4.20) on p. 81, we obtain

$$\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) \;=\; \Psi^*(\,\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau),\; \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k),\; \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k+\tau),\; \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k-\tau)\,) \,.$$

Using the integral representation of Ψ^* given by Theorem 4.6, we can evaluate $\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k)$ numerically.

In a similar manner, we obtain formulas for the evaluation of covariances between $C^{\boldsymbol{\tau}_1}(k)$ and $C^{\boldsymbol{\tau}_2}(l)$ for $\boldsymbol{\tau}_1 = (\tau_1, \tau_1)$ and $\boldsymbol{\tau}_2 = (\tau_2, \tau_2)$ with $\tau_1, \tau_2 \in \mathbb{N}$ and $\tau_1 \neq \tau_2$, which is useful for determining the asymptotic joint distribution of $\hat{c}_n^{\boldsymbol{\tau}_1}$ and $\hat{c}_n^{\boldsymbol{\tau}_2}$.

Asymptotic properties. Let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Next, we derive asymptotics of $\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k)$ as $k \to \infty$. Suppose there exists a function $f : \mathbb{N} \to \mathbb{R}$ with $\rho_{\boldsymbol{\vartheta}}(k) \sim f(k)$ and a constant $\beta \neq 0$ such that $f(k+1) \sim \beta f(k)$. Define

$$\kappa \;:=\; \frac{\sum_{l=-(\tau-1)}^{\tau-1} (\tau - |l|)\, \beta^l}{\tau + 2 \sum_{l=1}^{\tau-1} (\tau - l)\, \rho_{\boldsymbol{\vartheta}}(l)} \,.$$

According to formula (6.4), we obtain

$$\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) \;\sim\; \frac{\sum_{l=-(\tau-1)}^{\tau-1} (\tau - |l|)\, \beta^l\, f(k)}{\tau + 2 \sum_{l=1}^{\tau-1} (\tau - l)\, \rho_{\boldsymbol{\vartheta}}(l)} \;=\; \kappa\, f(k) \,.$$

Consequently, $\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k+1) \sim \beta \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k)$, and hence

$$\big(\,\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau),\ \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k),\ \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k+\tau),\ \rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k-\tau)\,\big) \ \sim\ \big(\,\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau),\ \kappa\, f(k),\ \beta^{\tau}\kappa\, f(k),\ \beta^{-\tau}\kappa\, f(k)\,\big)\,.$$

Since $(Z_t)_{t\in\mathbb{Z}}$ is Gaussian and non-degenerate, we have $|\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau)| < 1$ (see Theorem 2.6 (iv)). Thus, according to Corollary 4.9 (ii),

$$\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k) \ \sim\ \frac{(\kappa\, f(k))^2\,(2-\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau)\,(\beta^{\tau}+\beta^{-\tau}))^2}{2\pi^2(1-(\rho_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(\tau))^2)} + O\big((f(k))^4\big)\,.$$

Similar to Theorem 4.18, the latter expression allows to derive the asymptotics of the variance of $\hat{c}_n^{\boldsymbol{\tau}}$. Note that the rate of decreasing of $\gamma_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}}(k)$ does not depend on τ, and hence the same is true for the rate of decreasing of the variance of $\hat{c}_n^{\boldsymbol{\tau}}$.

6.3 Increasing delays

Next, we investigate ordinal pattern probabilities for increasing delays. In particular, we are interested in whether the sequence $p_{\mathbf{r}}^{\boldsymbol{\tau}}(\boldsymbol{\vartheta})$, $p_{\mathbf{r}}^{2\boldsymbol{\tau}}(\boldsymbol{\vartheta})$, $p_{\mathbf{r}}^{3\boldsymbol{\tau}}(\boldsymbol{\vartheta})$, ... (with fixed $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$) has a well-defined limit. In Section 6.4, we study this problem using results from the theory of renormalization groups. From the statistical viewpoint, considering ordinal patterns on large time scales is useful for studying the long term behavior of time series. For a mathematical description, we introduce a more specific model for time series.

Mixture models. Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ a sequence of measurable mappings from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. The increment process $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ is given by $Y_t := X_t - X_{t-1}$ for $t \in \mathbb{Z}$. Let $m \in \mathbb{N}$ and $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_m \neq \emptyset$. Suppose $(\Omega, \mathcal{A})$ is equipped with a family of probability measures $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}$ where

$$\boldsymbol{\Theta} \ :=\ \bigtimes_{l=1}^{m} \boldsymbol{\Theta}_l\,.$$

Furthermore, suppose there exist $w_1, w_2, \ldots, w_m > 0$ with $w_1^2 + w_2^2 + \ldots + w_m^2 = 1$ and measurable mappings $\mathbf{Y}^{(l)} = (Y_t^{(l)})_{t\in\mathbb{Z}}$ from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ for $l = 1, 2, \ldots, m$ such that

$$\mathbf{Y} \ =\ \sum_{l=1}^{m} w_l\, \mathbf{Y}^{(l)}\,.$$

We will assume that, for every $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \ldots, \boldsymbol{\vartheta}_m) \in \boldsymbol{\Theta}$, the following conditions are satisfied:

(M1') $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ are independent.

(M2') $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ are non-degenerate.

(M3') $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ are stationary.

(M4') $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ are zero-mean Gaussian and have unit variance.

(M5') For $l = 1, 2, \ldots, m$, the autocorrelation function of $\mathbf{Y}^{(l)}$, given by

$$\rho_{\boldsymbol{\vartheta}}^{(l)}(k) \;:=\; \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0^{(l)},\, Y_k^{(l)})$$

for $k \in \mathbb{Z}$, only depends on $\boldsymbol{\vartheta}_l$.

Condition (M5') means that, if $\boldsymbol{\vartheta}' = (\boldsymbol{\vartheta}_1', \boldsymbol{\vartheta}_2', \ldots, \boldsymbol{\vartheta}_m') \in \boldsymbol{\Theta}$ satisfies $\boldsymbol{\vartheta}_l' = \boldsymbol{\vartheta}_l$, we have

$$\rho_{\boldsymbol{\vartheta}'}^{(l)}(k) \;=\; \rho_{\boldsymbol{\vartheta}}^{(l)}(k)$$

for all $k \in \mathbb{Z}$. Assumptions (M1')-(M4') imply that $\mathbf{Y}$ is non-degenerate and stationary zero-mean Gaussian for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. In order to establish non-degeneracy, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $t_1 < t_2 < \ldots < t_k \in \mathbb{Z}$ with $k \in \mathbb{N}$. For $l = 1, 2$, let $\boldsymbol{\Sigma}^{(l)}$ denote the covariance matrix of $(Y_{t_1}^{(l)}, Y_{t_2}^{(l)}, \ldots, Y_{t_k}^{(l)})$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$. Since $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are independent, the covariance matrix of $(w_1 Y_{t_1}^{(1)} + w_2 Y_{t_1}^{(2)},\, w_1 Y_{t_2}^{(1)} + w_2 Y_{t_2}^{(2)},\, \ldots,\, w_1 Y_{t_k}^{(1)} + w_2 Y_{t_k}^{(2)})$ is given by $w_1^2 \boldsymbol{\Sigma}^{(1)} + w_2^2 \boldsymbol{\Sigma}^{(2)}$. Furthermore, because $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are non-degenerate, $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ are strictly positive definite. Thus, for any $\mathbf{x} \in \mathbb{R}^k$ with $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}\big(w_1^2 \boldsymbol{\Sigma}^{(1)} + w_2^2 \boldsymbol{\Sigma}^{(2)}\big)\mathbf{x}^T \;=\; w_1^2\, \mathbf{x}\, \boldsymbol{\Sigma}^{(1)} \mathbf{x}^T + w_2^2\, \mathbf{x}\, \boldsymbol{\Sigma}^{(2)} \mathbf{x}^T \;>\; 0\,,$$

which shows that $w_1^2 \boldsymbol{\Sigma}^{(1)} + w_2^2 \boldsymbol{\Sigma}^{(2)}$ is strictly positive definite and hence $w_1 \mathbf{Y}^{(1)} + w_2 \mathbf{Y}^{(2)}$ is non-degenerate. By repeating this argument, we obtain that $\mathbf{Y}$ is non-degenerate.

Note that, for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $t \in \mathbb{Z}$,

$$\mathrm{Var}_{\boldsymbol{\vartheta}}(Y_t) \;=\; \sum_{l=1}^{m} w_l^2\, \mathrm{Var}_{\boldsymbol{\vartheta}}(Y_t^{(l)})\,.$$

Since $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ have unit variance and $w_1^2 + w_2^2 + \ldots + w_m^2 = 1$, we obtain that $\mathbf{Y}$ has unit variance for every $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Thus, $\mathbf{Y}$ satisfies the model assumptions (M1)-(M3) on p. 42.

Since $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, $\ldots$, $\mathbf{Y}^{(m)}$ are independent, the autocorrelations of $\mathbf{Y}$ are given by

$$\rho_{\boldsymbol{\vartheta}}(k) \;:=\; \mathrm{Corr}_{\boldsymbol{\vartheta}}(Y_0, Y_k) \;=\; \sum_{l=1}^{m} w_l^2\, \rho_{\boldsymbol{\vartheta}}^{(l)}(k) \tag{6.5}$$

for $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ and $k \in \mathbb{Z}$.

Example. Let $m = 2$ and $\boldsymbol{\Theta} = (0,1) \times (-1,1)$. For $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) \in \boldsymbol{\Theta}$, suppose that $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ measured with respect to $\mathbb{P}_{\boldsymbol{\vartheta}}$ are FGN with the Hurst parameter $\boldsymbol{\vartheta}_1$ and an AR(1) process with the autoregressive coefficient $\boldsymbol{\vartheta}_2$, respectively. Then the process $\mathbf{X}$ can be regarded as the mixture of an equidistant discretization of FBM and an integrated AR(1) process. Same as in the previous chapters, we write $\boldsymbol{H}$ instead of $\boldsymbol{\vartheta}_1$ and $\boldsymbol{a}$ instead of $\boldsymbol{\vartheta}_2$.

For many practical applications, the actual value of $\boldsymbol{H}$ is of particular interest. For instance, if $\boldsymbol{H} > \frac{1}{2}$, then $\mathbf{Y}$ has *long memory* (in particular, the autocorrelations of $\mathbf{Y}$ given in (6.5) are not absolutely summable), and the long term behaviour of $\mathbf{X}$ is essentially determined by the contribution of $\mathbf{Y}^{(1)}$. Thus, we may regard $\mathbf{Y}^{(2)}$ as "noise" corrupting the "signal" $\mathbf{Y}^{(1)}$ and affecting the precision of estimates of $\boldsymbol{H}$. The ratio of the signal variance to the total variance of $\mathbf{Y}$ is given by

$$\frac{\mathrm{Var}_{\boldsymbol{\vartheta}}(w_1\, Y_t^{(1)})}{\mathrm{Var}_{\boldsymbol{\vartheta}}(w_1\, Y_t^{(1)} + w_2\, Y_t^{(2)})} \;=\; \frac{w_1^2}{w_1^2 + w_2^2}\;.$$

Figure 6.1 shows simulated sample paths of $\mathbf{X}$ with the Hurst parameter $\boldsymbol{H} = 0.75$ and the autoregressive coefficient $\boldsymbol{a} = -0.75$. The squared weights w_1^2 and w_2^2 are given by (a) $w_1^2 = 1$ and $w_2^2 = 0$, (b) $w_1^2 = \frac{1}{2}$ and $w_2^2 = \frac{1}{2}$, (c) $w_1^2 = \frac{1}{5}$ and $w_2^2 = \frac{4}{5}$. The corresponding ratios of the "signal" variance to the total variance of $\mathbf{Y}$ are 1, $\frac{1}{2}$ and $\frac{1}{5}$, respectively. Note that the weight $w_2 = 0$ is not admissible in the mixture model as defined above. The purpose of including it in (a) is to display the pure "signal" component of $\mathbf{X}$.

Typically for a sample path of FBM with a large Hurst parameter, the time series in (a) exhibits long monotone parts and only a small number of changes between "upwards" and "downwards". The addition of an integrated AR(1) process with a small autoregressive coefficient in (b) and (c) leads to an increase of the number of changes. On a larger time scale, however, the appearance of the three time series is very similar.

The graphs (d)-(f) are obtained by computing estimates of $\boldsymbol{H}$ for increasing delays. Let $\boldsymbol{\tau} = (1,1)$, and define

$$\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}} \;:=\; \max\left\{0,\, \log_2 \cos(\pi \hat{c}_n^{N\boldsymbol{\tau}}/2) + 1\right\} \tag{6.6}$$

for $n \in \mathbb{N}$ and $N \in \mathbb{N}$, with $\hat{c}_n^{N\boldsymbol{\tau}}$ as given in (6.3). For $N = 1$, $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ is the estimator of the Hurst parameter discussed in Section 5.3.

Figure 6.1 (d)-(f) shows the estimates obtained for $N = 1, 2, \ldots, 50$. The estimates are obtained for larger samples than displayed in (a)-(c), namely, for samples of size $n = 10\,000$ each. As graph (d) shows, the time series in (a) results in estimates of $\boldsymbol{H}$ approximately equal to 0.75 for any delay. Note that this is in accordance with equation (6.1), which states that, in equidistant discretizations of FBM, ordinal pattern probabilities (and functions thereof) are invariant with respect to a scaling of the delays.

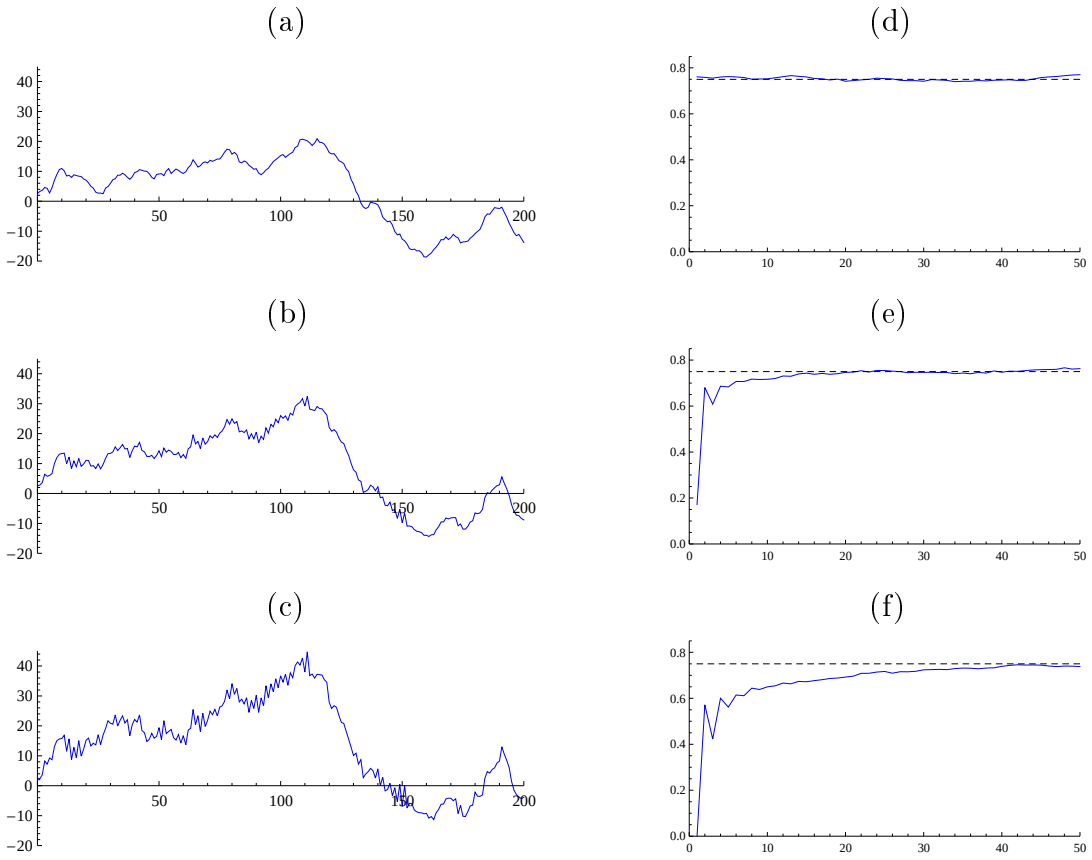

Figure 6.1: Mixture of FBM with the Hurst parameter $\boldsymbol{H} = 0.75$ and an integrated AR(1) process with the autoregressive coefficient $\boldsymbol{a} = -0.75$. The squared weights w_1^2 and w_2^2 are given by (a) $w_1^2 = 1$ and $w_2^2 = 0$, (b) $w_1^2 = \frac{1}{2}$ and $w_2^2 = \frac{1}{2}$, (c) $w_1^2 = \frac{1}{5}$ and $w_2^2 = \frac{4}{5}$. The resulting estimates of $\boldsymbol{H}$ for $N = 1, 2, \ldots, 50$ are shown in (d)-(f).

For the time series in (b) and (c), we obtain different estimates of $\boldsymbol{H}$ depending on the delays. For small delays, the Hurst parameter is underestimated, but as N increases, the estimates tend to 0.75.

The above results show that considering $\mathbf{X}$ on large time scales filters out the short range dependent noise added by the integrated AR(1) process and thus allows to estimate $\boldsymbol{H}$. One could also imagine the opposite situation where $\boldsymbol{H}$ is smaller than $\frac{1}{2}$ (and hence $\mathbf{X}$ does not exhibit long memory), but the addition of an integrated AR(1) process with a large autoregressive coefficient leads to an overestimation of $\boldsymbol{H}$ for small delays. In the following section, we study these phenomenons systematically. The application to the analysis of real-life time series is demonstrated in Section 6.5.

6.4 Asymptotic self-similarity

Throughout this section, any stochastic process is assumed to be defined on a probability space $(\Omega', \mathcal{A}', \mathbb{P})$. In fact, we are only interested in the distribution of the processes, so the structure of $(\Omega', \mathcal{A}')$ is irrelevant.

The renormalization group. Let $\boldsymbol{H} \in (0,1)$. For $N \in \mathbb{N}$, consider the operator $T_{N,\boldsymbol{H}}$ on $\mathbb{R}^{\mathbb{Z}}$ given by

$$T_{N,\boldsymbol{H}}\,\mathbf{z} \;=\; \big((T_{N,\boldsymbol{H}}\,\mathbf{z})_t\big)_{t\in\mathbb{Z}}$$

for $\mathbf{z} = (z_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$, where

$$(T_{N,\boldsymbol{H}}\,\mathbf{z})_t \;:=\; \frac{1}{N^{\boldsymbol{H}}} \sum_{k=(t-1)N+1}^{tN} z_k$$

for $t \in \mathbb{Z}$. Thus, the sequence $T_{N,\boldsymbol{H}}\,\mathbf{z}$ is obtained by summing the components of $\mathbf{z}$ over successive non-overlapping blocks of size N and scaling the sums by $\frac{1}{N^{\boldsymbol{H}}}$.

Since $T_{N,\boldsymbol{H}} \circ T_{M,\boldsymbol{H}} = T_{NM,\boldsymbol{H}}$ for $M, N \in \mathbb{N}$, the set of transformations $\{T_{N,\boldsymbol{H}} \,|\, N \in \mathbb{N}\}$ together with $\circ$ forms a semi-group, which is called *renormalization group with index* $\boldsymbol{H}$ (see Embrechts and Maejima [36], p. 15, Taqqu [96]).

A real-valued stationary process $\mathbf{Z} = (Z_t)_{t\in\mathbb{Z}}$ is called a *fixed point of the renormalization group with index* $\boldsymbol{H}$ iff $\mathbf{Z} \stackrel{\text{dist}}{=} T_{N,\boldsymbol{H}}\,\mathbf{Z}$ for all $N \in \mathbb{N}$. Note that if $\mathbf{Z}$ is almost surely equal to 0, then $\mathbf{Z}$ is a fixed point of the renormalization group with index $\boldsymbol{H}$ for all $\boldsymbol{H} \in (0,1)$. In order to avoid trivialities, we will require that fixed points of renormalization groups are non-degenerate.

A real-valued stationary process $\mathbf{Y} = (Y_t)_{t\in\mathbb{Z}}$ is said to be *asymptotically self-similar with index* $\boldsymbol{H}$ iff there exists a fixed point $\mathbf{Z}$ of the renormalization group with index $\boldsymbol{H}$ such that $T_{N,\boldsymbol{H}}\,\mathbf{Y}$ converges in distribution to $\mathbf{Z}$ as N goes to infinity.

The following result is well-known (see Theorem 2.2.1 in Embrechts and Maejima [36]). Recall that we use $\mathbf{G}_{\boldsymbol{H}}$ to denote the distribution of FGN with the Hurst parameter $\boldsymbol{H}$ (see Section 2.2.3).

Theorem 6.5. *A necessary and sufficient condition for a Gaussian process* $\mathbf{Y}$ *to be asymptotically self-similar with index* $\boldsymbol{H} \in (0,1)$ *is that, as* $N \to \infty$,

$$T_{N,\boldsymbol{H}}\,\mathbf{Y} \;\stackrel{\mathbb{P}}{\longrightarrow}\; \mathbf{G}_{\boldsymbol{H}}\,.$$

Implications for ordinal pattern probabilities. Suppose that $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ is a real-valued process on $(\Omega', \mathcal{A}', \mathbb{P})$ and $\mathbf{Y}$ is the increment process of $\mathbf{X}$, i.e., $Y_t = X_t - X_{t-1}$ for $t \in \mathbb{Z}$. For $d \in \mathbb{N}$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_d) \in \mathbb{N}^d$, let $(\Pi^{\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ denote the process of ordinal patterns of order d and with delays $\boldsymbol{\tau}$ in $\mathbf{X}$. The following corollary shows that if $\mathbf{Y}$ is Gaussian and asymptotically self-similar with index $\boldsymbol{H} \in (0,1)$, the distribution of $(\Pi^{N\boldsymbol{\tau}}(t))_{t\in\mathbb{Z}}$ converges to $\boldsymbol{\Pi}^{\boldsymbol{\tau}}_{\boldsymbol{H}}$ defined in the last paragraph of Section 6.1.

Corollary 6.6. *If $\mathbf{Y}$ is Gaussian and asymptotically self-similar with index $\boldsymbol{H} \in (0,1)$, then, as $N \to \infty$,*

$$(\Pi^{N\boldsymbol{\tau}}(Nt))_{t\in\mathbb{Z}} \xrightarrow{\mathbb{P}} \boldsymbol{\Pi}^{\boldsymbol{\tau}}_{\boldsymbol{H}} .$$

Proof. Let $\boldsymbol{H} \in (0,1)$ and suppose $\mathbf{Y}$ is Gaussian and asymptotically self-similar with index $\boldsymbol{H}$. Let $(\Pi^{\boldsymbol{\tau}}_{\boldsymbol{H}}(t))_{t\in\mathbb{Z}}$ be distributed according to $\boldsymbol{\Pi}^{\boldsymbol{\tau}}_{\boldsymbol{H}}$. Note that, for all $N \in \mathbb{N}$ and $t \in \mathbb{Z}$,

$$\begin{aligned}
\Pi^{N\boldsymbol{\tau}}(Nt) &= \tilde{\pi}^{N\boldsymbol{\tau}}(Y_{Nt+1}, Y_{Nt+2}, \ldots, Y_{Nt+N\boldsymbol{\tau}(d)}) \\
&= \tilde{\pi}^{\boldsymbol{\tau}}(N^{\boldsymbol{H}}(T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+1}, N^{\boldsymbol{H}}(T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+2}, \ldots, N^{\boldsymbol{H}}(T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+\boldsymbol{\tau}(d)}) \\
&= \tilde{\pi}^{\boldsymbol{\tau}}((T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+1}, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+2}, \ldots, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+\boldsymbol{\tau}(d)}) ,
\end{aligned}$$

where the last equality holds because $\tilde{\pi}^{\boldsymbol{\tau}}$ is invariant under positive scaling. According to Theorem 6.5, the distribution of $((T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+1}, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+2}, \ldots, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_{t+\boldsymbol{\tau}(d)})$ converges to that of $\boldsymbol{\tau}(d)$ subsequent values in FGN with the Hurst parameter $\boldsymbol{H}$. Furthermore, for every $\mathbf{r} \in S_d$, the boundary of $\{\mathbf{y} \in \mathbb{R}^{\boldsymbol{\tau}(d)} \,|\, \tilde{\pi}^{\boldsymbol{\tau}}(\mathbf{y}) = \mathbf{r}\}$ in $\mathbb{R}^{\boldsymbol{\tau}(d)}$ has Lebesgue measure 0 (this follows by the same argument showing that the boundary of $\{\mathbf{y} \in \mathbb{R}^d \,|\, \tilde{\pi}(\mathbf{y}) \in \bar{\mathbf{r}}\}$ in $\mathbb{R}^d$ has Lebesgue measure 0, see p. 53). Thus, as $N \to \infty$, the distribution of $\Pi^{N\boldsymbol{\tau}}(Nt)$ converges to that $\Pi^{\boldsymbol{\tau}}_{\boldsymbol{H}}(t)$. Analogously, we obtain that, for all $t_1, t_2, \ldots, t_k \in \mathbb{Z}$ with $k \in \mathbb{N}$, the distribution of $(\Pi^{N\boldsymbol{\tau}}(Nt_1), \Pi^{N\boldsymbol{\tau}}(Nt_2), \ldots, \Pi^{N\boldsymbol{\tau}}(Nt_k))$ converges to that of $(\Pi^{\boldsymbol{\tau}}_{\boldsymbol{H}}(t_1), \Pi^{\boldsymbol{\tau}}_{\boldsymbol{H}}(t_2), \ldots, \Pi^{\boldsymbol{\tau}}_{\boldsymbol{H}}(t_k))$, and hence the statement follows. □

Corollary 6.6 implies that when $\mathbf{Y}$ is Gaussian and asymptotically self-similar with index $\boldsymbol{H} \in (0,1)$, the ordinal pattern probability $\mathbb{P}(\Pi^{N\boldsymbol{\tau}}(t) = \mathbf{r})$ converges to the corresponding probability $\mathbb{P}_{\boldsymbol{H}}(\Pi^{\boldsymbol{\tau}}(t) = \mathbf{r})$ in an equidistant discretization of FBM with the Hurst parameter $\boldsymbol{H}$. For the probability of a change with delays $\boldsymbol{\tau} = (1,1)$, we obtain

$$\lim_{N\to\infty} \mathbb{P}(C^{N\boldsymbol{\tau}}(t) = 1) = 1 - \frac{2}{\pi} \arcsin 2^{\boldsymbol{H}-1}$$

(compare to formula (5.10) on p. 102). Thus, when N is "sufficiently large", $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ given in (6.6) can be regarded as a (slightly) biased estimator for the index of asymptotical self-similarity $\boldsymbol{H}$. Clearly, we do not know a priori how large N has to be in order to obtain an estimator which is only slightly biased. As an a posteriori criterion, the estimates of $\boldsymbol{H}$ should not vary "too much" if we further increase N. Compare to Figure 6.1 and the discussion in Section 6.5.

Conditions for asymptotic self-similarity. The following lemma relates the autocorrelation structure of $\mathbf{Y}$ to the index of asymptotical self-similarity.

Lemma 6.7. *Suppose* $\mathbf{Y}$ *is non-degenerate stationary Gaussian with zero means and unit variance. For* $k \in \mathbb{Z}$, *let* $\rho(k) := \mathrm{Corr}(Y_0, Y_k)$. *A sufficient condition for* $\mathbf{Y}$ *to be asymptotically self-similar with index* $\boldsymbol{H} \in (0,1)$ *is given as follows:*

(*i*) *For* $\boldsymbol{H} \in (0, \frac{1}{2})$: *there exists a* $c > 0$ *with*

$$\rho(k) \sim -ck^{2\boldsymbol{H}-2} \quad \textit{and} \quad \sum_{k=-\infty}^{\infty} \rho(k) = 0\,.$$

(*ii*) *For* $\boldsymbol{H} = \frac{1}{2}$: *there exists a* $c > 0$ *with*

$$\sum_{k=1}^{\infty} |\rho(k)| < \infty \quad \textit{and} \quad \sum_{k=-\infty}^{\infty} \rho(k) = c\,.$$

(*iii*) *For* $\boldsymbol{H} \in (\frac{1}{2}, 1)$: *there exists a* $c > 0$ *with*

$$\rho(k) \;\sim\; ck^{2\boldsymbol{H}-2}\,.$$

Proof. See Example 3.2 in Taqqu [96]. □

The following lemma shows that the sum of independent Gaussian asymptotically self-similar processes is again asymptotically self-similar, and the index of the sum process is the maximum of the indices of the summand processes.

Lemma 6.8. *Let* $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(m)}$ *with* $m \in \mathbb{N}$ *be independent Gaussian asymptotically self-similar processes with indices* $\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_m \in (0,1)$. *Then* $\mathbf{Y} := \sum_{l=1}^{m} \mathbf{Y}^{(l)}$ *is asymptotically selfsimilar with index* $\boldsymbol{H} = \max\{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_m\}$.

Proof. Note that it is sufficient to show the statement for $m = 2$. By induction, we then obtain the statement for $m > 2$. According to Theorem 6.5, in order to prove that $T_{N,\boldsymbol{H}}\,\mathbf{Y}$ converges in distribution to FGN with the Hurst parameter $\boldsymbol{H}$ as $N \to \infty$, we only need to show the existence of a $\sigma^2 > 0$ such that

$$\lim_{N\to\infty} \mathrm{Cov}\big((T_{N,\boldsymbol{H}}\,\mathbf{Y})_0,\, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_k\big) \;=\; \frac{\sigma^2}{2}\big(\,|k+1|^{2\boldsymbol{H}} - 2|k|^{2\boldsymbol{H}} + |k-1|^{2\boldsymbol{H}}\,\big) \tag{6.7}$$

for every $k \in \mathbb{Z}$ (see Theorem 2.6 (*iii*) and the definition of FGN in Section 2.2.3). Let $k \in \mathbb{Z}$. Since $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are independent, we have

$$\begin{aligned}\mathrm{Cov}\big((T_{N,\boldsymbol{H}}\,\mathbf{Y})_0,\, (T_{N,\boldsymbol{H}}\,\mathbf{Y})_k\big) \;&=\; \mathrm{Cov}\big((T_{N,\boldsymbol{H}}\,\mathbf{Y}^{(1)})_0,\, (T_{N,\boldsymbol{H}}\,\mathbf{Y}^{(1)})_k\big)\\ &+\; \mathrm{Cov}\big((T_{N,\boldsymbol{H}}\,\mathbf{Y}^{(2)})_0,\, (T_{N,\boldsymbol{H}}\,\mathbf{Y}^{(2)})_k\big)\end{aligned}$$

for every $N \in \mathbb{N}$. Furthermore, according to the asymptotic self-similarity of $\mathbf{Y}^{(l)}$ for $l = 1, 2$, we obtain that

$$\lim_{N\to\infty} \mathrm{Cov}\big((T_{N,\boldsymbol{H}_l}\,\mathbf{Y}^{(l)})_0,\,(T_{N,\boldsymbol{H}_l}\,\mathbf{Y}^{(l)})_k\big) \;=\; \frac{1}{2}\big(\,|k+1|^{2\boldsymbol{H}_l} - 2|k|^{2\boldsymbol{H}_l} + |k-1|^{2\boldsymbol{H}_l}\,\big)\,.$$

Thus, if $\boldsymbol{H}_1 = \boldsymbol{H}_2 = \boldsymbol{H}$, then (6.7) holds with $\sigma^2 = 2$. Now, without loss of generality, suppose that $\boldsymbol{H}_1 < \boldsymbol{H}_2$ and thus $\boldsymbol{H} = \boldsymbol{H}_2$. Since

$$\mathrm{Cov}\big((T_{N,\boldsymbol{H}_2}\,\mathbf{Y}^{(1)})_0,\,(T_{N,\boldsymbol{H}_2}\,\mathbf{Y}^{(1)})_k\big) \;=\; \frac{1}{N^{2(\boldsymbol{H}_2-\boldsymbol{H}_1)}}\,\mathrm{Cov}\big((T_{N,\boldsymbol{H}_1}\,\mathbf{Y}^{(1)})_0,\,(T_{N,\boldsymbol{H}_1}\,\mathbf{Y}^{(1)})_k\big)$$

for every $N \in \mathbb{N}$, we obtain

$$\lim_{N\to\infty} \mathrm{Cov}\big((T_{N,\boldsymbol{H}_2}\,\mathbf{Y}^{(1)})_0,\,(T_{N,\boldsymbol{H}_2}\,\mathbf{Y}^{(1)})_k\big) \;=\; 0\,.$$

Consequently, (6.7) holds with $\sigma^2 = 1$. The proof is complete. □

The following lemma gives examples of asymptotically self-similar Gaussian processes.

Lemma 6.9.

(*i*) *If* $\mathbf{Y}$ *is* FGN *with the Hurst parameter* $\boldsymbol{H} \in (0,1)$*, then* $\mathbf{Y}$ *is asymptotically self-similar with index* $\boldsymbol{H}$*.*

(*ii*) *If* $\mathbf{Y}$ *is an* ARFIMA(0,d,0) *process with the fractional differencing parameter* $\boldsymbol{d} \in (-\frac{1}{2}, \frac{1}{2})$*, then* $\mathbf{Y}$ *is asymptotically self-similar with index* $\boldsymbol{d} + \frac{1}{2}$*.*

(*iii*) *If* $\mathbf{Y}$ *is an* AR(1) *process with the autoregressive coefficient* $\boldsymbol{a} \in (-1,1)$*, then* $\mathbf{Y}$ *is asymptotically self-similar with index* $\frac{1}{2}$*.*

Proof. (*i*) See Theorem 2.2.1 in Embrechts and Maejima [36]). (*ii*) and (*iii*) follow by Lemma 6.7, Lemma 2.8 and Lemma 2.9, respectively. □

Now we have a mathematical explanation for the findings in Figure 6.1. According to Lemma 6.9, the processes $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are asymptotically self-similar with the indices $\boldsymbol{H}_1 = 0.75$ and $\boldsymbol{H}_2 = 0.5$, respectively. Therefore, Lemma 6.8 implies $\mathbf{Y} = \mathbf{Y}^{(1)} + \mathbf{Y}^{(2)}$ is asymptotically self-similar with the index $\boldsymbol{H} = 0.75$, and Corollary 6.6 shows that if N is sufficiently large, $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ is an only slightly biased estimator of $\boldsymbol{H}$. In fact, how large N needs to be depends on the ratio of the signal variance to the total variance.

6.5 Application to practical time series

Let us demonstrate the application of our method to the analysis of two real life time series: the yearly minimal water levels of the Nile River at the Roda Gauge and the NBS high precision weight measurements. Note that both data sets are well-known examples of long memory processes. (see Beran [15], pp. 20-29).

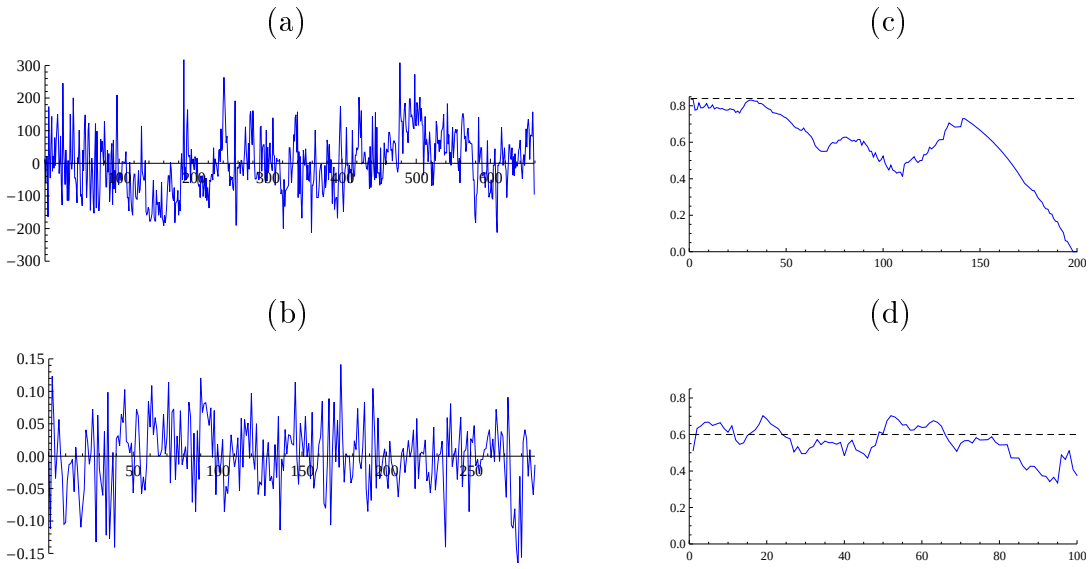

Figure 6.2: (a) Nile River minima. (b) NBS weight measurements. (c) and (d) Resulting estimates of the Hurst parameter.

Nile River minima. The yearly Nile River minima for the years 622-1281 recorded at the Roda Gauge near Cairo is one of the most famous data sets from hydrology. Figure 6.2 (a) shows this time series with the mean subtracted. Remarkably, there are long periods of "dryness" and "floods" where the measurements stay below and above 0, respectively, for many successive years. Evidence for long memory in the data has been first presented by the hydrologist H. E. Hurst. For a long time, the definition of a model which explains this phenomenon has been an open problem. B. M. Mandelbrot and his co-authors have introduced FGN exactly for this purpose. Therefore, the parameter $\boldsymbol{H}$ in FGN is now commonly referred to as the Hurst parameter (see Beran [15], pp. 32-34).

In fact, by a first diagnosis, there is no evidence against modelling the Nile River data by FGN: The time series does not exhibit a relevant deviation from (univariate) normality. Also, the spectral density of FGN gives a good fit of the periodogram of the Nile River data, where the best model fit is obtained for the Hurst parameter $\boldsymbol{H} = 0.84$ (in Figure 6.2 (b), this value is displayed by the black dashed line). The most important argument against using FGN is some evidence of non-stationarity. For details, see Beran [15], pp. 117-118, 125-126.

Figure 6.2 (b) shows the estimates of $\boldsymbol{H}$ obtained by $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ for $N = 1, 2, \ldots, 200$. For small N, the estimates are almost identical and close to the best model fit $\boldsymbol{H} = 0.84$. As N increases, the variation becomes larger and, finally, the sequence of estimates tends to 0. Note that we obtain qualitatively very similar graphs for simulations of FGN with the Hurst parameter $\boldsymbol{H} = 0.84$. For large N, the variance of the estimates increases due to a

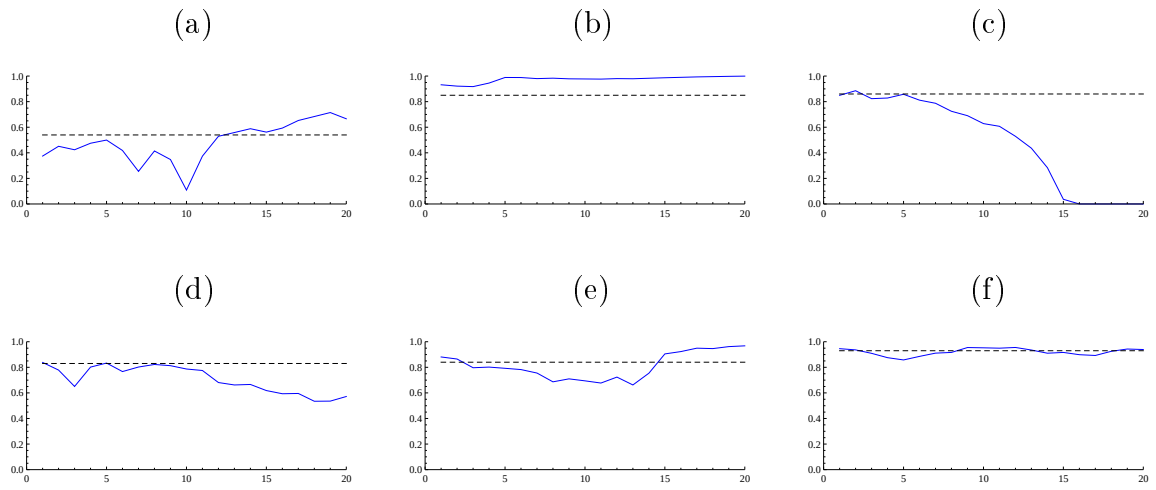

Figure 6.3: Estimates of the Hurst parameter for the Nile River data in successive non-overlapping time windows of length $n = 100$ each.

decreasing number of sample data points.

Now, consider the six time series obtained by dividing the first 600 observations of the Nile River data into parts of length 100. The best model fit of these time series is obtained for $\boldsymbol{H} = 0.54$, 0.85, 0.86, 0.83, 0.84 and 0.93 (see Beran [15], p. 207). Figure 6.2 (a)-(f) shows the estimates obtained by $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ for $N = 1, 2, \ldots, 20$. For small N, these estimates are close to the best model fit (except for the time series (a) and (b), where $\boldsymbol{H}$ is under- and overestimated, respectively). Given the small sample sizes, we find the variation of the estimates is surprisingly low.

NBS data. The NBS data set is given by high precision weight measurements of a 1kg standard weight, performed between June 1963 and October 1975 at the National Institute of Standards and Technology, Gaithersburg, USA. Figure 6.1 (c) displays this time series with the mean subtracted. Note that the measurements were not performed at equidistant dates. Treating them as if they were performed at equidistant dates, however, is likely to only slightly affect the long-range dependence structure (see Beran [15], p. 27). The data set is particularly interesting because it exhibits features of a long memory time series, even though the experimental setting was almost perfectly designed to obtain independent observations.

The HUBINC estimator yields $\boldsymbol{H} = 0.60$ as an estimate of the Hurst parameter, with an approximate standard deviation of 0.044 (see Beran [15], p. 140). Here, $\boldsymbol{H}$ is not necessarily interpreted as the Hurst parameter in FGN but, more generally, as the parameter in a model for long memory time series determining the rate of decay of the autocovariances.

Figure 6.1 (d) displays the estimates of $\boldsymbol{H}$ obtained by $\hat{\boldsymbol{H}}_n^{N\boldsymbol{\tau}}$ for $N = 1, 2, \ldots, 150$. The estimates are all close to the value $\boldsymbol{H} = 0.60$, except for very large values of N. After all, the variation of the estimates is small, in particular, smaller than the variation which we observe for simulations of FGN with the Hurst parameter $\boldsymbol{H} = 0.60$.

List of symbols

$\mathbb{B}(\mathcal{Y})$	Borel σ-field of the topological space $\mathcal{Y}$
$\sharp A$	cardinality of a set A
$A \Delta B$	symmetric difference of sets A and B
$\emptyset$	the empty set
∂A	boundary of a subset A of a topological space
$\max(A)$	maximimum of a subset $A \subset \mathbb{R}$
$\min(A)$	minimum of a subset $A \subset \mathbb{R}$
$\inf(A)$	infimum of a subset $A \subset \mathbb{R}$
$\sup(A)$	supremum of a subset $A \subset \mathbb{R}$
$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	set of natural numbers including 0
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integers
$\mathbf{A}^T$	transpose of the matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of the matrix $\mathbf{A}$
$\mathbf{diag}(a_1, a_2, \ldots, a_n)$	$n \times n$-diagonal matrix with the entries $a_1, a_2, \ldots, a_n$ on the diagonal
$\Gamma(x)$	value of the Gamma function at x
$\mathrm{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	(multidimensional) normal distribution with means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\overset{\text{dist}}{=}$	equality in distribution
$\xrightarrow{\mathbb{P}}$	convergence in distribution with respect to the probability measure $\mathbb{P}$
$\pi(\mathbf{x})$	ordinal pattern of the vector $\mathbf{x} \in \mathbb{R}^{d+1}$
$\tilde{\pi}(\mathbf{y})$	ordinal pattern of the vector of increments $\mathbf{y} \in \mathbb{R}^d$
$\boldsymbol{G_H}$	distribution of FGN with the Hurst parameter $\boldsymbol{H}$
$\boldsymbol{\Pi_H}$	distribution of ordinal patterns in FGN with the Hurst parameter $\boldsymbol{H}$
$\boldsymbol{\Pi_H^\tau}$	distribution of ordinal patterns with delays $\boldsymbol{\tau}$ in FGN with the Hurst parameter $\boldsymbol{H}$

Bibliography

[1] Abry, P. and Veitch, D. (1998): Wavelet Analysis of Long-Range-Dependent Traffic. *IEEE Transactions on Information Theory* 44, 1, 2-15.

[2] Abrahamson, I. G. (1964): Orthant probabilities for the quadrivariate normal distribution. *Ann. Math. Statist.* 35, 1685-708.

[3] Amigó, J. M. and Kennel, M. B. (2007): Topological permutation entropy. *Physica D* 231, 2, 137-42.

[4] Amigó, J. M., Kennel, M. B. and Kocarev, L. (2005): The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. *Physica D* 210, 1, 77-95.

[5] Amigó, J. M., Kocarev, L. and Tomovski, I. (2007): Discrete entropy. *Physica D* 228, 1, 77-85.

[6] Amigó, J. M., Zambrano, S. and Sanjuán, M. A. F. (2007): True and false forbidden patterns in deterministic and random dynamics. *Europhysics Letters* 79, 50001.

[7] Arcones, M. A. (1994): Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Annals of Probability* 22, 2242-74.

[8] Bacon, R. H. (1963): Approximation to multivariate normal orthant probabilities. *Ann. Math. Statist.* 34, 191-98.

[9] Bandt, C., Keller, G. and Pompe, B. (2002): Entropy of interval maps via permutations. *Nonlinearity* 15, 1595-602.

[10] Bandt, C. and Pompe, B. (2002): Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 88, 174102.

[11] Bandt, C. (2005): Ordinal time series analysis. *Ecological Modelling* 182, 229-38.

[12] Bandt, C. and Shiha, F. (2007): Order patterns in time series. *J. Time Ser. Anal.* 28, 646-65.

[13] Bardet, J.-M., Lang, G., Oppenheim, G., Phillipe, A. and Taqqu, M. S. (2003): *Generators of Long-Range Dependent Processes: A Survey.* In [35].

[14] Bauer, H. (1996): *Probability Theory.* de Gruyter, Berlin.

[15] Beran, J. (1994): *Statistics for Long-Memory Processes.* Chapman and Hall, London.

[16] Berman, S. M. (1992): *Sojourns and Extremes of Stochastic Processes.* Wadsworth and Brooks/Cole, Pacific Grove, California.

[17] Billingsley, P. (1995): *Probability and Measure. Third Edition.* Wiley, New York.

[18] Bretschneider, M., Kreuzer, M., Drexler, B., Hentschke, H., Antkowiak, B., Schwarz, C., Kochs, E. and Schneider, G. (2006): Coherence of in vitro and in vivo field potential activity and EEG, detected by ordinal analysis. *Journal of Neurosurgical Anesthesiology*, 18, 309.

[19] Brockwell, P. J. and Davis, R. A. (1991): *Time series: Theory and Methods, Second Edition.* Springer, New York.

[20] Bronshtein, I. N. and Semendyayev, K. A. (1979): *Handbook of Mathematics.* Springer, Berlin.

[21] Bruzzo, A. A., Gesierich, B., Santi, M., Tassinari, C. A., Birbaumer, N. and Rubboli, G. (2008): Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. A preliminary study. *Neurological Sciences* 29, 1, 3-9.

[22] Cammarota, C. and Rogora, E. (2005): Independence and symbolic independence of heartbeat intervals during atrial brillation. *Physica A* 353, 323-35.

[23] Cammarota, C. and Rogora, E. (2005): Testing independence in time series via universal distributions of permutations and words. *Int. J. Bifurcation Chaos* 15, 1-9.

[24] Cao, Y. H., Tung, W. W., Gao, J. B., Protopopescu, V. A. and Hively, L. M. (2004): Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* 70, 046217.

[25] Casella, G. and Berger, R. L. (2002): *Statistical Inference, 2nd edition.* Duxbury, California.

[26] Chang, S., Pihl, G. E. and Essigmann, M. W. (1951): Representations of speech sounds and some of their statistical properties. *Proc. IRE* 39, 147-53.

[27] Cheng, M. C. (1969): The orthant probability of four Gaussian variates. *Ann. Math. Statist.* 40, 152-61.

[28] Coeurjolly, J. F. (2000): Simulation and identification of the fractional Brownian motion: A bibliographical and comparative study. *J. Stat. Software* 5.

[29] Cornfeld, I. P., Fomin, S. V. and Sinai, Y. G. (1982): *Ergodic Theory*. Springer, Berlin.

[30] Craig, P. (2008): A new reconstruction of multivariate normal orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, 1, 227-43.

[31] Damsleth, E. and El-Shaarawi, A. H. (1988): Estimation of autocorrelation in a binary time series. *Stochastic Hydrol. Hydraul.* 2, 61-72.

[32] David, F. N. (1953): A note on the evaluation of the multivariate normal integral. *Biometrika* 40, 458-9.

[33] Davies, R. B. and Harte, D. S. (1987): Tests for Hurst effect. *Biometrika* 74, 95-101.

[34] Doukhan, P. (2003): *Models, Inequalities, and Limit Theorems for Stationary Sequences*. In [35].

[35] Doukhan, P., Oppenheim, G. and Taqqu, M. S. (Editors) (2003): *Theory and Applications of Long-Range-Dependence*. Birkhäuser, Boston.

[36] Embrechts, P. and Maejima, M. (2002): *Selfsimilar Processes*. Princeton University Press, New Jersey.

[37] Engelking, R. (1989): *General topology*. Heldermann, Berlin.

[38] Ewing, G. and Taylor, J. (1969): Computer recognition of speech using zero-crossing information. *IEEE Transactions on Audio and Electroacoustics* 17, 1, 37-40.

[39] Faul, S., Boylan, G., Connolly, S., Marnane, W. and Lightbody, G. (2005): Chaos theory analysis of the newborn EEG: Is it worth the wait? *IEEE International Workshop on Intelligent Signal Processing*, Faro (Portugal), 381-6.

[40] Ferguson, S. T., Genest, C. and Hallin, M. (2000): Kendall's tau for serial dependence. *Canadian J. Statist.* 28, 587-604.

[41] Frank, B., Pompe, B., Schneider, U. and Hoyer, D. (2006): Permutation entropy improves fetal behavioural state classification based on heart rate analysis from biomagnetic recordings in near term fetuses. *Med. Biol. Eng. Comput.* 44, 179-87.

[42] Galka, A. (2000): *Topics in Nonlinear Time Series Analysis. With Implications for EEG Analysis*. World Scientific, Singapore.

[43] Garel, B. and Hallin, M. (1999): Rank-based autoregressive order indentification. *J. Am. Stat. Assoc.* 94, 1357-71.

[44] Grassberger, P. and Procaccia, I. (1983): Measuring the strangeness of strange attractors. *Physica D* 9, 189-208.

[45] Groth, A. (2005): Visualization of coupling in time series by order recurrence plots. *Phys. Rev. E* 72, 046220.

[46] Hallin, M. and Jurečkova, J. (1999): Optimal tests for autoregressive models based on autoregression rank scores. *Annals Stat.* 27, 1385-414.

[47] Härdle, W. and Simar, L. (2003): *Applied Multivariate Statistical Analysis.* Springer, New York.

[48] Ho, H.-C. and Sun, T.-C. (1987): A central limit theorem for noninstantaneous filters of a stationary Gaussian process. *J. Multivariate Anal.* 22, 144-55.

[49] Horn, R. A. and Johnson, C. R. (1985): *Matrix Analysis*, Cambridge University Press, Cambridge.

[50] Hosking, J. R. M. (1996): Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics* 73, 261-84.

[51] Iber, C., Ancoli-Israel, S., Chesson, A., Quan, S. (2007): *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications.* American Academy of Sleep Medicine, Westchester.

[52] Kedem, B. (1980): *Binary Time Series.* Marcel Dekker, New York.

[53] Kedem, B. (1994): *Time Series Analysis by Higher Order Crossings.* IEEE Press, New York.

[54] Keenan, D. and MacRae, A. (1982): Time Series Analysis of Binary Data. *J. Am. Stat. Assoc.* 77, 380, 816-21.

[55] Keller, K. and Lauffer, H. (2003): Symbolic analysis of high-dimensional time series, *Int. J. Bifurcation Chaos* 13, 2657-68.

[56] Keller, K. and Wittfeld, K. (2004): Distances of time series components by means of symbolic dynamics, *Int. J. Bifurcation Chaos* 14, 693-704.

[57] Keller, K. and Sinn, M. (2005): Ordinal analysis of time series, *Physica A* 356, 114-20.

[58] Keller, K., Lauffer, H. and Sinn, M. (2007): Ordinal analysis of EEG time series, *Chaos and Complexity Letters* 2, 247-58.

[59] Keller, K. and Sinn, M. (2007): Ordinal pattern distributions in dynamical systems. *Proceedings ICDEA*, *submitted*.

[60] Keller, K., Sinn, M. and Emonds, J. (2007): Time series from the ordinal viewpoint. *Stochastics and Dynamics* 2, 247-72.

[61] Keller, K. and Sinn, M. (2009): Kolmogorov-Sinai entropy from the ordinal viewpoint. *submitted*.

[62] Keller, K. and Sinn, M. (2009): A standardized approach to the Kolmogorov-Sinai entropy. *submitted*.

[63] Kettani, H. and Gubner, J. A. (2006): A novel approach to the estimation of the Hurst parameter in self-similar traffic. *IEEE Trans. Circuits Syst. II*, 53, 463-7.

[64] Kitchens, B. P. (1998): *Symbolic Dynamics. One-sided, Two-sided and Countable State Markov Shifts.* Springer, Berlin.

[65] Larrondo, H. A., Martin, M. T., Gonzalez, C. M., Plastino, A. and Rosso, O. A. (2006): Random number generators and causality. *Physics Letters A* 352, 4-5, 421-5.

[66] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983): *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

[67] Lehmann, E. (1986): *Testing Statistical Hypotheses, Second Edition.* Springer, New York.

[68] Lehmann, E. (1999): *Elements of Large Sample Theory.* Springer, New York.

[69] Li, X., Cui, S. and Voss, L. J. (2008): Using permutation entropy to measure the electroencephalographic effects of sevoflurane. *Anesthesiology* 109, 448-56.

[70] Li, X., Ouyang, G. and Richards, D. A. (2007): Predictability analysis of absence seizures with permutation entropy. *Epilepsy research* 77, 70-4.

[71] Mandelbrot, B. (1997): *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk.* Springer, New York.

[72] Mantegna, R. N. and Stanley, H. E. (1996): Turbulence and financial markets. *Nature* 383, 587-88.

[73] Marković, D. and Koch, M. (2005): Sensitivity of Hurst parameter estimation to periodic signals in time series and filtering approaches. *Geophysical Research Letters* 32, L17401.

[74] Matilla-Garcia, M. (2007): A non-parametric test for independence based on symbolic dynamics. *Journal of Economic Dynamics and Control* 31, 12, 3889-903.

[75] Misiurewicz, M. (2003): Permutations and topological entropy for interval maps. *Nonlinearity* 16, 971-6.

[76] Molchan, G. M. (2003): *Historical comments related to fractional Brownian motion.* In [35].

[77] Olofsen, E., Sleigh, J. W. and Dahan, A. (2008): Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *British Journal of Anaesthesia* 101, 6, 810-21.

[78] Olver, F. W. J. (1997): *Asymptotics and Special Functions.* A. K. Peters, Natick, Massachusetts.

[79] Pfanzagl, J. (1994): *Parametric Statistical Theory.* De Gruyter, Berlin.

[80] Piterbarg, V. I. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* American Mathematical Society, Providence, Rhode Island.

[81] Plackett, R. L. (1954): A reduction formula for normal multivariate integrals. *Biometrika* 41, 351-60.

[82] Rabiner, L. R. and Schafer, R. W. (1978): *Digital processing of speech signals.* Prentice-Hall, London.

[83] Rechtschaffen, A. and Kales, A. (1968): *A manual on standardized terminology, techniques and scoring system for sleep stages of human subjects.* National Institute of Health Publications 204, US Gouvernment Printing Office, Washington DC.

[84] Rice, S. O. (1944): Mathematical analysis of random noise. *Bell Sys. Tech. Jour.* 23, 282-332.

[85] Robinson, P. M. (1995): Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of Statistics* 23, 5, 1630-61.

[86] Robinson, P. M. (Editor) (2003): *Time Series with Long Memory*, Oxford University Press, New York.

[87] Rosso, O. A., Zunino, L., Perez, D. G., Figliola, A., Larrondo, H. A., Garavaglia, M., Martin, M. T. and Plastino, A. (2007): Extracting features of Gaussian self-similar stochastic processes via the Bandt-Pompe approach. *Phys. Rev. E* 76, 061114.

[88] Salas, J. D., Chung, C. and Cancelliere, A. (2005): Correlations and crossing rates of periodic-stochastic hydrologic processes. *Journal of hydrologic engineering* 10, 4, 278-87.

[89] Samorodnitsky, G. (2006): Long range dependence. *Foundations and Trends in Stochastic Systems* 1, 3, 163-257.

[90] Shi, B., Vidakovic, B., Katul, G. and Albertson, J. D. (2005): Assessing the effects of atmospheric stability on the fine structure of surface layer turbulence using local and global multiscale approaches. *Physics of Fluids* 17, 055104.

[91] Shiha, F. A. (2004): *Distributions of order patterns in time series.* Ph.D. dissertation, Greifswald.

[92] Sinn, M. and Keller, K. (2008): Covariances of zero-crossings in Gaussian processes. *Submitted.*

[93] Staniek, M. and Lehnertz, K. (2007): Parameter selection for permutation entropy measurements. *Int. J. Bifurcation Chaos* 17, 3729-33.

[94] Surgailis, D. (2003): *CLTs for Polynomials of Linear Sequences: Diagram Formula with Illustrations.* In [35].

[95] Taqqu, M. S. (1975): Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* 31, 287-302.

[96] Taqqu, M. S. (2003): *Fractional Brownian Motion and Long-Range-Dependence.* In [35].

[97] Taqqu, M. S., Teverovsky, W. and Willinger, W. (1995): Estimators for long-range dependence: an empirical study. *Fractals* 3, 4, 785-98.

[98] Veisi, I., Pariz, N. and Karimpour, A. (2007): Fast and robust detection of epilepsy in noisy EEG signals using permutation entropy. *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, 200-3.

[99] Walters, P. (1982): *An Introduction to Ergodic Theory.* Springer, New York.