Ordinal Patterns: Entropy Concepts and Dependence Between Time Series

DISSERTATION
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
M.Sc. Angelika Silbernagel

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen

Siegen, Oktober 2024

gedruckt auf alterungsbeständigem holz- und säurefreiem Papier

Betreuer und erster Gutachter:
Prof. Dr. Alexander Schnurr
Department Mathematik
Universität Siegen
57068 Siegen, Germany

Zweiter Gutachter:
Prof. Dr. Christian H. Weiß
Fächergruppe Mathematik/Statistik
Helmut-Schmidt-Universität
22008 Hamburg, Germany

Tag der mündlichen Prüfung: 07.02.2025

Für meine Großmutter

Abstract

Since their introduction, ordinal patterns have proven to be a powerful tool not only in the context of dynamical systems, but also in time series analysis. Even though working with ordinal patterns leads to a loss of information, they bring many advantages which justify this loss. In this thesis, we contribute to ordinal pattern analysis in various ways.

With regard to the basics, we provide a comparative analysis of different representations of (multivariate generalizations of) ordinal patterns. Furthermore, we give a historical overview of the applications of ordinal patterns in data analysis and mathematical statistics. However, since there is already an extensive amount of literature available, we do not claim completeness.

Afterwards, we consider a specific measure of complexity in a time series or dynamical system, namely the symbolic correlation integral. We investigate it by providing limit theorems for an estimator of this quantity which is based on U-statistics under the assumption of short-range dependence. This also covers limit theorems for the Rényi-2 permutation entropy due to the close relation between these two. To this end, we slightly generalize existing limit theorems in the framework of approximating functionals. Afterwards we derive an estimator for the limit variance to lay the foundation for possible hypothesis tests.

Then, we turn our attention from the structure *within* a univariate time series to the structure *between the components* of a bivariate time series. Ordinal pattern dependence has been introduced in order to capture how strong the co-movement between two data sets or two time series is. Betken et al. (2021) aimed to show that ordinal pattern dependence fits into the axiomatic framework for multivariate measures of dependence between random vectors of the same dimension which had been proposed by Grothe et al. (2014). We reconsider the results by Betken et al. (2021). We show that there is an error with regard to the concordance ordering and that this cannot be verified in general for ordinal pattern dependence. Furthermore, we show that ordinal pattern dependence satisfies a modified set of axioms instead. In addition, we also consider ordinal pattern dependence in the context of supermodular ordering.

Finally, we prove general limit theorems for the distributions of multivariate ordinal patterns under the assumption of not only serial but also componentwise independence. We use our results to propose novel tests for cross-dependence. These include a test based on ordinal pattern dependence. We compare their performance with three competitors, namely classical Pearson's and Spearman's correlations and Chatterjee's correlation coefficient. To this end, we conduct a comprehensive simulation study. Two real-world data examples complete this thesis.

Zusammenfassung

Seit ihrer Einführung haben sich ordinale Muster nicht nur im Zusammenhang mit dynamischen Systemen, sondern auch in der Zeitreihenanalyse als ein mächtiges Werkzeug erwiesen. Auch wenn die Arbeit mit ordinalen Mustern zu einem Informationsverlust führt, bringen sie viele Vorteile mit sich, die diesen Verlust rechtfertigen. In dieser Dissertation leisten wir auf verschiedene Weise einen Beitrag zur ordinalen Muster-Analyse.

Im Hinblick auf die Grundlagen geben wir eine komparative Analyse verschiedener Repräsentationen von (multivariaten Verallgemeinerungen von) ordinalen Mustern. Außerdem geben wir einen historischen Überblick über die Anwendungen von ordinalen Mustern in der Datenanalyse und der mathematischen Statistik. Da aber bereits eine Fülle an Literatur existiert, erheben wir keinen Anspruch auf Vollständigkeit.

Anschließend betrachten wir ein spezifisches Maß für die Komplexität in einer Zeitreihe oder einem dynamischen System, nämlich das sogenannte symbolische Korrelationsintegral. Wir untersuchen es, indem wir Grenzwertsätze unter der Annahme der Kurzzeitabhängigkeit für einen auf U-Statistiken basierenden Schätzer dieser Größe liefern. Dies schließt auch Grenzwertsätze für die Rényi-2-Permutationsentropie ein, da diese beiden Größen eng miteinander verwandt sind. Zu diesem Zweck verallgemeinern wir bestehende Grenzwertsätze im Rahmen von approximierenden Funktionalen leicht. Anschließend leiten wir einen Schätzer für die Varianz im Grenzwert her, um damit die Grundlage für mögliche Hypothesentests zu legen.

Anschließend richten wir unsere Aufmerksamkeit von der Struktur *innerhalb* einer univariaten Zeitreihe auf die Struktur *zwischen den Komponenten* einer bivariaten Zeitreihe. Die ordinale Muster-Abhängigkeit wurde eingeführt, um zu ermitteln, wie stark die Ko-Bewegung zwischen zwei Datensätzen oder zwei Zeitreihen ist. Betken et al. (2021) wollten zeigen, dass die ordinale Muster-Abhängigkeit in den von Grothe et al. (2014) vorgeschlagenen axiomatischen Rahmen für multivariate Abhängigkeitsmaße zwischen Zufallsvektoren der gleichen Dimension passt. Wir revidieren die Ergebnisse von Betken et al. (2021). Wir zeigen, dass es einen Fehler in Bezug auf die Konkordanzordnung gibt und dass diese im Allgemeinen nicht für ordinale Muster-Abhängigkeit verifiziert werden kann. Außerdem zeigen wir, dass die ordinale Muster-Abhängigkeit stattdessen einen modifizierten Satz von Axiomen erfüllt. Darüber hinaus betrachten wir die ordinale Muster-Abhängigkeit im Kontext der supermodularen Ordnung.

Schließlich beweisen wir allgemeine Grenzwertsätze für die Verteilungen von multivariaten Verallgemeinerungen ordinaler Muster unter der Annahme von nicht nur serieller, sondern auch komponentenweiser Unabhängigkeit. Wir nutzen unsere Ergebnisse, um neue Tests für Kreuzabhängigkeit vorzuschlagen. Darunter ist ein auf ordinaler Muster-Abhängigkeit basierender Test. Wir vergleichen die Performanz unserer Tests mit drei Konkurrenten, nämlich den klassischen Korrelationen von Pearson und Spearman sowie dem Korrelationskoeffizienten von Chatterjee. Zu diesem Zweck führen wir eine umfassende Simulationsstudie durch. Zwei reale Datenbeispiele vervollständigen diese Arbeit.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Alexander Schnurr. He has been a constant source of encouragement and was always there for advice and support. Many thanks go to him and my other co-authors, Christian Weiß and Manuel 'Manolo' Ruiz Marín, for successful collaborations, insightful discussions and helpful advice at all times. I deeply appreciate their patience when helping me with numerous problems and challenges, especially when dealing with POS - particularly outstanding science. A rogue who thinks POS means something else...

Special thanks go to Alfred Müller for his insights and helpful advice on stochastic order relations and dependence in general, as well as Annika Betken and Giorgio Micali. I am very thankful for the discussions and their valuable feedback on the proof regarding the shift-invariance of the r-approximating condition. Furthermore, I would like to express thanks to Nils Heerten, Katharina Hees, Robin Mey and Manuel Schäfer for their valuable comments and proofreading.

I am also grateful to the "Deutsche Forschungsgemeinschaft (DFG)" for funding my research. Moreover, I would like to especially thank Annika Betken and Herold Dehling for planting the idea of doing a PhD in my head. I would not have done this without them.

Finally, my greatest thanks go to my family and friends, for their infinite patience and care, especially during the last months - I know they have not always had it easy with me. Even though I cannot name everyone individually, I would like to make a few exceptions: I am grateful to my 'parents-in-law' Beate and Jürgen for their faith in me and the encouragement to strive for more. There are never enough words to thank my parents Ida and Johann for their unconditional love and support. Even if I did not ask for help, they always knew when it was needed - a kind of seventh sense that only parents have.

Lastly, I would like to thank my boyfriend Tom: For every paper he proofread, for each of my rehearsed talks he had to listen to and of course for proofreading this dissertation of mine - all in spite of being an algebraist (i.e., on the dark side of mathematics) and therefore having nothing to do with stochastics. I know you are reading this, so thank you for always listening to my blabbering and enduring me all these years. Thank you for cheering me up whenever I feel down. Thank you for being you. I hope I have not hidden to manny mistakes.

Danksagung

Zuallererst möchte ich meinem Betreuer, Alexander Schnurr, meinen Dank aussprechen. Er war eine stetige Quelle der Ermutigung und stand mir immer mit Rat und Tat zur Seite. Ihm und meinen anderen beiden Co-Autoren, Christian Weiß und Manuel 'Manolo' Ruiz Marín, schulde ich großen Dank für die erfolgreiche Zusammenarbeit, die aufschlussreichen Diskussionen und die hilfreichen Ratschläge zu jeder Zeit. Ich bin ihnen sehr dankbar für ihre Geduld, mit der sie mir bei zahlreichen Problemen und Herausforderungen geholfen haben, vor allem, wenn es um POS ging - 'particularly outstanding science' (besonders herausragende Wissenschaft). Ein Schelm, wer glaubt, dass POS für etwas anderes steht...

Besonderer Dank gilt Alfred Müller für seine Anregungen und hilfreichen Ratschläge zu stochastischen Ordnungsrelationen und Abhängigkeit im Allgemeinen, sowie Annika Betken und Giorgio Micali. Ich bin den beiden sehr dankbar für die Diskussionen und ihr wertvolles Feedback zum Beweis der Verschiebungsinvarianz der r-Approximationsbedingung. Des Weiteren möchte ich mich bei Nils Heerten, Katharina Hees, Robin Mey und Manuel Schäfer für ihre wertvollen Kommentare und Korrekturlesen bedanken.

Mein Dank gilt auch der Deutschen Forschungsgemeinschaft (DFG) für die Finanzierung meiner Forschung. Darüber hinaus möchte ich mich insbesondere bei Annika Betken und Herold Dehling bedanken, die mich überhaupt erst dazu ermutigt haben, zu promovieren. Ohne sie hätte ich mir das nicht zugetraut.

Mein größter Dank gilt letztlich meiner Familie und meinen Freunden für ihre unendliche Geduld und Fürsorge, besonders in den letzten Monaten - ich weiß, dass sie es nicht immer leicht mit mir hatten. Auch wenn ich nicht jeden einzeln nennen kann, so möchte ich doch einige Ausnahmen machen: Ich bin meinen 'Schwiegereltern' Beate und Jürgen dankbar für ihr Vertrauen in mich und die Ermutigung, nach mehr zu streben. Es gibt nicht genug Worte, um meinen Eltern Ida und Johann für ihre bedingungslose Liebe und Unterstützung zu danken. Auch wenn ich nicht immer um Hilfe gebeten habe, so wussten sie doch immer, wann sie gebraucht wurde - eine Art siebter Sinn, den nur Eltern haben.

Und schließlich möchte ich meinem Partner Tom danken: Für jedes Manuskript, das er Korrektur gelesen hat, für jeden meiner geprobten Vorträge, den er sich anhören musste, und natürlich für das Korrekturlesen meiner Dissertation - und das alles, obwohl er Algebraiker ist (also auf der dunklen Seite der Mathematik steht) und daher nichts mit der Stochastik zu tun hat. Ich weiß, dass du das hier liest, also danke, dass du dir mein Geplapper immer angehört und mich all die Jahre ertragen hast. Danke, dass du mich immer aufmunterst, wenn ich mich schlecht fühle. Danke, dass du so bist, wie du bist. Ich hoffe, ich habe hier nicht zu fiele Vehler versteckt.

Contents

1 Introduction

Ordinal patterns are defined as the description of the relation of the elements of d consecutive data points in terms of position and rank order. This means that an ordinal pattern is fully specified by relations $x_s < x_t$ or $x_s > x_t$ for all pairs (s, t) of distinct points in time.

With regard to a data set $(x_t)_{t\in\mathbb{N}}$ and a fixed length or order $d \geq 2$ (or embedding dimension as it is sometimes called), the ordinal structure is captured by a sliding window approach, i.e., ordinal patterns for windows $(x_t, x_{t+\Delta} \ldots, x_{t+(d-1)\Delta})$ of length d are determined, where $\Delta \in \mathbb{N} := \{1, 2, \ldots\}$ denotes the delay or lag parameter. Thus, the original process is transformed into a categorical ordinal pattern process.

Note that the delay is often denoted by τ in the literature. However, in order to keep the notation as consistent as possible, we have decided to use Δ here. (Furthermore, we implicitly assume $\Delta = 1$ from Chapter 2 onwards. Hence, we will not have to explicitly take it into account any further.) In practice, recommendations for the order d are made depending on the field of application, but in general a choice $d \in \{2, \ldots, 7\}$ is reasonable. For illustration, all possible ordinal patterns of length $d = 3$ are depicted in Fig. 1.1.

The actual values x_t are not taken into account, so working with ordinal patterns leads to a loss of information. (The actual amount of information lost is considered more closely in Section 1.1.2.) However, analyzing data via ordinal patterns has many advantages which justify this loss. First of all, ordinal patterns are invariant with regard to (not necessarily linear) monotone transformations, which particularly implies that ordinal patterns are scale-invariant. Therefore, the choice of measuring instrument used for data collection has no influence on the result. Furthermore, ordinal patterns are robust with regard to small noise which can be possibly caused by measurement errors. Overall, the concept of ordinal patterns is very intuitive and extremely fast algorithms are available (see for instance Keller et al. [44], Unakafova and Keller [85] and Berger et al. [14]).

Due to these qualities, ordinal patterns have been used extensively in many contexts where random or 'random-like' systems are considered, the most prominent ones being dynamical systems or symbolic dynamics, information theory and time series analysis. In fact, there is already an extensive amount of literature on ordinal patterns available which covers a wide range of applications (see Section 1.1).

This dissertation ties in at various points of ordinal pattern analysis, though its main focus lies in the context of time series analysis and dependence. There it is a key problem to determine whether dependence is present at all in the data or underlying time series (either within one or between two time series). However, in order to be able to properly recognize our contributions, we first provide an overview of the existing literature. Admittedly, as we can only consider the tip of the iceberg, we do not claim completeness.

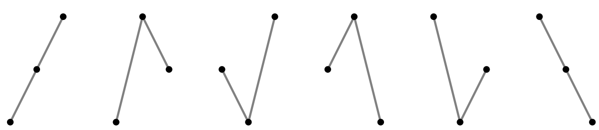


Figure 1.1: Ordinal patterns of length $d = 3$.

1.1 Historical Background

We begin by delving into the origins of ordinal patterns which lie in the theory of dynamical systems. From there we explore in different directions. Section 1.1.2 deals with data analysis and the estimation of parameters of the underlying processes. Ties between values are treated in Section 1.1.3. Section 1.1.4 deals with serial dependence and dependence between time series or data sets, followed by a survey on recent developments with regard to multivariate extensions of ordinal patterns in Section 1.1.5. We conclude this overview by considering some practical applications.

The historical overview (as well as parts of the above introductory section) are based on the joint work [81] with A. Schnurr.

1.1.1 Dynamical Systems and Entropy

Ordinal patterns have been introduced in the seminal paper by Bandt and Pompe [10] in the context of dynamical systems. At this point, the main purpose of ordinal patterns was to define so-called *permutation entropy*, which is the Shannon entropy for the distribution of ordinal patterns, as a measure of complexity in a dynamical system or time series. Most of the main types of complexity parameters, which are given by entropies, fractal dimensions, and Lyapunov exponents, break down as soon as noise is added to the series (see Bandt and Pompe [10]). In contrast, permutation entropy can be calculated for arbitrary real-valued time series. As permutation entropy is based on ordinal patterns, it inherits their advantages when compared to other complexity measures. Furthermore, for the family of logistic maps it behaves similarly to Lyapunov exponents [10].

Shortly thereafter, Bandt et al. [13] considered permutation entropy more theoretically and showed that, at least for piecewise monotone interval maps, limits exist and coincide with the Kolmogorov-Sinai entropy. A similar statement holds for topological entropy. In applications, this means that the complicated (but important) Kolmogorov-Sinai entropy can be approximated in several contexts by the much simpler permutation entropy. This has prompted further investigation into the relationship between these two entropy concepts in greater generality. Amigó et al. [5] have considered a definition of permutation entropy slightly differing from the one originally given. The authors have proved the result of [13] for their concept of permutation entropy under the additional requirement of ergodicity (Theorem 2), and extended it to maps on d-dimensional intervals (Theorem 4). Moreover, they showed that the permutation entropy of ergodic one-dimensional interval maps, as originally defined by [13], is an upper bound for the Kolmogorov-Sinai entropy. Considering the more general case of one-dimensional measure-preserving dynamical systems, Keller and Sinn [43] have shown that the Kolmogorov-Sinai entropy is not larger than permutation entropy under the sole assumption of measurability, which generalizes the result by Amigó et al. [5]. The relationship between these entropies has been further investigated by Keller et al. [45] and Amigó [1], where the latter has been able to extend the equality between the altered permutation entropy as proposed in [5] and the Kolmogorov-Sinai entropy also to (not necessarily ordered) measure-preserving dynamical systems without further assumptions.

Compare in this context also the recent PhD thesis [33]. There, Gutjahr has shown equality between the Kolmogorov-Sinai entropy and permutation entropy as defined by Bandt and Pompe [10] for one-dimensional dynamical functions which are monotone on a countable partition of the domain of definition (Theorem 3.17). Note that this generalizes the result given by Bandt et al. [13]. Furthermore, the author has investigated whether this equality also holds true if the Shannon entropy used in the definition of permutation entropy is replaced by the Rényi entropy. (Note that the Rényi entropy is a generalization of the Shannon entropy by definition, see Section 2.2.) In general this is not the case (see [33, Section 3.4.8]).

1.1.2 Data Analysis and Parameter Estimation

As already mentioned, the actual values of a time series are not taken into account in ordinal pattern analysis, but, as the name already suggests, only the ordinal structure. Hence, the question arises how much information on the structure of the underlying process has been neglected. Bandt and Shiha [11] have shown that for stationary ergodic processes, all finite-dimensional distributions can be recovered from the ordinal structure complemented by the one-dimensional distribution (Theorem 2). Even though this is a rather theoretical statement, it further justifies the use of ordinal patterns in practice.

However, before we further consider values based on pattern probabilities, let us take a step back to simply observing the sequence of patterns in a system. Already the visualization of ordinal patterns of fixed length in time can give useful insights into the nature of a time series. Amigó et al. [3] even recommend it to be the first step in analysis. To this end, ordinal patterns are symbolized by integers in a natural way. (Later, we will refer to them as *number representations* of ordinal patterns, see Chapter 3.2.) Usually they are numbered lexicographically according to their degree of deviation from a monotonically increasing pattern. On this basis, Keller and Sinn [42] have introduced so-called 'ordinal transformation' where the ordinal information contained in a time series up to some point in time (order) is extracted by some number in the interval $[0, 1]$. The higher the order, the more information is extracted, and for the (theoretical) case of an order equal to infinity, all information is extracted. This approach can be used, e.g., to detect change points in data sets or to classify different states (see, e.g., [42] for detection of epileptic seizures).

Now returning to values based on pattern probabilities, data analysis with regard to ordinal patterns is by no means limited to permutation entropy and variants of it. Rather, the consideration of frequencies of ordinal patterns already provides a lot of interesting insight on the underlying process. Sinn and Keller [83] were the first to study estimators based on ordinal pattern frequencies. They have thus derived an estimation of the Hurst parameter H of the data generating process. However, their considerations were restricted to the case $H < \frac{3}{4}$. Betken et al. [16] have complemented their work by considering the long-range dependent case $H > \frac{3}{4}$.

Bandt and Shiha [11] have proposed comparing the balance of monotone increasing and decreasing patterns (of length $d = 2$), which they refer to as *up-down balance*, in a time series as a function of the delay Δ. They also recommend consideration of relative frequencies of ordinal patterns of length $d = 3$, again depending on the delay (in accordance with the autocorrelation function). On this basis, they define the indicators of *persistence* and *rotational symmetry*. While the latter is self-explanatory, persistence describes how often a pattern $x_t < x_{t+\Delta}$ or $x_t > x_{t+\Delta}$ will persist in subsequent comparisons between $x_{t+\Delta}$ and $x_{t+2\Delta}$. It can be used to detect periodicities [11] and for small delays Δ, it describes the degree of smoothness present in the data [7]. *Turning rate*, which has been discussed by Bandt [7], counts the number of turning points and can be understood as complementary to persistence as it is a measure of roughness. We refer to these as well as so-called up-down scaling as *pattern contrasts*. They have been discussed further in terms of describing certain asymmetries and periodicities [7], change points [8] and independence of (most) pattern contrasts under certain assumptions [9]. Moreover, at this point we would like to emphasize applicational results with regard to EEG data from sleep research in particular: The work by Bandt [9] indicates a very tight connection between sleep stages and turning rate. Furthermore, it shows that slow oscillations, which, as the author states, are hardly accessible by conventional methods, can be detected by turning rate.

1.1.3 Ordinal Patterns and Ties

Since their introduction, a critical assumption for the definition of ordinal patterns stemming from a time series has been the continuity of the underlying distribution such that no or just a few equal values are present. We refer to these equalities by *ties*. However, if ties are present in the data, three different approaches are common [72]: In the first approach, windows containing ties are omitted from the analysis which might lead to a great loss of information in case of many ties being present. For the second approach a small noise is added to the data having the drawback of disregarding constant patterns or possibly underestimating the co-movement between multiple time series. For the last approach, the definition of ordinal patterns is altered in favor of increasing patterns, i.e., ties $x_s = x_t$ with $s < t$ are treated as if '$x_s < x_t$'. However, this maps the vectors $(1, 1, 1)$ and $(1, 10, 100)$ onto the same ordinal pattern, that is, they are considered to exhibit the same up and down movement, which is clearly not the case here.

Therefore, Schnurr and Fischer [72] introduced so-called *generalized ordinal patterns* where they explicitly allow for ties by referring to a larger set of possible patterns. Generalized ordinal patterns clearly overcome the drawbacks of the previously mentioned classical approaches, but depending on the order of ordinal patterns, a lot more patterns need to be considered. In fact, the patterns are no longer classical permutations, but Cayley permutations, whose number is equal to the ordered Bell number of order d. This can result in a greater computational cost and more underlying parameters in statistical estimation. Therefore, the recommendation is to use them with regard to time series where many ties are to be expected, as it is, e.g., the case for categorical time series with few categories. For more information on capturing the ordinal structure in presence of ties, see Section 3.4.

So far, generalized ordinal patterns have been (mostly) considered in the context of dependence.

1.1.4 Dependence

As serial independence is a critical assumption for most statistical tests, there is a great interest in testing for dependence. As a result, an extensive amount of literature is available on this matter (see for instance [55] and the references mentioned therein). Ordinal patterns have also been widely applied to derive tests for (possibly non-linear) serial dependence. Compared to other tests, tests based on ordinal patterns have many advantages. Essentially, the main characteristics of ordinal patterns are preserved, that is, they are robust with regard to small noise and invariant with regard to monotone transformations. Furthermore, the tests are non-parametric. Only the order of the ordinal patterns considered has to be fixed beforehand, though there are recommendations available based on the selected test and the length of the data set at hand. Computational simplicity and, resulting from this, short computational times make tests based on ordinal patterns very appealing.

One of the first tests on serial dependence has been provided by Matilla-García [54]. There, the idea was to compare the relative frequencies of ordinal patterns (of fixed length) with their distribution under the null of serial independence following Pearson's approach in [65]. Later, Matilla-García and Marín [55] have introduced a consistent test based on the permutation entropy and maximum likelihood estimation. Caballero-Pintado et al. [21] have defined the symbolic correlation integral which is highly inspired by the classical correlation integral, but avoids the necessity of choosing an adequate distance parameter beforehand by being based on ordinal patterns. On this basis, the authors have proposed a new "nonparametric test for independence that overcomes some of the pitfalls of the BDS statistic, which is a test based on the standard correlation integral and therefore is distance-dependent" [21, p. 548].

Weiß [87] has derived tests in real-valued time series not only based on the permutation entropy as well as some modifications of it including the symbolic correlation integral, but also on most of the pattern contrasts by investigating their asymptotic distributions under the null of independence. Weiß and Schnurr [89] have proposed a test for independence in time series where ties are explicitly allowed, using generalized ordinal patterns as introduced in Schnurr and Fischer [72]. They have shown that the pattern distribution is not uniform under the null of serial independence, but depends strongly on the underlying distribution. Therefore, they have chosen a data-driven approach based on the classical Efron's bootstrap procedure in order to dispose of parametric assumptions. On the other hand, Weiß and Testik [90] have extended the theory in terms of monitoring for the existence of serial dependence in real-valued and continuously distributed processes. There, 'monitoring' means that deviations with regard to independence can already be detected during data collection. In this regard, the authors have proposed control charts based on ordinal patterns, which overcome the typical drawbacks of control charts discussed so far. This means, that their approach is fully non-parametric and distribution-free, no parameter estimation is required and it can be used almost immediately at the start of process monitoring, i.e., there is (almost) no delay.

With regard to dependence *between* time series, Matilla-García et al. [57] have proposed a test for independence between time series which is based on permutation entropy. As a result, the authors obtain a non-parametric test for not-necessarily linear processes, which does not require restrictive assumptions and is consistent for those processes where dependence structure is within the (a priori fixed) order of ordinal patterns. The authors' approach is extended by Matilla-García et al. [56], who have derived a non-parametric test for linear and nonlinear causality.

With regard to *measuring* dependence between time series, so-called ordinal pattern dependence has been proposed by Schnurr [70] and Schnurr and Dehling [71]. The idea is to count the number of coincident patterns in both time series and compare this with the expected number in case of (hypothetical) independence. Then, numerous coincident patterns would indicate similar up-and-down behavior and hence, co-monotonic behavior of the two time series.

Ordinal pattern dependence has several advantages when compared to classical measures of dependence: Apart from the advantages that ordinal patterns entail, ordinal pattern dependence does not require the existence of second moments. The concept is very intuitive and again fast algorithms are available. Betken et al. [17] have investigated how ordinal pattern dependence fits into the axiomatic framework of multivariate measures of dependence proposed by Grothe et al. [32].

Schnurr and Dehling [71] have derived limit theorems for ordinal pattern dependence and a test on structural breaks under stationarity and some mild assumptions on short-range dependence. Later, Betken et al. [16] and Nüßgen and Schnurr [62] have supplemented limit theorems under long-range dependence with the latter also considering the mixed cases of short- and long-range dependence.

Schnurr and Fischer [73] have proposed a method to detect change points in the dependence structure between possibly non-stationary time series which is based on ordinal patterns. They have demonstrated that their method overcomes problems of classical techniques in case of non-stationarity.

1.1.5 Multivariate Extensions

Recently, there has been a growing interest in extending ordinal patterns to multivariate time series. An overview of early approaches has been provided by Mohr et al. [58], but, as the authors have pointed out, none of those take potential correlations between the movement of the variables into account. Instead, Mohr et al. [58] have proposed two approaches of their own. Their first idea is what we refer to as *multivariate ordinal patterns* in the remaining and is based on the idea of storing the classical univariate ordinal patterns of both components in one vector/matrix (depending on the chosen representation, see Section 6.1). The second idea is based on principal component analysis. There, the multivariate time series is transformed into a univariate time series from which univariate ordinal patterns then can be obtained as usual. Even though both of the approaches lead to the avoidance of a loss of information with regard to the co-movement of the components, the advantage of the second method over the first is that a smaller number of patterns has to be taken into account there.

Later, Bandt and Wittfeld [12] have proposed a method where even more information is preserved and which we refer to as *spatial ordinal patterns*. There the patterns of ups and downs are no longer determined component by component, but across all dimensions. In particular, the authors have considered 2×2-patterns and categorized them into three types. Then, they have proposed statistics based on these types in order to describe and distinguish textures in images and therefore, highlighted the application of spatial ordinal patterns with regard to image processing. Weiß and Kim [88] have considered spatial ordinal patterns under the assumption of (spatial) independence. They have derived the asymptotics under the null and proposed tests for spatial dependence.

Restricting themselves to two dimensions, Fischer et al. [29] have introduced the concept of *motion patterns*. There, the authors considered the definition of multivariate ordinal patterns, i.e., vectors of univariate patterns, and reduced their number by categorization in order to obtain a computationally efficient procedure which retains a sufficient amount of information. This categorization is done by depicting both univariate ordinal patterns of order $d = 3$ in a grid and considering the nature of the resulting movement. Then, the ordinal behavior of the time series can be analyzed in terms of classes of movements. Both cases with and without ties have been considered by the authors, and limit theorems for motion pattern distributions for stationary strongly mixing processes have been derived.

Betken and Schnurr [15] have established a definition for ordinal patterns in more than one dimension based on the concept of Tukey's halfspace depth. "The basic idea of statistical depth is to measure how deep a specific element in a multidimensional space lies in a given, multivariate reference distribution, and therefore naturally leads to a center-outward ordering of sample points in multivariate data" [15]. This way univariate ordinal patterns can be determined which represent the multivariate ordinal structure of the data (dimension reduction technique). The authors have provided limit theorems for the cases where the reference function is known and where it is unknown under the assumption of weak dependence of the underlying time series.

1.1.6 Practical Applications

Ordinal pattern analysis is a "popular method in biology and medicine, especially when it comes to distinguishing abnormal from normal health conditions in real time" [3, p. 2]. In fact, "it is not surprising that permutation entropy and similar measures are applied to physical data because such data are characterized by special underlying patterns often related to certain states of a system (e.g. spike-and-wave patterns related to epilepsy, special sleep patterns such as sleep spindles, and burst suppression patterns related to inactivated brain states, or Wolff–Parkinson–White patterns in abnormal electrocardiograms). It seems that ordinal patterns are appropriate for capturing structures containing such patterns and also abrupt changes in their distributions" [3, p. 13]. For a throughout survey on applications of ordinal pattern analysis in biomedicine, we refer to [3] and the references therein. Apart from the applications mentioned so far (including the previous sections), from the very beginning, ordinal patterns have been used to analyze speech signals [10]. Further applications include finance [8, 55, 56]. In the context of negative dependence between financial index data, ordinal pattern dependence was used in [70]. Multivariate patterns were applied to climate data (center of rain events) in [29] and patterns with ties to hydrological data in [72] and [73]. Furthermore, extremal events were treated in [63].

1.2 Goals of this Thesis

As already mentioned, this work picks up on various points in the existing literature. We mainly contribute to the theory of dependence between two time series. There, one of the intended contributions is to propose non-parametric tests for dependence *between* time series based on multivariate extensions of ordinal patterns and conduct a comparative performance analysis taking into account some classic competitors. Furthermore, we reconsider the results of Betken et al. [17] with regard to ordinal pattern dependence and the axiomatic framework for multivariate measures of dependence proposed by Grothe et al. [32]. Silbernagel [79, p. 89ff] has noted that there is an error in the proof of the fifth axiom. Therefore, a question which naturally arises is whether this axiom (concordance ordering) can be verified for ordinal pattern dependence at all. In addition, the question arises as to how ordinal pattern dependence relates to other stochastic order relations.

Another main goal of this work is to complement the work of Caballero-Pintado et al. [21] by establishing limit theorems for the symbolic correlation integral under the assumption of short-range dependence. This would pave the way for several possible applications, which are not feasible with the results limited to the independent and identically distributed (i.i.d.) case. In particular, this would form the basis for the development of hypothesis tests on whether two (short-range dependent) time series have the same underlying structure/complexity/model.

Another aim is given by the small but non-negligible contribution of a survey on ordinal pattern representations existing in the literature, with a special focus on their applicability from different perspectives.

1.3 Outline of this Thesis

In Chapter 2, we introduce the reader to most of the mathematical tools and concepts necessary for the remaining of this work: We provide basic definitions in the context of time series analysis and dependence properties, as well as entropy, which has its origin in information theory, but has also become an important tool especially in the context of ordinal patterns and dynamical systems. Furthermore, we introduce the reader to U-statistics and approximating functionals in the context of short-range dependent stochastic processes. We derive or rather generalize limit theorems for U-statistics of approximating functionals given by Borovkova et al. [18], which will be useful later. We conclude this chapter with a short introduction to regular conditional distributions.

The main part of this work can be divided into two parts. In the first part, we contribute to ordinal pattern analysis by considering classical (univariate) ordinal patterns. We begin with an extensive introduction to ordinal patterns in Chapter 3. There we describe and analyze different approaches to represent them by deriving their advantages and disadvantages in different contexts. Namely, we consider digital implementation of ordinal patterns, inverse patterns and ties between values. Chapter 4 deals with the symbolic correlation integral and the Rényi entropy with regard to ordinal pattern distributions. We introduce a natural estimator on the basis of U-statistics and derive limit theorems under the assumption of short-range dependence. Here we make use of the limit theorems considered in Chapter 2. We complete this chapter with simulations to support and discuss our results in more detail. In Chapter 5 we reconsider the results of Betken et al. [17] with regard to the axiomatic framework of multivariate measures of dependence proposed by Grothe et al. [32]. We establish a counterexample which shows that ordinal pattern dependence does not satisfy the fifth axiom, and derive a proof under different but arguably similar assumptions. Moreover, we consider ordinal pattern dependence with regard to other stochastic order relations. We conclude this chapter (as well as the first part of this thesis) with a critical look at axiomatic approaches with regard to measures of dependence.

The second part of this thesis focuses on multivariate extensions of ordinal patterns. In the same way as in the first part, in Chapter 6 we discuss representations of two of the extensions, namely, multivariate ordinal patterns and spatial ordinal patterns. With regard to these, we continue with the introduction of a general framework for dependence tests between time series under the assumption of serial independence (Chapter 7). This also includes ordinal pattern dependence as it can be embedded into the context of multivariate ordinal patterns. To this end, we prove general limit theorems of multivariate pattern distributions. These encompass some existing results. We compare the performance of the proposed ordinal pattern-based statistics with three competitors, namely classical Pearson's and Spearman's correlation as well as the rank-based Chatterjee's correlation coefficient, in a simulation study and real-world data examples.

2 Mathematical Preliminaries

In this chapter, we give an overview on the mathematical tools and concepts necessary for our results. This does not yet encompass ordinal patterns, which are treated in the subsequent chapter. First of all, we provide a brief introduction into time series and short-range dependence. Then we introduce the concepts of entropy and U-statistics, before we discuss a broad class of short-range dependent processes in more detail, namely r-approximating functionals. There we derive limit theorems for U-statistics of r-approximating functionals which constitute generalizations of the limit theorems established by Borovkova et al. [18]. Those will be of great importance to us in Chapter 4. We conclude this chapter with a short introduction to regular conditional distributions, which will be relevant in Chapter 5.

2.1 Time Series Analysis and Dependence

In order to define time series, let us recall the definition of stochastic processes first.

Definition 2.1.1 (Brockwell and Davis [20, Definition 1.2.1])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space, $(S, \mathcal{S})$ a measurable space and T an index set. If $X_t : \Omega \to S$ is a (possibly multivariate) random variable for each $t \in T$, then we call $(X_t)_{t\in T}$ a *stochastic process* and $(S, \mathcal{S})$ its *state space.*

Based on the above definition, a stochastic process is an indexed family of random variables. As the name might already suggest, in time series analysis the index set T consists of time points - we then call $(X_t)_{t\in T}$ a *time series.* Such index sets are, for example, given by the natural numbers $\mathbb{N} := \{1, 2, \ldots, \}$ or positive real numbers $\mathbb{R}^+ :=]0, \infty[$. If the past is also to be taken into account, a two-sided process can be considered by setting the set of integers $\mathbb{Z}$ or the entirety of the real numbers $\mathbb{R}$ as the index set. In this thesis, we focus on discrete time.

A further distinction must be made in the choice of the state space $(S, \mathcal{S})$: We distinguish between continuous- and discrete-valued processes, depending on whether S is continuous or discrete. This distinction is crucial when dealing with ordinal patterns: The nature of the state space largely determines whether ties have to be taken into account, which in turn have an enormous influence on the resulting ordinal pattern distribution. Recall that the probability of coinciding values equals zero in the continuous case. In this thesis, most of the time, that is, unless indicated otherwise, we deal with the admittedly easier continuous case.

Due to its great importance, we remind the reader of the definition of stationarity:

Definition 2.1.2 (Brockwell and Davis [20, Definitions 1.3.2/1.3.3.])**.**

1. A time series $(X_t)_{t\in\mathbb{Z}}$ is called *weakly stationary* if it satisfies the following three conditions:

   a) $\mathbb{E}\,|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$,

   b) $\mathbb{E}X_s = \mathbb{E}X_t$ for all $s, t \in \mathbb{Z}$,

   c) $\operatorname{Cov}(X_{s+h}, X_{t+h}) = \operatorname{Cov}(X_s, X_t)$ for all $s, t, h \in \mathbb{Z}$.

2. A time series $(X_t)_{t\in\mathbb{Z}}$ is called *(strictly) stationary* if

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{D}{=} (X_{t_1+h}, \ldots, X_{t_k+h})$$

for all $k \geq 1$ and $t_1, \ldots, t_k, h \in \mathbb{Z}$, where $\stackrel{D}{=}$ denotes equality in distribution.

Note that the definition of strict stationarity is equivalent to the statement that

$$(X_1, \ldots, X_k) \stackrel{D}{=} (X_{1+h}, \ldots, X_{k+h}) \qquad \text{for all } k \geq 1 \text{ and } h \in \mathbb{Z}$$

(see, e.g., [20, Remark 4]). Furthermore, note that a time series $(X_t)_{t\in\mathbb{Z}}$ with finite variance is weakly stationary (cf. [20, Chapter 1.3]).

Dependence properties of time series play a crucial role for the application of limit theorems. In general, a distinction is drawn between short- and long-range dependent processes. These are characterized by properties of their autocovariance functions.

Definition 2.1.3 (Brockwell and Davis [20, Definition 1.3.1/Remark 2])**.** Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary time series. We define the *autocovariance function* γ_X associated with $(X_t)_{t\in\mathbb{Z}}$ by

$$\gamma_X(h) := \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_0, X_h) \qquad \text{for } t, h \in \mathbb{Z}.$$

If the association with the time series $(X_t)_{t\in\mathbb{Z}}$ is clear from the context, we omit the index and simply write $\gamma(\cdot) := \gamma_X(\cdot)$.

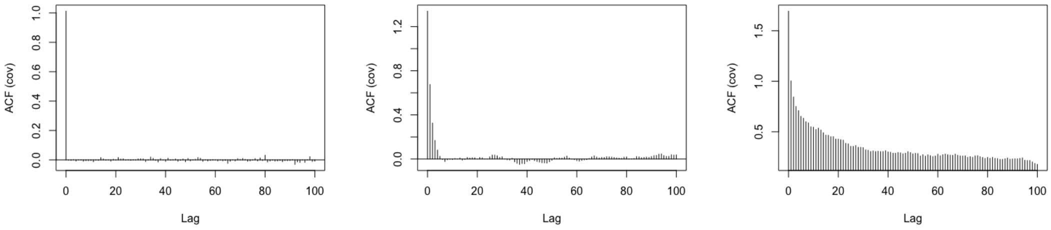
Definition 2.1.4 (Pipiras and Taqqu [66, Definition 2.3.1/Section 2.1])**.** A weakly stationary time series $(X_t)_{t\in\mathbb{Z}}$ is called

- *short-range dependent* if its autocovariance functions are absolutely summable, that is, $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$,
- and *long-range dependent* else, i.e., if $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$.

Note that there are various definitions of long-range dependence which can be equivalent depending on the choice of the slowly varying functions [66]. For more details, we refer the reader to [66, Chapters 2.1 – 2.2].

Classical examples for short-range and long-range dependent processes are *autoregressive moving average (ARMA)*-processes and *autoregressive fractionally integrated moving average (ARFIMA or FARIMA)*-processes, respectively [66, p. 15]. (A proper definition for ARMA-processes is given in Section 2.4.1.) Consider in this regard the autocovariance functions of the AR(1)- and FAR(1)-process depicted in Fig. 2.1. The autocovariance function of the AR(1)-process converges very quickly to 0. This is not the case for the FAR(1)-process. There, the dependence seems to reach infinitely far – or at least it is decaying very slowly.

In this thesis, we restrict ourselves to short-range dependent processes. There are many kinds of short-range dependent processes which are varying in strength of the assumptions. We do not intend to discuss all of them in detail here. Instead we provide a brief overview in Fig. 2.2. There, the arrows imply, e.g., that absolute regular processes are also r-approximating functionals, that is, absolute regularity is stronger than the r-approximating condition. Furthermore, note that "mixing" in the lower left part of this illustration is meant in the ergodic-theoretic sense, and let us emphasize that the r-approximating condition and L_r-near-epoch dependence (L_r-NED) are equivalent only for stationary processes. Note that



(i) **(ii)** **(iii)**

Figure 2.1: Autocovariance functions of (i) an i.i.d. series, (ii) an AR(1)-process with autoregressive parameter $\phi = 0.5$ and (iii) an FAR(1)-process with $\phi = 0.5$ and $d = 0.4$.

this overview constitutes only a selection and is by no means a complete list of all concepts available in the literature. Furthermore, let us emphasize that we refrain from giving a definition for each of the concepts mentioned, since this is beyond the scope of this thesis. Instead, we refer the interested reader for more details, e.g., to Bradley [19] and Davidson [24, Chapters 13 – 14, 17], as well as Dehling et al. [26] with regard to near-epoch dependence in probability ($\mathbb{P}$-NED).

In this thesis, we mainly focus on r-approximating functionals (Section 2.4). It is in fact no easy task to find/construct a short-range dependent stochastic process which does not fulfill the r-approximating condition (see Section 2.4.1), which already speaks for the sheer size of this class of short-range dependent processes.

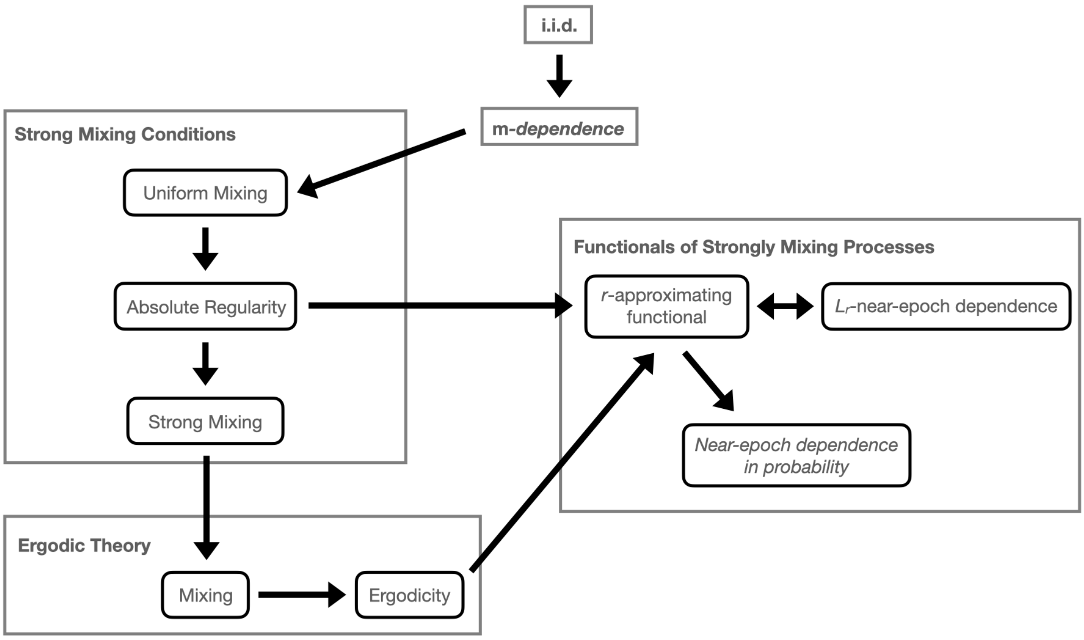


Figure 2.2: Illustration of the relations between different kinds of short-range dependent processes under the assumption of stationarity.

2.2 Entropy

The concept of entropy was first introduced in 1865 by Clausius [23] in the context of thermodynamics. Later, Shannon [78] used entropy as a basis to build the theory of information and communication on. For a historical overview and recent applications, we refer, e.g., to the review by Amigó et al. [4]. However, entropy has also become highly relevant in the context of ordinal patterns (recall that ordinal patterns have been originally introduced as a tool for so-called permutation entropy in [10]).

In general, entropies measure inhomogeneity, impurity, complexity and uncertainty or unpredictability [58]. In this thesis, we are going to employ the concept of entropy to perform classification tasks on the degree of complexity present within a single time series (Chapter 4) as well as tests for dependence between two time series (Chapter 7), since dependence is strongly connected to how unpredictable or uncertain an event is. To this end, we first give a mathematical definition of uncertainty (taken from the field of information theory) and then, we indicate how entropy is a measure of uncertainty. Thereafter, we state the definition of (Shannon) entropy and discuss some of its most characteristic properties. Lastly, we give the definition of the Rényi entropy as one example of so-called generalized entropies and discuss its relation to the classical (Shannon) entropy.

Definition 2.2.1 (Martin and England [53, Defintion 2.1])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. The *information*, or *uncertainty*, $I : [0,1] \to \mathbb{R}$ is a real-valued function of probabilities of measurable events which satisfies the following three properties:

(I1) An event $A \in \mathcal{F}$ which occurs almost surely, that is, with probability $\mathbb{P}(A) = 1$, has zero uncertainty: $I(\mathbb{P}(A)) = 0$.

(I2) If an event $A_1 \in \mathcal{F}$ is less probable than another event $A_2 \in \mathcal{F}$, then the first event A_1 has more uncertainty than the second A_2. That is, $p^{(1)} := \mathbb{P}(A_1) \leq \mathbb{P}(A_2) =: p^{(2)}$ implies $I(p^{(1)}) \geq I(p^{(2)})$.

(I3) The uncertainty of the simultaneous occurrence of two independent events is the sum of their individual uncertainties. In more detail, for events $A_1, A_2 \in \mathcal{F}$ with $p_{12} := \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) = p^{(1)} \cdot p^{(2)}$, it follows $I(p_{12}) = I(p^{(1)} \cdot p^{(2)}) = I(p^{(1)}) + I(p^{(2)})$.

As explained by Martin and England [53], it is quite easy to find functions which satisfy the first two properties. The third condition is more restrictive. In fact, it is a classical result that the only measurable function $f : [0,1] \to \mathbb{R}$ which satisfies the functional equation $f(xy) = f(x) + f(y)$ for $x, y \in [0,1]$ is given by a constant multiple of the natural logarithm, i.e., $f(x) = C \log(x)$ for $C \in \mathbb{R}$ constant (see [53, p. 52]). Note that f vanishes at 1 for any constant C, but it is monotonously decreasing if and only if $C < 0$. Therefore, if we require the uncertainty to be measurable with regard to the probabilities of events, then I is uniquely defined by

$$I(\mathbb{P}(A)) = \begin{cases} -b \log(\mathbb{P}(A)) & \text{if } \mathbb{P}(A) > 0, \\ \infty & \text{if } \mathbb{P}(A) = 0, \end{cases} \tag{2.1}$$

where $b > 0$ is some positive real number and $A \in \mathcal{F}$ denotes some measurable event.

Unless stated otherwise, we use natural logarithms (which we denote by log), though logarithms to base 2 are often a more natural choice with regard to information and entropy (see, e.g, [4, 53]).

Let the m-dimensional vector $\mathbf{p} = (p^{(1)}, \dots, p^{(m)})^\top \in [0,1]^m$ refer to the probability mass function (pmf) of some discrete random variable X with finite range of possible states. Note that, in particular, the components of $\mathbf{p}$ have to sum up to one, i.e., $p^{(1)} + \cdots + p^{(m)} = 1$. Then, the *Shannon entropy* H of $\mathbf{p}$ is defined as the linear average of the information function I defined in (2.1) or, equivalently, as the expected value of the random variable $I(p(X)) = -b\log(p(X))$, where p denotes the function which assigns the respective probability $p^{(j)}$ to each realization x_j of X:

$$H(\mathbf{p}) = \mathbb{E}(I(p(X))) = \sum_{j=1}^{m} p^{(j)} I(p^{(j)}) = -b \sum_{j=1}^{m} p^{(j)} \log p^{(j)} \tag{2.2}$$

with the convention $x \log x = 0$ for $x = 0$ (see [4, 53]). Hence, entropy can be interpreted as the expected uncertainty or randomness. It holds

$$0 \leq H(\mathbf{p}) \leq b \log m$$

(see, e.g., [53, Theorem 2.3]), and thus, setting $b = 1/\log(m)$ yields a natural standardization.

Definition 2.2.2. Let $\mathbf{p} \in [0,1]^m$ be a pmf. Then, its *(standardized) Shannon entropy* is defined as

$$H(\mathbf{p}) := -\frac{1}{\log(m)} \sum_{j=1}^{m} p^{(j)} \cdot \log(p^{(j)}),$$

with the convention $0 \cdot \log(0) := 0$.

Remark 2.2.3. The Shannon entropy is often referred to as just *entropy* (see, e.g., [4, 53]) due to its uniqueness (see below). If $\mathbf{p}$ expresses the ordinal pattern probabilities of a time series, then $\mathrm{PE}(\mathbf{p}) := H(\mathbf{p})$ is referred to as the *(standardized) permutation entropy*, which has been introduced by Bandt and Pompe [10].

For a uniform distribution, that is, all possible states appear with the same probability $1/m$, the standardized Shannon entropy attains its upper bound given by 1 as the time series is then the most uncertain or unpredictable. In contrast, a time series is the least complex if it follows a one-point distribution, i.e., $p^{(j)} = 1$ for one $1 \leq j \leq m$ and $p^{(k)} = 0$ otherwise, which yields $\mathrm{PE}(\mathbf{p}) = 0$ in consequence (see, e.g., [47, p. 4]).

Remark 2.2.4. Note that discrete pmf vectors $\mathbf{p}$ may also arise if some originally continuously distributed time series is transformed into a sequence of ordinal patterns. Now, let us make some remarks about what the above means with regard to ordinal patterns. First of all, even though the pattern distribution is uniform for (classical) univariate ordinal patterns stemming from an i.i.d. time series, this is not the case for generalized ordinal patterns (see Weiß and Schnurr [89]). Next, a one-point ditribution with regard to ordinal pattern distributions corresponds to a monotonically increasing or decreasing time series. This is, in fact, the only case where the ordinal patterns of a time series follow a one-point distribution, because the ordinal patterns are obtained from a time series in an overlapping nature. In more detail, the windows $(x_t, \dots, x_{t+d-1})$ consisting of the original data $(x_t)_{t\in\mathbb{N}}$ potentially share some components for $t \in \mathbb{N}$. Consequently, the respective ordinal patterns must indicate the same ordinal structure for the overlapping segments. For illustration, consider Figure 2.3. There, an ordinal pattern of length $d = 3$ which is neither monotonically increasing nor decreasing is depicted in black. All options for subsequent patterns (of length $d = 3$) are shown in green,

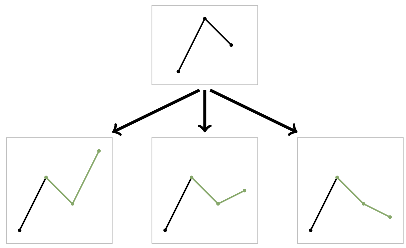


Figure 2.3: An ordinal pattern of length $d = 3$ in black, and its possible subsequent ordinal patterns in green.

but none of them agrees with the original ordinal pattern, making a one-point distribution impossible (see Silbernagel et al. [82]).

Shannon entropy is unique in the sense that it is the only function defined on pmfs which satisfies the so-called *Khinchin-Shannon axioms* [4]:

(KS1) *Continuity:* $H(\mathbf{p})$ depends continuously on all variables for each m.

(KS2) *Maximality:* For all m,

$$H(\mathbf{p}) \leq H\Big(\frac{1}{m}, \dots, \frac{1}{m}\Big).$$

(KS3) *Expansibility:* For all m and $1 \leq j \leq m$,

$$H(0, p^{(1)}, \dots, p^{(m)}) = H(p^{(1)}, \dots, p^{(j)}, 0, p^{(j+1)}, \dots, p^{(m)}) = H(\mathbf{p}).$$

(KS4) *Separability* (or *strong additivity*): For all $m, n \geq 2$ it holds

$$\begin{aligned} &H(p_{11}, \dots, p_{1n}, p_{21}, \dots, p_{2n}, \dots, p_{m1}, \dots, p_{mn}) \\ &\qquad = H(p^{(1)}, \dots, p^{(m)}) + \sum_{i=1}^{m} p^{(i)} H\Big(\frac{p_{i1}}{p^{(i)}}, \dots, \frac{p_{in}}{p^{(i)}}\Big), \end{aligned}$$

where $p^{(i)} = \sum_{j=1}^{n} p_{ij}$.

If

$$\mathbf{p}_{X,Y} = (p_{11}, \dots, p_{1n}, p_{21}, \dots, p_{2n}, \dots, p_{m1}, \dots, p_{mn})^\top$$

corresponds to the joint pmf of some random variables X and Y with marginal distributions

$$\mathbf{p}_X = (p_X^{(1)}, \dots, p_X^{(m)})^\top = (p^{(1)}, \dots, p^{(m)})^\top \quad \text{and} \quad \mathbf{p}_Y = (p_Y^{(1)}, \dots, p_Y^{(n)})^\top,$$

respectively, where $p_Y^{(j)} = \sum_{i=1}^{m} p_{ij}$, then axiom (KS4) is equivalent to $H(\mathbf{p}_{XY}) = H(\mathbf{p}_X) + H(\mathbf{p}_{Y|X})$ where $H(\mathbf{p}_{Y|X})$ is the entropy of Y conditional on X. If X and Y are independent, i.e., $p_{ij} = p_X^{(i)} p_Y^{(j)}$, then $H(\mathbf{p}_{Y|X}) = H(\mathbf{p}_Y)$ and

$$H(\mathbf{p}_{X,Y}) = H(\mathbf{p}_X) + H(\mathbf{p}_Y) \tag{2.3}$$

(see [4], [47, p. 4f]). Property (2.3) is referred to by *additivity.* Note the connection to (I3).

Non-negative functions defined for pmfs which satisfy (KS1) – (KS3) are referred to as *generalized entropies* [4]. Classical examples are the Tsallis entropy and Rényi entropy. Here we focus on the latter due to its relevance in Chapter 4.

Definition 2.2.5 (Rényi [68])**.** The *Rényi-q entropy* of some pmf vector $\mathbf{p} = (p^{(1)}, \ldots, p^{(m)})^\top$ is defined as

$$R_q(\mathbf{p}) := \frac{1}{1-q} \log\Big(\sum_{j=1}^{m} (p^{(j)})^q\Big)$$

where $q \geq 0, q \neq 1$.

Remark 2.2.6. If $\mathbf{p}$ expresses the ordinal pattern probabilities of a time series, then $R_q(\mathbf{p})$ is called *Rényi-q permutation entropy* [2].

Hence, the Rényi entropy is actually a one-parameter family of entropies. It was introduced with the motivation in mind to find the most general class of information measures that preserved the additivity of statistically independent systems and were compatible with Kolmogorov's probability axioms (for more details, see Principe [67, Chapter 2.2]). At this point, let us remark that the Shannon entropy is the simplest of those information measures [67].

Furthermore, one can show that the Rényi entropy is related to the Shannon entropy by its limit:

$$\begin{aligned}
\lim_{q\to 1} \frac{1}{1-q} \log\Big(\sum_{j=1}^{m} (p^{(j)})^q\Big) &= -\lim_{q\to 1} \frac{\partial}{\partial q} \log\Big(\sum_{j=1}^{m} (p^{(j)})^q\Big) \\
&= -\lim_{q\to 1} \frac{1}{\sum_{j=1}^{m} (p^{(j)})^q} \sum_{j=1}^{m} (p^{(j)})^q \log p^{(j)} \\
&= -\sum_{j=1}^{m} p^{(j)} \log p^{(j)},
\end{aligned}$$

where we have used L'Hopital's rule in the first equation and $\sum_{j=1}^{m} p^{(j)} = 1$ in the last [4]. For further information on the Rényi entropy we refer the reader to Principe [67, Chapter 2].

The empirical versions are obtained by using relative frequencies $\hat{p}^{(j)}$, $1 \leq j \leq m$, instead of probabilities.

2.3 U-statistics

Even though certain optimal properties of U-statistics have been proved before by Halmos [34], U-statistics have been formally introduced by Hoeffding [35]. Many classical estimators have been recognized as U-statistics, including the sample mean, sample variance and Kendall's τ. However, many recently proposed estimators are also based on U-statistics as, e.g., the estimator for the symbolic correlation integral and Rényi-2 permutation entropy (see Chapter 4).

Definition 2.3.1 (Hoeffding [35])**.** Let $(X_t)_{t\in\mathbb{N}}$ be a stationary sequence of d-dimensional random vectors each with cumulative distribution function (cdf) F. For $m \geq 1$, let $h : \mathbb{R}^{m\times d} \to \mathbb{R}$ be a measurable function which is invariant under permutation of its arguments. The *U-statistic* of *order* (or *degree*) m with kernel h is defined by

$$U_n := U_n(h) := \binom{n}{m}^{-1} \sum_{1\leq t_1<\cdots<t_m\leq n} h(X_{t_1}, \ldots, X_{t_m}).$$

U-statistics were named after their property of being unbiased estimators for so-called regular functionals

$$\theta(F) = \int \cdots \int h(x_1, \ldots, x_m) dF(x_1) \ldots dF(x_m)$$

(therefore the letter 'U'), even though their "symmetric nature [. . .] plays a more important role in the characterization of their optimality properties" [76, p. 301].

A crucial tool for analyzing the asymptotic behavior of U-statistics is the Hoeffding-decomposition [35], where a U-statistic $U_n(h)$ with kernel h is decomposed into a linear and a degenerate part. For $m = 2$, this is

$$U_n(h) = \theta + \frac{2}{n} \sum_{t=1}^{n} h^*(X_t) + \frac{2}{n(n-1)} \sum_{1 \le t_1 < t_2 \le n} J(X_{t_1}, X_{t_2}), \tag{2.4}$$

where $\theta = \mathbb{E}h(X_1, X_2)$, $h_1(x_1) = \mathbb{E}h(x_1, X_2)$, $h^*(x_1) = h_1(x_1) - \theta$ and $J(x_1, x_2) = h(x_1, x_2) - h_1(x_1) - h_1(x_2) + \theta$. Then, it directly follows

$$U_n(h) = \theta + 2U_n(h^*) + U_n(J),$$

where $U_n(h^*)$ and $U_n(J)$ denote the U-statistics with regard to the respective kernels. One then shows that the degenerate part $U_n(J)$ is asymptotically negligible such that it remains to consider the asymptotic distribution of the linear part $U_n(h^*)$.

Hoeffding has established asymptotic normality and a law of large numbers for i.i.d. sequences in [35] and [36], respectively. Since then, asymptotical results for U-statistics (including invariance principles for empirical processes of U-statistics structure) have been proved for cases where the underlying stochastic process $(X_t)_{t \in \mathbb{N}}$ exhibits different kinds of dependence; for a brief historical overview, we refer to [18, Sections 4.1, 4.2 and 5.1].

2.4 Approximating Functionals

Now, we give an introduction to r-approximating functionals as considered by Borovkova et al. [18]. First of all, we provide mathematical definitions and discuss some basic properties (Section 2.4.1). Afterwards, we generalize some theorems of [18] in Sections 2.4.2–2.4.3 in order to establish generalizations of two of their main results, namely a law of large numbers (LLN) [18, Theorem 6] and a central limit theorem (CLT) [18, Theorem 7] for U-statistics of r-approximating functionals of absolutely regular processes (Section 2.4.4). If not indicated otherwise, our derivations are based on the joint work [75] with A. Schnurr and M. R. Marín and supplemented with further examples.

2.4.1 Mathematical Definitions

First of all, let us give the definition of a functional.

Definition 2.4.1 (Borovkova et al. [18, Definition 1.3]). Let $(Z_t)_{t \in \mathbb{Z}}$ be a real-valued stationary process. We call a sequence $(X_t)_{t \in \mathbb{Z}}$ a *functional of* $(Z_t)_{t \in \mathbb{Z}}$ if there is a measurable function f defined on $\mathbb{R}^{\mathbb{Z}}$ such that

$$X_t = f((Z_{t+k})_{k \in \mathbb{Z}}).$$

Note that $(X_t)_{t\in\mathbb{Z}}$ is necessarily a stationary process.

The idea is that even though functionals of mixing processes are not necessarily mixing themselves, as long as they can be 'approximated' almost entirely by the 'near epoch' of the mixing process, they still often have properties which allow for the application of limit theorems (cf. Davidson [24, p. 261]). In what follows, we will resort to absolute regular processes as 'background processes'. Recall that absolute regularity constitutes a certain type of mixing condition (see Fig. 2.2).

Definition 2.4.2 (Borovkova et al. [18, Definition 1.2])**.** For a time series $(Z_t)_{t\in\mathbb{Z}}$ and $k, l \in \mathbb{Z}$ with $k \leq l$, define $\mathcal{A}_k^l := \sigma(Z_k, ..., Z_l)$ as the σ-algebra generated by $Z_k, \dots, Z_l$. $(Z_t)_{t\in\mathbb{Z}}$ is called *absolutely regular* if $\beta_k \to 0$ where

$$\begin{aligned}\beta_k &= 2\sup_n \left\{ \sup_{A\in\mathcal{A}_{n+k}^{\infty}} (\mathbb{P}(A|\mathcal{A}_1^n) - \mathbb{P}(A)) \right\} \\ &= \sup_n \left\{ \sup \sum_{i=1}^{I}\sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \right\},\end{aligned}$$

where the last supremum is over all finite $\mathcal{A}_1^n$-measurable partitions $(A_1, ..., A_I)$ and all finite $\mathcal{A}_{n+k}^{\infty}$-measurable partitions $(B_1, ..., B_J)$.

Let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ denote the natural numbers including zero and $\|\cdot\|$ a norm on $\mathbb{R}^d$. Note that the background process $(Z_t)_{t\in\mathbb{Z}}$ is not further specified in the subsequent definition of r-approximating functionals. Furthermore, note that the r-approximating condition does not necessarily require that the r-th order moments of X_0 exist.

Definition 2.4.3 (Borovkova et al. [18, Definition 1.4]/Davidson [24, p. 261f])**.** Let $(X_t)_{t\in\mathbb{Z}}$ be a functional of $(Z_t)_{t\in\mathbb{Z}}$ and let $r \geq 1$. Suppose that $(a_k)_{k\in\mathbb{N}_0}$ are constants with $a_k \to 0$. We say that $(X_t)_{t\in\mathbb{Z}}$ satisfies the *r-approximating condition* or that it is an *r-approximating functional* if

$$\mathbb{E}\left\|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\right\|^r \leq a_k \tag{2.5}$$

for all $k \in \mathbb{N}_0$. The sequence $(a_k)_{k\in\mathbb{N}_0}$ of *approximating constants* is said to be *of size* $-\lambda$ if $a_k = \mathcal{O}(k^{-\lambda-\varepsilon})$ for some $\varepsilon > 0$, where $\mathcal{O}$ denotes the Landau Big O.

r-approximating functionals include one of the most important classes for parametric models of weakly stationary time series, namely *ARMA-processes*. In a nutshell, an ARMA(p, q)-process $(X_t)_{t\in\mathbb{Z}}$ is a weakly stationary process which satisfies the so-called ARMA-equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{2.6}$$

for some constants $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$ and white noise $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$ [20, Definition 3.1.2]. Thereby, *white noise* is defined as a weakly stationary process which has zero mean, whose second moments exist and whose components are pairwise uncorrelated [20, Definition 3.1.1].

Example 2.4.4. First, let us consider the special case of an ARMA(0, 1)-process, namely a *moving average-process of order 1 (short: MA(1)-process)* $(X_t)_{t\in\mathbb{Z}}$ on some white noise $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, 1)$ defined by

$$X_t = \theta Z_{t-1} + Z_t$$

for some constant $\theta \in \mathbb{R}$. We assume that the white noise is Gaussian. For $k \geq 1$, it holds $\mathbb{E}(X_0|\mathcal{A}_{-k}^k) = X_0$ due to measurability of X_0 with respect to $\mathcal{A}_{-k}^k$ and thus, $\mathbb{E}|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)|^r = 0$. Hence, the approximating constants $(a_k)_{k\in\mathbb{N}_0}$ converge to 0 as k tends to infinity, and in fact, they do so very fast. What remains now is to show the boundedness of the r-approximating condition in case of $k = 0$. Note that for jointly normal distributed random variables, uncorrelatedness and independence are equivalent [64, 9.1.2] and all absolute moments exist [64, 2.2.1]. Then, it follows

$$\mathbb{E}(X_0|Z_0) = \theta \cdot \mathbb{E}(Z_{-1}|Z_0) + \mathbb{E}(Z_0|Z_0) = \theta \cdot \mathbb{E}Z_{-1} + Z_0 = Z_0$$

by definition of X_0 and properties of the conditional expectation, and hence,

$$\mathbb{E}|X_0 - \mathbb{E}(X_0|Z_0)|^r = |\theta|^r \cdot \underbrace{\mathbb{E}|Z_0|^r}_{<\infty} =: a_0 < \infty.$$

Example 2.4.5. Now, we return to the more complex case of general ARMA(p, q)-processes, though we only consider the case of 1-approximating functionals ($r = 1$) for illustration. First, let us set the more compact notation

$$\phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p$$
$$\theta(z) := 1 + \theta_1 z + \cdots + \theta_q z^q$$

for the p^{th} and q^{th} degree polynomials and let B denote the backward shift operator defined by

$$B^j X_t := X_{t-j}, \quad j \in \mathbb{Z}.$$

Then, the ARMA-equation (2.6) is equivalent to

$$\phi(B)X_t = \theta(B)Z_t, \quad t \in \mathbb{Z}$$

[20, p. 78]. Let $(Z_t)_{t\in\mathbb{Z}} \sim \text{WN}(0, 1)$ again denote some Gaussian white noise. If $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$, then [20, Theorem 3.1.3] implies that the ARMA-equations have the unique stationary solution

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

where the coefficients ψ_j are defined by

$$\psi(z) := \theta(z)\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j, \quad c^{-1} < |z| < c,$$

for some $c > 1$. With regard to the 1-approximating condition, that is Eq. (2.5) with $r = 1$, for $k \in \mathbb{N}_0$ it then follows

$$\begin{aligned}\mathbb{E}\big|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\big| &= \mathbb{E}\Big|\sum_{j=-\infty}^{\infty} \psi_j Z_{-j} - \mathbb{E}\Big(\sum_{j=-\infty}^{\infty} \psi_j Z_{-j}\big|\mathcal{A}_{-k}^k\Big)\Big| \\ &= \mathbb{E}\Big|\sum_{|j|>k} \psi_j Z_{-j} - \mathbb{E}\Big(\sum_{|j|>k} \psi_j Z_{-j}\big|\mathcal{A}_{-k}^k\Big)\Big|\end{aligned}$$

$$= \mathbb{E}\Big|\sum_{|j|>k} \psi_j Z_{-j} - \mathbb{E}\Big(\sum_{|j|>k} \psi_j Z_{-j}\Big)\Big|$$

$$= \mathbb{E}\Big|\sum_{|j|>k} \psi_j Z_{-j} - \sum_{|j|>k} \psi_j \cdot \mathbb{E}Z_{-j}\Big|$$

due to measurability, independence and the dominated convergence theorem. Note that the coefficients ψ_j are absolutely summable. Since $\mathbb{E}Z_0 = 0$, it follows

$$\mathbb{E}\big|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\big| = \mathbb{E}\Big|\sum_{|j|>k} \psi_j Z_{-j}\Big| \leq \sum_{|j|>k} |\psi_j| \cdot \mathbb{E}\,|Z_{-j}| = \sqrt{\frac{2}{\pi}} \cdot \sum_{|j|>k} |\psi_j| := a_k.$$

These constants are bounded for all $k \in \mathbb{N}_0$ and $a_k \to 0$ as $k \to \infty$. Hence, ARMA-processes which are defined on, e.g., some Gaussian white noise $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,1)$ and satisfy $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$ are 1-approximating functionals.

The class of r-approximating functionals is in fact quite vast. It includes, e.g., Lipschitz functionals (cf. Borovkova et al. [18, Section 1]), but is not limited to them as the subsequent example shows:

Example 2.4.6. A functional defined by a map $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is called *Lipschitz* if there are constants $C > 0$ and $0 < \alpha < 1$ such that for any two sequences $(b_t)_{t\in\mathbb{Z}}$ and $(b'_t)_{t\in\mathbb{Z}}$ satisfying $b_t = b'_t$ for $-k \leq t \leq k$ for some $k \in \mathbb{N}_0$ it holds

$$\big|f((b_t)_{t\in\mathbb{Z}}) - f((b'_t)_{t\in\mathbb{Z}})\big| \leq C\alpha^k$$

[18, Definition 1.3]. Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ denote the extended real number line. We consider the functional $(X_t)_{t\in\mathbb{Z}}$ of $(Z_t)_{t\in\mathbb{Z}}$ defined by the map $X_t = f((Z_{t+k})_{k\in\mathbb{Z}})$ with

$$f : [0,1]^{\mathbb{Z}} \to \overline{\mathbb{R}}, (..., z_{-1}, z_0, z_1, ...) \mapsto \begin{cases} \infty & \text{if } z_1 = 1, \\ 0 & \text{else.} \end{cases}$$

Furthermore, suppose that $(Z_t)_{t\in\mathbb{Z}}$ is i.i.d. and $\mathbb{P}(Z_t \neq 1) = 1$ such that $\mathbb{P}(X_t = 0) = 1$. For $k = 0$ and sequences $(b_t)_{t\in\mathbb{Z}}$ and $(b'_t)_{t\in\mathbb{Z}}$ defined by $b_1 = 1$, $b'_1 \neq 1$ and $b_t = b'_t$ for all $t \in \mathbb{Z} \setminus \{1\}$, it holds

$$\big|f((b_t)_{t\in\mathbb{Z}})) - f((b'_t)_{t\in\mathbb{Z}}))\big| = \infty,$$

that is, $(X_t)_{t\in\mathbb{Z}}$ is not Lipschitz. On the contrary, it holds

$$\mathbb{E}\,|X_0 - \mathbb{E}(X_0|Z_0)| = \mathbb{E}\,|X_0 - \mathbb{E}X_0| \leq 2\mathbb{E}\,|X_0| = 0$$

due to independence, and

$$\mathbb{E}\big|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\big| = \mathbb{E}\,\big|X_0 - \infty \cdot 1_{\{1\}}(Z_1)\big| \leq \mathbb{E}\,|X_0| + \infty \cdot \mathbb{P}(Z_1 = 1) = 0$$

for $k \geq 1$. Therefore, $(X_t)_{t\in\mathbb{Z}}$ is a 1-approximating functional.

Including the examples given by Borovkova et al. [18], so far we have only seen processes that satisfy the r-approximating condition. Therefore, it might be interesting to consider a non-trivial (i.e. well-defined in the sense of existing conditional expectations/first moments) example of a functional which does not satisfy the r-approximating condition.

Example 2.4.7. Let Z and $\widetilde{Z}$ be two independent, identically Pareto(2, 1)-distributed random variables defined on the same probability space and $(Z_t)_{t\in\mathbb{Z}}$ a time series defined by

$$Z_t := \begin{cases} Z & \text{if } t \text{ even,} \\ \widetilde{Z} & \text{if } t \text{ odd.} \end{cases}$$

Note that

$$\mathbb{E}Z^n = \begin{cases} \frac{2}{2-n} & \text{if } 2 > n, \\ \infty & \text{if } 2 \leq n, \end{cases}$$

(see, e.g., [48, Chapter 24]). For another time series $(X_t)_{t\in\mathbb{Z}}$, suppose that $(X_t)_{t\in\mathbb{Z}} = (Z_{t+1})_{t\in\mathbb{Z}}$. For $r > 0$ and $k \geq 0$, then it holds

$$\begin{aligned} \mathbb{E}\big|X_t - \mathbb{E}(X_t|\mathcal{A}_{t-1}^{t+1})\big|^r &= \mathbb{E}|Z_{t+1} - \mathbb{E}(Z_{t+1}|Z_{t-k}, ..., Z_{t+k})|^r \\ &= \begin{cases} \mathbb{E}|Z_{t+1} - \mathbb{E}Z_{t+1}|^r & \text{if } k = 0, \\ \mathbb{E}|Z_{t+1} - Z_{t+1}|^r = 0 & \text{else.} \end{cases} \end{aligned}$$

Considering the case $k = 0$ and $r = 2$, it follows

$$\mathbb{E}|Z_{t+1} - \mathbb{E}Z_{t+1}|^2 = \mathbb{E}(Z_{t+1} - 2)^2 = \mathbb{E}Z_{t+1}^2 - 2 \cdot 2 \cdot \mathbb{E}Z_{t+1} + 2^2 = \infty.$$

Therefore, Example 2.4.7 violates the 2-approximating condition with regard to upper bounds. Admittedly, it is debatable whether a finite number of approximating constants should be allowed to be equal to infinity by definition (see the discussion below). However, the next example gives a sufficient condition on the existence of such bounds.

Example 2.4.8. Let $r \geq 1$ and suppose that $(X_t)_{t\in\mathbb{Z}}$ is a stationary sequence of random variables such that $\mathbb{E}|X_0|^r < \infty$. Then, it holds

$$\begin{aligned} \big|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\big|^r &\leq 2^{r-1}\Big(|X_0|^r + \big|\mathbb{E}(X_0|\mathcal{A}_{-k}^k)\big|^r\Big) \\ &\leq 2^{r-1}\Big(|X_0|^r + \mathbb{E}\big(|X_0|^r\big|\mathcal{A}_{-k}^k\big)\Big) \end{aligned}$$

for all $k \geq 0$, where we used the well-known c_r-inequality given by

$$|U + V|^r \leq c_r \cdot (|U|^r + |V|^r), \tag{2.7}$$

for random variables U and V and

$$c_r := \begin{cases} 2^{r-1} & \text{if } r > 1, \\ 1 & \text{if } 0 < r \leq 1 \end{cases}$$

[50, p. 157] and the conditional Jensen-inequality. By the law of total expectation it holds

$$\mathbb{E}\bigg(2^{r-1}\Big(|X_0|^r + \mathbb{E}\big(|X_0|^r\big|\mathcal{A}_{-k}^k\big)\Big)\bigg) = 2^{r-1}\Big(\mathbb{E}|X_0|^r + \mathbb{E}\big(\mathbb{E}(|X_0|^r|\mathcal{A}_{-k}^k)\big)\Big) = 2^r \cdot \mathbb{E}|X_0|^r < \infty.$$

In particular, for $k \geq 0$, $\mathbb{E}|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)|^r$ is bounded by a series of constants.

The concepts of L_r-near-epoch dependence (for a definition, see, e.g., Davidson [24, Definition 17.1]) and r-approximating functionals are closely linked: Under the assumption of stationarity, possible trends as considered in the definition of L_r-near-epoch dependence are omitted, hence, L_r-near-epoch dependence and the r-approximating condition are equivalent for stationary time series. The only (notational) difference is the size of the constants under consideration: If $(X_t)_{t\in\mathbb{Z}}$ is a stationary L_r-near-epoch dependent time series on $(Z_t)_{t\in\mathbb{Z}}$ with constants $(a_k)_{k\in\mathbb{N}_0}$, then $(X_t)_{t\in\mathbb{Z}}$ is an r-approximating functional of $(Z_t)_{t\in\mathbb{Z}}$ with constants $(a_k^r)_{k\in\mathbb{N}_0}$.

The following lemma is very important for the subsequent reasoning. As we have seen, (2.5) is only demanded to hold at time zero. For stationary time series, this equation is 'shift invariant'. This fact seems to be clear for most authors, but as we could not find a proof in the existing literature, we have decided to close this gap.

Lemma 2.4.9. *Let $(X_t)_{t\in\mathbb{Z}}$ be a (possibly $\mathbb{R}^d$-valued) functional of a stationary time series $(Z_t)_{t\in\mathbb{Z}}$ and $r > 0$. Then, for all $l \geq 0$ it holds*

$$\mathbb{E}\left\|X_t - \mathbb{E}(X_t|\mathcal{A}_{t-l}^{t+l})\right\|^r = \mathbb{E}\left\|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-l}^{l})\right\|^r \tag{2.8}$$

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

Proof. Define the vector $\mathbf{Z}_k^l := (Z_k, ..., Z_l)$, $k < l$, consisting of $l - k + 1$ consecutive components of the time series $(Z_t)_{t\in\mathbb{Z}}$. Let $l \geq 0$ and $t \in \mathbb{Z}$ be fixed. By the Doob–Dynkin lemma (or factorization lemma) [41, Lemma 1.14], there is a map $g : \mathbb{R}^{2l+1} \to \mathbb{R}^d$ such that

$$\mathbb{E}(X_t|\mathbf{Z}_{t-l}^{t+l}) = g\big(\mathbf{Z}_{t-l}^{t+l}\big)$$

holds almost surely. Since g depends only on the distribution of $(X_t, \mathbf{Z}_{t-l}^{t+l})$ [41, p. 167] and $(X_t, \mathbf{Z}_{t-l}^{t+l}) \overset{D}{=} (X_0, \mathbf{Z}_{-l}^{l})$ by stationarity and definition of X_t, it follows

$$\mathbb{E}(X_t|\mathbf{Z}_{t-l}^{t+l}) = \mathbb{E}(X_0|\mathbf{Z}_{-l}^{l}) \quad \text{ a.s.}$$

Note that the random variables do not necessarily need to be measurable with respect to the same σ-algebra. Then, it holds

$$\mathbb{P}\big(X_t - \mathbb{E}(X_t|\mathbf{Z}_{t-l}^{t+l}) \in B\big) = \mathbb{P}\big(X_t - \mathbb{E}(X_0|\mathbf{Z}_{-l}^{l}) \in B\big) = \mathbb{P}\big(X_0 - \mathbb{E}(X_0|\mathbf{Z}_{-l}^{l}) \in B\big)$$

for all $B \in \mathcal{B}(\mathbb{R})$, which concludes the proof. □

At this point we would like to discuss the possibility of some approximating constants a_k being equal to infinity in more detail. In general, the r-approximating condition is not violated by definition if there is a finite number of approximating constants which are equal to infinity as long as the sequence still converges to 0. Therefore, with regard to limit theorems a problem can only arise if an additional summability condition is imposed on the approximating constants, i.e. something along the lines of $\sum_{k=0}^{\infty} a_k < \infty$. This is precisely the case for the limit theorems of Borovkova et al. [18] which we would like to employ in Chapter 4 in the more generalized setting mentioned before. However, the summability condition in their limit theorems can be weakened such that summability is only required for all but a finite number of approximating constants. Even though our proofs follow mainly those of Borovkova et al. [18], they are still very technical and not straightforward.

Without loss of generality, in what follows we assume that the first d approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of some time series $(X_t)_{t\in\mathbb{Z}}$ are infinite, i.e., $a_0 = \cdots = a_{d-1} = \infty$ for some finite constant $d > 0$. Therefore, we consider altered summability conditions of the form $\sum_{k=d} a_k < \infty$ or $\sum_{k=d} k^2 a_k < \infty$. The proof in case of $a_{i_0} = \cdots = a_{i_{d-1}} = \infty$ for some indices $i_0, \ldots, i_{d-1} \in \mathbb{N}$ works analogously.

In order to consider the main theorems, which are given by a law of large numbers (Theorem 2.4.16) and a central limit theorem (Theorem 2.4.17) for U-statistics of r-approximating functionals, first we consider some of the preceding results of Borovkova et al. [18] with regard to our generalized summability condition. Following the authors' ideas, we are going to use these for the proofs of the main results.

However, the reader may also skip the subsequent sections initially and look up the relevant theorems later.

2.4.2 Near-epoch Dependence and Near Regularity

We begin by defining blocking, which constitutes a useful tool, and near regularity.

Definition 2.4.10 (Borovkova et al. [18, Definitions 2.3/2.6])**.**

1. Let $(X_t)_{t\in\mathbb{N}}$ be a stochastic process, let $M, N \in \mathbb{N}$ be positive integers and moreover assume that M is even. An (M,N)*-blocking of* $(X_t)_{t\in\mathbb{N}}$ is defined as the sequence of blocks $B_1, B_2, \ldots$ each consisting of N consecutive X_t's, where each two consecutive blocks are separated by blocks of length M, more precisely,

$$B_s = (X_{(s-1)(M+N)+\frac{M}{2}+1}, \ldots, X_{s(M+N)-\frac{M}{2}}),\ s \geq 1.$$

   The set of indices in a block B_s is denoted by $I_s := \{(s-1)(M+N) + \frac{M}{2} + 1, \ldots, s(M+N) - \frac{M}{2}\}$.

2. A stochastic process $(X_t)_{t\in\mathbb{N}}$ is called *nearly regular* if for any $\varepsilon, \delta > 0$ there exists an $M \in \mathbb{N}$ such that for all $N \in \mathbb{N}$ we can find a sequence $(B'_s)_{s\in\mathbb{N}}$ consisting of independent $\mathbb{R}^N$-valued random vectors which satisfy the following two conditions:

   a) Let B_s denote the s-th (M,N)-block of $(X_t)_{t\in\mathbb{N}}$. Then, B'_s has the same distribution as B_s.

   b) It holds $\mathbb{P}(\|B_s - B'_s\| \leq \delta) \geq 1 - \varepsilon$, where $\|\cdot\|$ denotes the L_1-norm on $\mathbb{R}^N$, that is, $\|x\| = \sum_{i=1}^N |x_i|$.

"Absolute regularity of a process implies that the sequence of (M,N)-blocks can be perfectly coupled with the sequence of independent long blocks, which have the same distribution as those of the original process" [18, p. 4271]. Near regularity in turn refers to the similarity to absolute regularity in terms of this property, that is, near regularity implies closeness to such a process.

Theorem 3 of [18] shows that 1-approximating functionals with summable approximating constants are nearly regular. However, since the authors have not used the assumed summability condition in their proof, their result remains valid even if the summability condition is not satisfied at all, though it then becomes trivial. For our purpose it is enough to consider the following generalization:

Theorem 2.4.11. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process with mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$. If there are integers $K, L, N \in \mathbb{N}$, K even, such that $\sum_{k=L}^{\infty} a_k < \infty$, then we can approximate the sequence of $(K+2L, N)$-blocks $(B_s)_{s\in\mathbb{N}}$ by a sequence of independent blocks $(B'_s)_{s\in\mathbb{N}}$ with the same marginal distribution in such a way that*

$$\mathbb{P}(\|B_s - B'_s\| \leq 2\alpha_L) \geq 1 - \beta_K - 2\alpha_L, \tag{2.9}$$

where

$$\alpha_L := (2\sum_{k=L}^{\infty} a_l)^{1/2}, \tag{2.10}$$

so $(X_t)_{t\in\mathbb{N}}$ is nearly regular by definition.

2.4.3 Moment Inequalities and a Central Limit Theorem for Partial Sums

Now we generalize some inequalities for second and fourth order moments of partial sums $\xi_n := X_1 + \cdots + X_n$ of 1-approximating functionals of absolutely regular processes which have been given in [18]. Note that such partial sums are often denoted by S_n in the literature. However, in order to avoid confusion with regard to the notation S_d, which will be introduced in the subsequent chapter and is highly accepted and standard in the ordinal pattern community, we decided to opt for ξ_n instead.

Our proofs are closely linked to the original proofs of Borovkova et al. [18], but our basic idea is to split the respective sums appearing in the proofs according to our altered summability condition. Hence, the difference is that we need to find some other bounds for the first summands.

Lemma 2.4.12. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process with mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$. Moreover, suppose that $\mathbb{E}X_0 = 0$ and that one of the following conditions holds for a fixed integer $d \geq 0$:*

1. *X_0 is bounded a.s. and $\sum_{k=d}^{\infty}(a_k + \beta_k) < \infty$.*
2. *$\mathbb{E}|X_0|^{2+\delta} < \infty$ for some $\delta > 0$ and $\sum_{k=d}^{\infty}(a_k^{\frac{\delta}{1+\delta}} + \beta_k^{\frac{\delta}{2+\delta}}) < \infty$.*

Then it follows

$$\frac{1}{n}\mathbb{E}\xi_n^2 \to \mathbb{E}X_0^2 + 2\sum_{k=1}^{\infty}\mathbb{E}(X_0X_k) \tag{2.11}$$

as $n \to \infty$ and the sum on the r.h.s. converges absolutely.

Proof. The proofs under the different conditions are basically the same using the respective results of Lemma 2.18 in [18]. Therefore, we just give a proof under the second condition. It holds

$$\begin{aligned}\mathbb{E}\xi_n^2 &= \sum_{1\leq i,j\leq n}\mathbb{E}X_iX_j \\ &= n\mathbb{E}X_0^2 + 2\sum_{k=1}^{n}(n-k)\mathbb{E}X_0X_k\end{aligned}$$

$$= n\Big(\mathbb{E}X_0^2 + 2\Big(\sum_{k=1}^{3d-1}(1-\frac{k}{n})\mathbb{E}X_0X_k + \sum_{k=3d}^{n}(1-\frac{k}{n})\mathbb{E}X_0X_k\Big)\Big)$$

by stationarity of $(X_t)_{t\in\mathbb{N}}$.
Due to $\mathbb{E}|X_0|^{2+\delta} < \infty$, $\mathbb{E}X_0^2$ as well as $\mathbb{E}X_0X_k$ are bounded by a constant $C_1 > 0$ for $1 \le k < 3d$, so in particular, $\sum_{k=1}^{3d-1}(1-\frac{k}{n})\mathbb{E}X_0X_k$ is bounded by $(3d-1)C_1$. (Note the difference to [18] at this point.). Since $\mathbb{E}X_0 = 0$ by assumption, for $k \ge 3d$, Lemma 2.18 (ii) of [18] yields

$$|\mathbb{E}X_0X_k| \le 4\,\|X_0\|_{2+\delta}^{\frac{\delta}{1+\delta}}\,(a_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{1+\delta}} + 2\,\|X_0\|_{2+\delta}^2\,(\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{2+\delta}}.$$

Hence,

$$\sum_{k=3d}^{\infty}|\mathbb{E}X_0X_k| \le C_2\sum_{k=3d}^{\infty}\Big((a_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{1+\delta}} + (\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{2+\delta}}\Big) = 3C_2\sum_{k=d}^{\infty}\Big(a_k^{\frac{\delta}{1+\delta}} + \beta_k^{\frac{\delta}{2+\delta}}\Big)$$

converges absolutely by assumption, where $C_2 := \max\{4\,\|X_0\|_{2+\delta}^{\frac{\delta}{1+\delta}}, 2\,\|X_0\|_{2+\delta}^2\}$. Then the dominated convergence theorem yields

$$\sum_{k=3d}^{n}(1-\frac{k}{n})\mathbb{E}X_0X_k \to \sum_{k=3d}^{\infty}\mathbb{E}X_0X_k$$

as $n \to \infty$, which proves our claim. □

We continue with an inequality for fourth order moments of partial sums:

Lemma 2.4.13. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process with mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$. Moreover, suppose that $\mathbb{E}X_0 = 0$ and that one of the following conditions holds for a fixed integer $d \ge 0$:*

1. *X_0 is bounded a.s. and $\sum_{k=d}^{\infty}k^2(a_k+\beta_k) < \infty$.*
2. *$\mathbb{E}|X_0|^{4+\delta} < \infty$ for some $\delta > 0$ and $\sum_{k=d}^{\infty}k^2(a_k^{\frac{\delta}{3+\delta}} + \beta_k^{\frac{\delta}{4+\delta}}) < \infty$.*

Then, there exists a constant $C > 0$ such that

$$\mathbb{E}\xi_n^4 \le Cn^2. \tag{2.12}$$

The proof follows the same idea, but is slightly more involved.

Proof. Again, the proofs are very similar using the respective results of Lemma 2.18 as well as Lemma 2.21 and Lemma 2.22 of [18], respectively, so we only give the proof with regard to the second condition. By stationarity it holds

$$\begin{aligned}\mathbb{E}\xi_n^4 &\le 4!\sum_{1\le i_1\le i_2\le i_3\le i_4\le n}|\mathbb{E}(X_{i_1}X_{i_2}X_{i_3}X_{i_4})| \\ &\le 4!n\sum_{\substack{i,j,k\ge 0\\ i+j+k\le n}}|\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|. \end{aligned}\tag{2.13}$$

We can split the sum appearing in (2.13) in the following way:

$$
\begin{aligned}
&\sum_{\substack{i,j,k\geq 0\\ i+j+k\leq n}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\leq \sum_{0\leq j,k\leq i\leq n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| + \sum_{0\leq i,j\leq k\leq n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\qquad + \sum_{0\leq i,k\leq j\leq n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&= \underbrace{\sum_{0\leq j,k\leq i<3d} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|}_{\text{(I)}} + \sum_{\substack{3d\leq i\leq n\\ 0\leq j,k\leq i}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\qquad + \underbrace{\sum_{0\leq i,j\leq k<3d} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|}_{\text{(II)}} + \sum_{\substack{3d\leq k\leq n\\ 0\leq i,j\leq k}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\qquad + \underbrace{\sum_{0\leq i,k\leq j<3d} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|}_{\text{(III)}} + \underbrace{\sum_{\substack{3d\leq j\leq n\\ 0\leq i,k\leq j\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|}_{\text{(IV)}} \\
&\qquad\qquad + \underbrace{\sum_{3d\leq i,k\leq j\leq n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})|}_{\text{(V)}}.
\end{aligned}
$$

Due to $\mathbb{E}|X_0|^{4+\delta} < \infty$ and the assumed stationarity, there exists a constant $C_1 > 0$ such that $|\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| < C_1$ (as well as a constant $C_2 > 0$ such that $|\mathbb{E}(X_0X_i)| < C_2$), and hence, the finite sums (I), (II) and (III) which are independent of n are obviously finite, too. In particular it follows

$$
\sum_{0\leq j,k\leq i<3d} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \leq (3d)^3 \cdot C_1 = 27d^3 \cdot C_1
$$

for sums of this type. By adding a zero-valued term we obtain

$$
\begin{aligned}
|\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| = |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| &- |\mathbb{E}(X_0X_i)| \cdot |\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&+ |\mathbb{E}(X_0X_i)| \cdot |\mathbb{E}(X_{i+j}X_{i+j+k})|
\end{aligned}
$$

such that we can find the following bound for (IV) + (V):

$$
\begin{aligned}
&\sum_{\substack{3d\leq j\leq n\\ 0\leq i,k\leq j\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| + \sum_{3d\leq i,k\leq j\leq n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\qquad = \sum_{\substack{3d\leq j\leq n\\ 0\leq i,k\leq j\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| - |\mathbb{E}(X_0X_i)| \cdot |\mathbb{E}(X_{i+j}X_{i+j+k})|
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{\substack{3d\le j\le n\\ 0\le i,k\le j\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&+ \sum_{3d\le i,k\le j\le n} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| - |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&+ \sum_{3d\le i,k\le j\le n} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
\le& \sum_{\substack{3d\le j\le n\\ 0\le i,k\le j}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| - |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&+ n\sum_{3d\le i,k\le n} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| + n\sum_{\substack{0\le i,k\le n\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)|\,.
\end{aligned}
$$

Note that there we have summed up the first and the third term and we have bounded the second and fourth term by using stationarity. Accordingly it follows

$$
\begin{aligned}
&\sum_{\substack{i,j,k\ge 0\\ i+j+k\le n}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\quad\le 3\cdot 27d^3\cdot C_1 + \sum_{\substack{3d\le i\le n\\ 0\le j,k\le i}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| + \sum_{\substack{3d\le k\le n\\ 0\le i,j\le k}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| \\
&\qquad + \sum_{\substack{3d\le j\le n\\ 0\le i,k\le j}} |\mathbb{E}(X_0X_iX_{i+j}X_{i+j+k})| - |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&\qquad + n\cdot\sum_{3d\le i,k\le n} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| + n\cdot\sum_{\substack{0\le i,k\le n\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)|\,.
\end{aligned}
\tag{2.14}
$$

The idea is now to apply different upper bounds on the remaining summands as it has been done by Borovkova et al. [18]. In more detail: Since $\mathbb{E}X_0 = 0$, Lemma 2.22 of [18] implies

$$
\begin{aligned}
|\mathbb{E}(X_0(X_iX_{i+j}X_{i+j+k}))| &\le 6(\beta_{\lfloor \frac{i}{3}\rfloor})^{\frac{\delta}{4+\delta}}\|X_0\|_{4+\delta}^4 + 8(a_{\lfloor \frac{i}{3}\rfloor})^{\frac{\delta}{3+\delta}}\|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}} \\
|\mathbb{E}((X_0X_iX_{i+j})X_{i+j+k})| &\le 6(\beta_{\lfloor \frac{k}{3}\rfloor})^{\frac{\delta}{4+\delta}}\|X_0\|_{4+\delta}^4 + 8(a_{\lfloor \frac{k}{3}\rfloor})^{\frac{\delta}{3+\delta}}\|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}\,.
\end{aligned}
$$

Furthermore, application of Lemma 2.22 and Lemma 2.18 of [18], respectively, yields

$$
\begin{aligned}
&|\mathbb{E}((X_0X_i)(X_{i+j}X_{i+j+k}))| - |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_{i+j}X_{i+j+k})| \\
&\qquad\le 6(\beta_{\lfloor \frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\|X_0\|_{4+\delta}^4 + 8(a_{\lfloor \frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}}\|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}
\end{aligned}
$$

and

$$
\begin{aligned}
|\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| \le& \left(2\|X_0\|_{4+\delta}^2(\beta_{\lfloor \frac{i}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + 4\|X_0\|_{4+\delta}^{\frac{4+\delta}{3+\delta}}(a_{\lfloor \frac{i}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right) \\
&\times\left(2\|X_0\|_{4+\delta}^2(\beta_{\lfloor \frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + 4\|X_0\|_{4+\delta}^{\frac{4+\delta}{3+\delta}}(a_{\lfloor \frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right).
\end{aligned}
$$

Hence, we obtain

$$
\begin{aligned}
&\sum_{\substack{0\le i,k\le n\\ i<3d \text{ or } k<3d}} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| \\
&\qquad \le \sum_{0\le i,k\le 3d} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| + 2\cdot \sum_{\substack{0\le i<3d\\ 3d\le k\le n}} |\mathbb{E}(X_0X_i)|\cdot|\mathbb{E}(X_0X_k)| \\
&\qquad \le (3d)^2C_2^2 + 2\cdot 3d\cdot C_2\sum_{k=3d}^{n}|\mathbb{E}(X_0X_k)| \\
&\qquad \le 9d^2C_2^2 + 6d\cdot C_2\sum_{k=3d}^{n}\left(2\,\|X_0\|_{4+\delta}^2\,(\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + 4\,\|X_0\|_{4+\delta}^{\frac{4+\delta}{3+\delta}}\,(a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)
\end{aligned}
$$

with regard to the last summand appearing in (2.14). (Note that we have used Lemma 2.18 of [18] for the last inequality.) Therefore, there is a constant $C_3>0$ such that

$$
\begin{aligned}
\mathbb{E}\xi_n^4 \le{}& (4!\cdot 81d^3\cdot C_1)n + (4!\cdot 9d^2C_2^2)n^2 + C_3 n\Bigg(\sum_{\substack{3d\le i\le n\\ 0\le j,k\le i}}\left((a_{\lfloor\frac{i}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{i}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) \\
&+ \sum_{\substack{3d\le k\le n\\ 0\le i,j\le k}}\left((a_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) + \sum_{\substack{3d\le j\le n\\ 0\le i,k\le j}}\left((a_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) \\
&+ n\cdot\sum_{3d\le i,k\le n}\left((\beta_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right) \\
&+ n\sum_{k=3d}^{n}\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)\Bigg) \\
\le{}& (4!\cdot 81d^3\cdot C_1)n + (4!\cdot 9d^2C_2^2)n^2 + C_3 n\Bigg(3\sum_{\substack{3d\le j\le n\\ 0\le i,k\le j}}\left((a_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) \\
&+ n\sum_{3d\le i,k\le n}\left((\beta_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right) \\
&+ n\sum_{3d\le k\le n}\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)\Bigg).
\end{aligned}
$$

Due to the flooring and the assumed summability of the constants it holds

$$
\begin{aligned}
\sum_{\substack{3d\le j\le n\\ 0\le i,k\le j}}\left((a_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) &= \sum_{j=3d}^{n}\sum_{i,k=0}^{j}\left((a_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) \\
&= \sum_{j=3d}^{n}(j+1)^2\left((a_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{3+\delta}} + (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{\delta}{4+\delta}}\right) \\
&\le 3\sum_{j=d}^{n/3}(3j+1)^2\left(a_j^{\frac{\delta}{3+\delta}} + \beta_j^{\frac{\delta}{4+\delta}}\right)
\end{aligned}
$$

$$\leq 3\sum_{j=d}^{\infty} 9(j+1)^2 \left(a_j^{\frac{\delta}{3+\delta}} + \beta_j^{\frac{\delta}{4+\delta}}\right) < \infty.$$

In a similar manner we obtain

$$\begin{aligned}
\sum_{i,k=3d}^{n} &\left((\beta_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{i}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right) \\
&= \left(\sum_{j=3d}^{n} (\beta_{\lfloor\frac{j}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{j}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right)^2 \\
&\leq 9\left(\sum_{j=d}^{\infty} \beta_j^{\frac{2+\delta}{4+\delta}} + a_j^{\frac{2+\delta}{3+\delta}}\right)^2 < \infty
\end{aligned}$$

and $\sum_{3d\leq k\leq n}\left((\beta_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{4+\delta}} + (a_{\lfloor\frac{k}{3}\rfloor})^{\frac{2+\delta}{3+\delta}}\right) < \infty$, which concludes the proof. □

The central limit theorem for partial sums of functionals of absolutely regular processes given by Borovkova et al. [18, Theorem 4] can be directly extended with regard to our slightly weaker summability condition. The respective proofs are very much the same: First of all, the introduced $(K_n + 2L_n, N_n)$-blocking which has to satisfy conditions (3.5)-(3.8) of [18] can be chosen such that $L_n \geq d$. Then, using Lemma 2.4.12 and 2.4.13 instead of Lemma 2.23 and 2.24 of [18], respectively, already yields the desired result:

Theorem 2.4.14. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process with mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$. Furthermore, suppose that $\mathbb{E}X_0 = 0$, $\mathbb{E}|X_0|^{4+\delta} < \infty$ and*

$$\sum_{k=d}^{\infty} k^2 \left(a_k^{\frac{\delta}{3+\delta}} + \beta_k^{\frac{\delta}{4+\delta}}\right) < \infty$$

for some $\delta > 0$ and a fixed integer $d \geq 0$. Then, as $n \to \infty$,

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

where $\sigma^2 = \mathbb{E}X_0^2 + 2\sum_{k=1}^{\infty}\mathbb{E}(X_0X_k)$. In case $\sigma^2 = 0$, we adopt the convention that $N(0,0)$ denotes the point mass at the origin.

2.4.4 Law of large numbers and central limit theorem for U-statistics

In order to establish the generalizations of the LLN and CLT of U-statistics of r-approximating functionals, that is [18, Theorem 6] and [18, Theorem 7], respectively, first we need to give the following technical condition:

Definition 2.4.15 (Borovkova et al. [18, Definition 2.12])**.** Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary time series and $h : \mathbb{R}^{2d} \to \mathbb{R}$ be a measurable, symmetric kernel. Then we say that h is *p-continuous* if there exists a function $\phi :]0,\infty[\to]0,\infty[$ with $\phi(\varepsilon) = o(1)$ as $\varepsilon \to 0$ such that

$$\mathbb{E}\left|h(U,V) - h(\tilde{U},V)\right|^p \mathbb{1}_{\{\|U-\tilde{U}\|\leq\varepsilon\}} \leq \phi(\varepsilon) \tag{2.15}$$

for all random variables U, $\tilde{U}$, V with marginal distribution $F = F_{Y_0}$ and such that (U, V) either has distribution $F \times F$ (independent case) or $\mathbb{P}_{(Y_0,Y_t)}$ for some $t \in \mathbb{N}$, where the latter denotes the joint distribution of Y_0 and Y_t.

Note that we state this definition in the multivariate version. The Sections 1 to 5 of [18] are all written down for the one-dimensional case. However, in the sixth section, the one dealing with entropy concepts, the authors state that all the previous results hold true in a d-dimensional setting as well. Instead of calling a kernel h p-continuous, some authors say that h satisfies the *p-Lipschitz condition.*

Now, utilizing the proof given by Borovkova et al. [18] though using Theorem 2.4.11 instead of their Theorem 3, we already obtain a law of large numbers for U-statistics of functionals of absolutely regular processes under our weaker summability condition:

Theorem 2.4.16. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process, where $\sum_{k=d}^{\infty} a_k < \infty$ for some integer $d \geq 0$. Furthermore, suppose that $h : \mathbb{R}^{2d} \to \mathbb{R}^d$ is a measurable and symmetric function which is* 1*-continuous, and that the family of random variables $\{h(X_s, X_t) : s, t \geq 1\}$ is uniformly integrable. Then,*

$$U_n(h) = \frac{1}{n(n-1)} \sum_{\substack{1\leq s,t\leq n \\ s\neq t}} h(X_s, X_t) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^{2d}} h(x, y)dF(x)dF(y) =: \theta,$$

as $n \to \infty$.

What remains to show in order to complete our generalized theory on limit theorems for U-statistics for r-approximating functionals is a central limit theorem:

Theorem 2.4.17. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional with approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of an absolutely regular process with mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$, and let h be a bounded and 1-continuous kernel. Suppose that the sequences $(a_k)_{k\in\mathbb{N}_0}$, $(\beta_k)_{k\in\mathbb{N}_0}$ and $(\phi(a_k))_{k\in\mathbb{N}_0}$ satisfy the summability condition*

$$\sum_{k=d}^{\infty} k^2(\beta_k + a_k + \phi(a_k)) < \infty \tag{2.16}$$

for some fixed integer $d \geq 0$. Then the series

$$\sigma^2 = \operatorname{Var}(h_1(X_0)) + 2\sum_{k=1}^{\infty} \operatorname{Cov}(h_1(X_0), h_1(X_k)) \tag{2.17}$$

converges absolutely and, as $n \to \infty$,

$$\sqrt{n}(U_n(h) - \theta) \xrightarrow{D} N(0, 4\sigma^2).$$

Note that there is a typo in the long-run variance given in [18, Theorem 7]: In fact, the variance should not be squared. This can be verified by considering the proof, which, among others, is based on [18, Theorem 4]. There, the long-run variance with respect to centered 1-approximating functionals of absolute regular processes is given correctly.

The proof is again very similar to the one given by Borovkova et al. [18]. The main idea is to make use of the Hoeffding decomposition (2.4). Then we use Theorem 2.4.14 (instead of [18, Theorem 4]) to show the convergence of the linear part, that is

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{n}(h_1(X_t)-\theta)\xrightarrow{D}N(0,4\sigma^2).$$

Let us emphasize at this point that $h_1(X_t)-\theta$ is bounded, since h is a bounded kernel. Hence, the moment assumption in Theorem 2.4.14 is satisfied. We do not use Theorem 2.4.14 on $(X_t)_{t\in\mathbb{N}}$ itself, therefore we do not need any additional assumptions on the original process. For more details on this part of the proof, we refer the reader to [18, p. 4301].

Defining

$$R_n := \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}J(X_i,X_j)$$

as the degenerate part of the Hoeffding decomposition (2.4), it then remains to show that $\sqrt{n}R_n$ is asymptotically negligible under our altered summability condition instead of using [18, Lemma 4.4]:

Lemma 2.4.18. *Under the conditions of Theorem 2.4.17, it holds*

$$\sup_n \mathbb{E}\left(\frac{1}{n}\sum_{1\leq i<j\leq n}J(X_i,X_j)\right)^2<\infty, \tag{2.18}$$

and therefore, $\sqrt{n}R_n\xrightarrow{\mathbb{P}}0$*, as* $n\to\infty$.

In comparison to the analogous result [18, Lemma 4.4], we need to ensure convergence of series for which we cannot find a bound using our summability condition. Nevertheless, our proof is still similar to the one given by the authors.

Proof of Lemma 2.4.18. For $k\in\{i_1,i_2,j_1,j_2\}$, let $\mathbb{E}_{X_k}(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2}))$ denote the expected value of $J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})$ taken with respect to the random variable X_k, with the remaining variables kept fixed. (We adopt this notation from [18].) Since $J(x,y)$ is a degenerate kernel, i.e., $\int J(x,y)dF(x)=0$ for all $y\in\mathbb{R}^d$ where F denotes the cdf of the margins X_t, it follows

$$\mathbb{E}_{X_{j_2}}(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2}))=\int_{\mathbb{R}^d}J(X_{i_1},X_{j_1})J(X_{i_2},y)dF(y)=0, \tag{2.19}$$

and similarly for $\mathbb{E}_{X_{i_1}}$, $\mathbb{E}_{X_{i_2}}$ and $\mathbb{E}_{X_{j_1}}$. As both h and h_1 are bounded by definition and 1-continuous by [18, Lemma 2.15]), the same holds for $g(x_1,x_2,x_3,x_4)=J(x_1,x_2)J(x_3,x_4)$ due to [18, Lemma 2.14]. Hence, there is a constant $C_1>0$ such that

$$\mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)<C_1, \tag{2.20}$$

$$\mathbb{E}_{X_{i_1},X_{i_2}}\mathbb{E}_{X_{j_1},X_{j_2}}(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2}))<C_1. \tag{2.21}$$

By linearity of the expected value it holds

$$\mathbb{E}\left(\sum_{1\leq i<j\leq n}J(X_i,X_j)\right)^2=\mathbb{E}\left(\sum_{\substack{1\leq i_1<j_1\leq n\\1\leq i_2<j_2\leq n}}J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)$$

$$= \sum_{\substack{1\le i_1<j_1\le n\\ 1\le i_2<j_2\le n\\ i_1=i_2 \text{ and } j_1=j_2}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)$$

$$+ \sum_{\substack{1\le i_1<j_1\le n\\ 1\le i_2<j_2\le n\\ i_1\ne i_2 \text{ or } j_1\ne j_2}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right). \tag{2.22}$$

With regard to the first sum given by $\sum_{1\le i<j\le n}\mathbb{E}\left(J(X_i,X_j)J(X_i,X_j)\right)$ there are at most $n(n-1)/2$ summands, so

$$\sum_{1\le i<j\le n}\mathbb{E}\left(J(X_i,X_j)J(X_i,X_j)\right)\le n^2C_1. \tag{2.23}$$

In the remainder, we consider the sum given by (2.22). First, let us assume that at least one index is different from all the others, e.g., j_2, and suppose that $i_1\le i_2\le j_1<j_2$. Let Δ_i denote the i-th largest difference between two consecutive indices. In contrast to [18] we need to split the sum according to some values of Δ_i and find a different bound for it for small values of Δ_i. If $\Delta_1=j_2-j_1$, then

$$\sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)$$

$$= \sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1<3d}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right) + \sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1\ge 3d}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right). \tag{2.24}$$

With regard to the first sum appearing in (2.24), there are $3d-1$ values possible for Δ_1 and each of them can be obtained less than n times by shifting j_1 and $j_2=j_1+\Delta_1$. Furthermore, for fixed j_1, at most $3d$ choices for i_2 are possible by definition of $0\le\Delta_3\le\Delta_2\le\Delta_1$. The same holds true for i_1 and fixed i_2, which leads to

$$\sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1<3d}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)\le n(3d)^3C_1.$$

Now we consider second sum appearing in (2.24). Eq. (2.19) yields

$$\begin{aligned}\mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right) &= \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)\\ &\quad - \mathbb{E}_{X_{i_1},X_{j_1},X_{i_2}}\mathbb{E}_{X_{j_2}}(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})),\end{aligned}$$

so application of Lemma 4.3 of [18] (with $r=\infty$, $s=1$ and $M=C_1$) yields

$$\begin{aligned}&\sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1\ge 3d}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)\\ &\le 4C_1 \sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1=j_2-j_1\ge 3d}} \left(\beta_{\lfloor\Delta_1/3\rfloor}+a_{\lfloor\Delta_1/3\rfloor}+2\phi\left(a_{\lfloor\Delta_1/3\rfloor}\right)\right)\\ &= 4C_1 \sum_{\substack{1\le j_1<j_2\le n\\ \Delta_1=j_2-j_1\ge 3d}} (\Delta_1+1)^2\left(\beta_{\lfloor\Delta_1/3\rfloor}+a_{\lfloor\Delta_1/3\rfloor}+2\phi\left(a_{\lfloor\Delta_1/3\rfloor}\right)\right),\end{aligned}$$

since there are $\Delta_1 + 1$ possibilities for i_2 for each j_1 due to $0 \leq j_1 - i_2 \leq \Delta_1$ by assumption and, dependent on the choice of i_2, the same holds true for i_1, which results in a total of $(\Delta_1 + 1)^2$ possibilities for i_1 and i_2 for each j_1. Furthermore, for each $\Delta_1 \geq 3d$ there are $n - \Delta_1$ possible choices for pairs (j_1, j_2) such that $1 \leq j_1 < j_2 \leq n$ and $\Delta_1 = j_2 - j_1$, which yields

$$
\begin{aligned}
4C_1 \sum_{\substack{1\leq j_1<j_2\leq n \\ \Delta_1=j_2-j_1\geq 3d}} & (\Delta_1+1)^2 \left(\beta_{\lfloor \Delta_1/3\rfloor} + a_{\lfloor \Delta_1/3\rfloor} + 2\phi\left(a_{\lfloor \Delta_1/3\rfloor}\right)\right) \\
&\leq 4C_1 n \sum_{\Delta_1=3d}^{n-3d} (\Delta_1+1)^2 \left(\beta_{\lfloor \Delta_1/3\rfloor} + a_{\lfloor \Delta_1/3\rfloor} + 2\phi\left(a_{\lfloor \Delta_1/3\rfloor}\right)\right) \\
&\leq 4C_1 n \sum_{k=d}^{\lfloor \frac{n-3d}{3}\rfloor} (3k+1)^2 \left(\beta_k + a_k + 2\phi\left(a_k\right)\right) \\
&\leq 36 C_1 n \sum_{k=d}^{n} (k+1)^2 \left(\beta_k + a_k + 2\phi\left(a_k\right)\right).
\end{aligned}
$$

If $\Delta_1 \neq j_2 - j_1$, then, depending on whether Δ_1 is located between i_1 and i_2 or i_2 and j_1, by applying Lemma 4.3 of [18] twice, we obtain a bound for $\mathbb{E}(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2}))$. E.g., if $\Delta_1 = j_1 - i_2$, then

$$
\begin{aligned}
&\mathbb{E}(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})) \\
&\quad = \mathbb{E}(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})) - \mathbb{E}_{X_{i_1},X_{i_2}}\mathbb{E}_{X_{j_1},X_{j_2}}(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})) \\
&\qquad\quad + \mathbb{E}_{X_{i_1},X_{i_2}}\mathbb{E}_{X_{j_1},X_{j_2}}(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})) \\
&\quad \leq 4C_1\big(\beta_{\lfloor \frac{\Delta_1}{3}\rfloor} + a_{\lfloor \frac{\Delta_1}{3}\rfloor}\big) + 2\phi\big(a_{\lfloor \frac{\Delta_1}{3}\rfloor}\big) + 4C_1\big(\beta_{\lfloor \frac{\Delta_2}{3}\rfloor} + a_{\lfloor \frac{\Delta_2}{3}\rfloor}\big) + 2\phi\big(a_{\lfloor \frac{\Delta_2}{3}\rfloor}\big), \qquad (2.25)
\end{aligned}
$$

where we have used (2.19) as well as (2.20) and (2.21). Note that due to Fubini's theorem, we can always change the order of the expected values in Lemma 4.3 of [18] such that we always obtain the result with respect to Δ_2. Furthermore, note that $\Delta_2 \geq 3d$ implies $\Delta_1 \geq 3d$. Now considering (2.22) again, in this case it holds

$$
\begin{aligned}
&\sum_{\substack{1\leq i_1\leq i_2\leq j_1<j_2\leq n \\ j_2-j_1\neq\Delta_1}} \mathbb{E}\left(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})\right) \\
&= \sum_{\substack{1\leq i_1\leq i_2\leq j_1<j_2\leq n \\ \Delta_1\neq j_2-j_1 \\ \Delta_2<3d}} \mathbb{E}\left(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})\right) + \sum_{\substack{1\leq i_1\leq i_2\leq j_1<j_2\leq n \\ \Delta_1\neq j_2-j_1 \\ \Delta_2\geq 3d}} \mathbb{E}\left(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})\right).
\end{aligned}
$$

The first sum is bounded by

$$
\sum_{\substack{1\leq i_1\leq i_2\leq j_1<j_2\leq n \\ \Delta_1\neq j_2-j_1 \\ \Delta_2<3d}} \mathbb{E}\left(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})\right) \leq \sum_{\substack{1\leq i_1\leq i_2\leq j_1<j_2\leq n \\ \Delta_1\neq j_2-j_1 \\ \Delta_2<3d}} C_1 \leq n^2(3d)^2 C_1,
$$

as there are $3d$ possible values for Δ_2, which can be each obtained less than n times, and for each Δ_2 there are $3d$ possible choices for Δ_3. For a given Δ_2, we do not have much information on Δ_1, so we multiply with the factor n. On the other hand, by application of (2.25) for the second sum it holds

$$\begin{aligned}
&\sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1\neq j_2-j_1\\ \Delta_2\ge 3d}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)\\
&\quad\le 4C_1 \sum_{\substack{1\le i_1\le i_2\le j_1<j_2\le n\\ \Delta_1\neq j_2-j_1\\ \Delta_2\ge 3d}} \Big(\beta_{\lfloor\frac{\Delta_1}{3}\rfloor}+a_{\lfloor\frac{\Delta_1}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_1}{3}\rfloor}\big)+\beta_{\lfloor\frac{\Delta_2}{3}\rfloor}+a_{\lfloor\frac{\Delta_2}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_2}{3}\rfloor}\big)\Big)\\
&\quad\le 4C_1 \sum_{\Delta_1=3d}^{n-3d-1}\Bigg(n\cdot(\Delta_1+1)^2\cdot\Big(\beta_{\lfloor\frac{\Delta_1}{3}\rfloor}+a_{\lfloor\frac{\Delta_1}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_1}{3}\rfloor}\big)\Big)\\
&\qquad\qquad + \sum_{\Delta_2=3d}^{\Delta_1} n\cdot(\Delta_2+1)\cdot\Big(\beta_{\lfloor\frac{\Delta_2}{3}\rfloor}+a_{\lfloor\frac{\Delta_2}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_2}{3}\rfloor}\big)\Big)\Bigg)\\
&\quad\le 4C_1\Bigg(\bigg(\sum_{\Delta_1=3d}^{n} n\cdot(\Delta_1+1)^2\cdot\Big(\beta_{\lfloor\frac{\Delta_1}{3}\rfloor}+a_{\lfloor\frac{\Delta_1}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_1}{3}\rfloor}\big)\Big)\bigg)\\
&\qquad\qquad + \bigg(\sum_{\Delta_2=3d}^{n} n^2\cdot(\Delta_2+1)\cdot\Big(\beta_{\lfloor\frac{\Delta_2}{3}\rfloor}+a_{\lfloor\frac{\Delta_2}{3}\rfloor}+\phi\big(a_{\lfloor\frac{\Delta_2}{3}\rfloor}\big)\Big)\bigg)\Bigg)\\
&\quad\le 8C_1 n(n+1)\sum_{k=d}^{n}(3k+1)\cdot(\beta_k+a_k+\phi(a_k)),
\end{aligned}$$

since $\Delta_1+1\le n+1$ for $3d\le\Delta_1\le n$. Proceeding in an analogous way in the other cases, by our assumed summability condition (2.16) we get that

$$\sum_{\substack{1\le i_1<j_1\le n\\ 1\le i_2<j_2\le n\\ i_1\neq i_2 \text{ or } j_1\neq j_2}} \mathbb{E}\left(J(X_{i_1},X_{j_1})J(X_{i_2},X_{j_2})\right)\le C_2 n^2,$$

for some constant $C_2>0$. Combining this with (2.23) we directly obtain (2.18). □

2.5 Conditional Distributions

In this section, we give a very brief introduction to regular conditional probabilities stating all results we will make use of in Chapter 5, so that we can refer to them in a proper manner.

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\mathcal{A}\subset\mathcal{F}$ a sub-σ-algebra. A *conditional probability* given $\mathcal{A}$ is defined by

$$\mathbb{P}(B,\mathcal{A})(\omega):=\mathbb{E}(\mathbb{1}_{\{B\}}|\mathcal{A})(\cdot)$$

for each measurable set $B\in\mathcal{F}$ and $\omega\in\Omega$. It would now be desirable if these conditional distributions fulfilled the definition of a probability measure. However, even though countable additivity holds for almost all $\omega\in\Omega$ by definition of the conditional expectation and the monotone convergence theorem, in general it is not directly given. This is due to the fact that the null sets might depend on the sequence of distinct measurable sets $B_i\in\mathcal{F}$, $i\in\mathbb{N}$, and the union of these null sets might cover Ω (cf. Dudley [27, Chapter 10.2]). Therefore, the need for the additional condition of regularity arises: Let $\mathbb{P}_{|\mathcal{A}}$ be the restriction of $\mathbb{P}$ to $\mathcal{A}$. A *regular conditional probability* is a function $\mathbb{P}_{|\mathcal{A}}(\cdot,\cdot) : \Omega \times \mathcal{F} \to [0,1]$ such that for each $B \in \mathcal{F}$, $\mathbb{P}_{|\mathcal{A}}(\cdot, B) := \mathbb{P}(B|\mathcal{A})(\cdot)$ is a conditional probability, and for $\mathbb{P}_{|\mathcal{A}}$-almost all $\omega \in \Omega$, $\mathbb{P}(\cdot|\mathcal{A})(\omega)$ is a probability measure on $\mathcal{F}$ (cf. Dudley [27, Chapter 10.2]).

Based on these regular conditional probabilities, now the general definition of the conditional distribution of a random variable X given a σ-algebra $\mathcal{A}$ can be given.

Definition 2.5.1 (Dudley [27, Chapter 10.2])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(S, \mathcal{S})$ a measurable space. Let $X : \Omega \to S$ be a measurable function. Furthermore, let $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra. A *conditional distribution for X, given $\mathcal{A}$,* is a function $\mathbb{P}_{X|\mathcal{A}}(\cdot,\cdot) : \Omega \times \mathcal{S} \to [0,1]$ such that

(1) for $\mathbb{P}$-almost all $\omega \in \Omega$, $\mathbb{P}_{X|\mathcal{A}}(\omega, \cdot)$ is a probability measure on $\mathcal{S}$,

(2) for each $B \in \mathcal{S}$, $\mathbb{P}_{X|\mathcal{A}}(\cdot, B)$ is $\mathcal{A}$-measurable, and

(3) for each $B \in \mathcal{S}$, it holds $\mathbb{P}_{X|\mathcal{A}}(\cdot, B) = \mathbb{P}(X^{-1}(B)|\mathcal{A})(\cdot)$ almost surely.

Remark 2.5.2. (a) Regarding the existence and uniqueness of such a distribution, the fulfillment of the first condition (regularity) is the crucial point. The necessary requirements for this are stated in Theorem 2.5.3.

(b) The above definition is the most generalized one as it is given in terms of conditioning on a σ-algebra $\mathcal{A}$. However, when conditioning on a random variable $Z : \Omega' \to S'$, the conditional distribution can be defined equivalently as a function

$$\mathbb{P}_{X|Z}(\cdot,\cdot) = \mathbb{P}_{X|\sigma(Z)}(\cdot,\cdot) : S' \times \mathcal{S} \to [0,1]$$

such that

(1a) for $\mathbb{P}$-almost all $z \in S'$, $\mathbb{P}_{X|Z}(z, \cdot) := \mathbb{P}_{X|Z=z}(\cdot)$ is a probability measure on $\mathcal{S}$

in addition to conditions (2) and (3) of Definition 2.5.1 (cf. Dudley [27, Chapter 10.2]). The equivalence holds due to

$$\begin{aligned}\mathbb{P}(X^{-1}(B)|Z)(\omega) &= \mathbb{E}(\mathbb{1}_{\{X^{-1}(B)\}}|Z)(\omega)\\ &= \mathbb{E}(\mathbb{1}_{\{X^{-1}(B)\}}|Z = z) \circ Z(\omega)\\ &= \mathbb{P}(X^{-1}(B)|Z = z) \circ Z(\omega),\end{aligned}$$

for all $\omega \in \Omega'$.

Theorem 2.5.3 (Dudley [27, Theorem 10.2.2])**.** *Let S be a Polish space, $\mathcal{S}$ its σ-algebra of Borel sets, $X : \Omega \to S$ a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$ a sub-σ-algebra. Then, a conditional distribution $\mathbb{P}_{X|\mathcal{A}}(\cdot,\cdot)$ on $\Omega \times \mathcal{S}$ exists. It is unique in the sense that if $\tilde{\mathbb{P}}(\cdot,\cdot)$ also satisfies the definition of $\mathbb{P}_{X|\mathcal{A}}$, then for $\mathbb{P}$-almost all ω, the two laws $\tilde{\mathbb{P}}(\omega,\cdot)$ and $\mathbb{P}_{X|\mathcal{A}}(\omega,\cdot)$ are identical.*

Note that it is possible to construct more general spaces on which regular conditional probabilities do not exist.

On the other hand, note that the real line is a Polish space and that countable products of Polish spaces provided with the product topology are again Polish (cf., e.g., [86, Example 13.15/Theorem 13.16]). Hence, $\mathbb{R}^d$ is Polish for any dimension $d \geq 1$ and therefore, existence and uniqueness of conditional distributions $\mathbb{P}_{X|\mathcal{A}}(\cdot,\cdot)$ on $\Omega \times \mathcal{B}(\mathbb{R}^d)$ is guaranteed.

The last theorem of this chapter states that if a conditional distribution exists, conditional expectations can be written as integrals for each $\omega \in \Omega$.

Theorem 2.5.4 (Dudley [27, Theorem 10.2.5])**.** *Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ be a random variable, $\mathcal{A} \subset \mathcal{F}$ a sub-σ-algebra and $\mathbb{P}_{X|\mathcal{A}}(\cdot,\cdot)$ a conditional distribution on $\Omega \times \mathcal{S}$. Furthermore, let $g : S \to \mathbb{R}^d$ be a measurable function with $\mathbb{E}|g(X)| < \infty$. Then for $\mathbb{P}$-almost all $\omega \in \Omega$, g is integrable with respect to $\mathbb{P}_{X|\mathcal{A}}(\omega,\cdot)$, and*

$$\mathbb{E}(g \circ X|\mathcal{A})(\omega) = \int g(x)\mathbb{P}_{X|\mathcal{A}}(\omega, dx). \tag{2.26}$$

3 Univariate Ordinal Patterns and their Representations

In this chapter we introduce (univariate) ordinal patterns by describing and analyzing different approaches to represent them. Most of them can be found in the literature, though usually one gets the impression that the authors have chosen the respective representations randomly or 'just because others have used them before'. Therefore, we compare the most important representations (plus sub-classes) with regard to their applicability from different angles, namely digital implementation, inverse patterns and ties between values in the data.

This chapter is structured as follows: In the first section, we present the main mathematical concepts. Sections 3.2–3.4 constitute a comparison of the previously defined ordinal pattern representations from different perspectives followed by a short case study in Section 3.5. We conclude this chapter with a short guideline on which occasions which representation should be used, on what we base our choice of representations used in the remaining of this thesis.

This chapter is essentially a revised version of the publication [74], which is a joint work with A. Schnurr.

3.1 Definitions and Mathematical Framework

Let χ be a totally ordered space and $\mathbf{x} = (x_1, ..., x_d) \in \chi^d$ with $d \geq 2$.

Definition 3.1.1. The *(univariate) ordinal pattern* (of length d) of $\mathbf{x}$ is defined as the description of the relation of the elements of $\mathbf{x}$ in terms of their position and rank order.

Considering the vector $(5, 3, 7) \in \mathbb{N}^3$ as an example, its ordinal pattern is fully specified by: *"Of three elements, the third is the largest, while the second is the least"*. Note that this description applies to any vector (x_1, x_2, x_3) satisfying the relation $x_2 < x_1 < x_3$, hence all of them correspond to the same ordinal pattern. For now, we focus on the restriction that $\mathbf{x}$ consists of pairwise distinct elements. Allowing for ties, i.e., equalities between some elements, requires more refined representations, which are the subject of Section 3.4. With this restriction, for $d = 2$ there are only 2 ordinal patterns, namely, the upward pattern $x_1 < x_2$ and the downward pattern $x_1 > x_2$. For $d = 3$, there are already 6 possible patterns, which are depicted in Fig. 1.1.

Now, this specification via text or illustration is quite cumbersome in practice, so the need for other representations arises. Let S_d denote the set consisting of the permutations of $\{1, \ldots, d\}$ and note that $|S_d| = d!$, where $|\cdot|$ denotes the cardinality of a set.

Definition 3.1.2. The *permutation representation* of the ordinal pattern of $\mathbf{x}$ is defined as the permutation $\pi = (\pi_1, ..., \pi_d) \in S_d$ satisfying

$$x_{\pi_1} < ... < x_{\pi_d}. \tag{3.1}$$

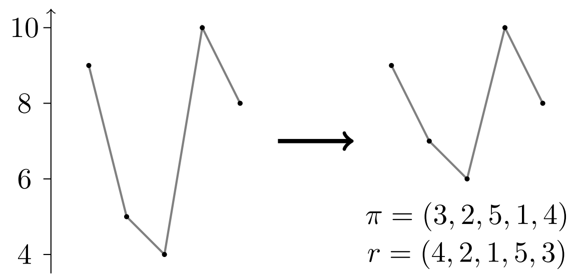


Figure 3.1: Ordinal pattern representations for $\mathbf{x} = (9, 5, 4, 10, 8)$.

This representation is used, e.g., in [13, 87, 90]. The tuple defined above consists of the indices of the respective elements of $\mathbf{x}$ sorted from the least to the largest value. Instead of an increasing order, the permutation representation of ordinal patterns is sometimes defined in a decreasing order, i.e., the pattern π has to satisfy $x_{\pi_1} > ... > x_{\pi_d}$ (see, e.g., [73]). Moreover, even though for our purpose consideration of the length d of an ordinal pattern is more convenient, regarding ordinal time series analysis it is more common to count the number of increments instead of d [17], so, starting in zero instead of one, it is prevalent to consider permutation representations $(\pi_0, \pi_1, ..., \pi_d)$ of vectors $(x_0, x_1, ..., x_d)$ with $d \in \mathbb{N}$ (see, e.g., [10, 17, 44, 71]). This already results in four different possibilities to represent an ordinal pattern which are all closely linked to each other. Let us now come to a contrasting but complementary representation.

Definition 3.1.3. The *rank representation* $r = (r_1, ..., r_d) \in S_d$ of the ordinal pattern of $\mathbf{x}$ is defined by the condition

$$r_j < r_k \iff x_j < x_k \quad \text{for all } j, k \in \{1, ..., d\}. \tag{3.2}$$

This representation is used, e.g., in [7, 11, 87, 89]. Originally, the permutation representation has been most prevalent in the literature. Now it seems that there is a shift in recent publications as the rank representation seems to become more popular. A reason for this may be the intuitiveness of the concept of ranks in general.

For the rank representation, the entries are given by the ranks of the respective values, where 1 denotes the minimum and d the maximum. Again, an inverted definition in terms of ranks is possible such that 1 denotes the maximum, while d denotes the minimum, though it is not very common.

For better illustration of the ideas, consider Fig. 3.1. There the ordinal pattern of the vector $\mathbf{x} = (9, 5, 4, 10, 8)$ is depicted together with the respective permutation and rank representations. Note the dichotomy: While for the rank representation unique ranks are assigned to the indices of $\mathbf{x}$, for the permutation representation of ordinal patterns indices are assigned to ranks. Nevertheless, the definitions are equivalent: Any pattern $\pi = (\pi_1, ..., \pi_d)$ can be determined by a distinct permutation function $\sigma : \mathbb{N} \to \mathbb{N}$ defined by $\sigma(j) = \pi_j$, that is, the condition $x_{\sigma(1)} < ... < x_{\sigma(d)}$ is satisfied. Moreover, its inverse function $\sigma^{-1} : \mathbb{N} \to \mathbb{N}$ satisfies

$$r_j = \sigma^{-1}(j) < \sigma^{-1}(k) = r_k \iff x_j < x_k,$$

which already constitutes the rank representation [14].

Definition 3.1.4. Define the set

$$\mathcal{I}_d = \bigtimes_{k=1}^{d} \{0, 1, ..., d-k\}$$

for $d \geq 2$. The *inversion representation* of the ordinal pattern of $\mathbf{x}$ is defined as the tuple $\eta = (\eta_1, ..., \eta_d) \in \mathcal{I}_d$ consisting of the (right) inversion counts, that is,

$$\eta_j = \sum_{k=j+1}^{d} \mathbb{1}_{\{x_j > x_k\}} \tag{3.3}$$

for $j \in \{1, ..., d\}$.

This representation is used, e.g., in [14]. Note that it holds $\eta_d = 0$ for the rightmost inversion count, which is why most authors omit it in their definitions of the inversion representation. We leave it as it is, so that all representations using tuples have the same length.

Unlike the other two representations using tuples, the inversion representation has no meaningful pictorial interpretation from which one could immediately read the pattern of up and down within the vector. Originating in discrete mathematics, this representation is typically used with regard to digital implementation of ordinal patterns as it is very convenient there with regard to numerical encoding of ordinal patterns (see Section 3.2), though, instead of the (right) inversion counts, sometimes variations in terms of left inversion counts or non-inversion counts are used as, e.g., in [42, 44].

The question now arises whether it is possible to find explicit maps which allow to switch from one representation to another one. For the permutation and rank representations, the question is already settled. Noting that the inversion counts η_j, $j \in \{1, ..., d\}$, can be obtained via

$$\eta_j = \sum_{k=j+1}^{d} \mathbb{1}_{\{r_j > r_k\}}, \tag{3.4}$$

where r_k denotes the rank of x_k, it is sufficient to show that the permutation representation π can be obtained by the inversion representation η. For this, let $\pi^{(1)}, \pi^{(2)}, ..., \pi^{(d)} = \pi$ be a sequence of permutations of $\{d\}, \{d, d-1\}, ..., \{d, d-1, ..., 1\}$, respectively. Since $\{d\}$ consists of one element, $\pi^{(1)} = (d)$ denotes the trivial permutation. Suppose $\pi^{(l-1)} = (\rho_1, \rho_2, ..., \rho_{l-1})$ is given for an $l \in \{2, ..., d\}$. Then $\pi^{(l)}$ can be obtained from $\pi^{(l-1)}$ by inserting $d+1-l$ into $(\rho_1, ..., \rho_{l-1})$:

1. If $\eta_{d+1-l} = 0$, then $d+1-l$ is inserted to the left of ρ_1, such that $\pi^{(l)} = (d+1-l, \rho_1, ..., \rho_{l-1})$.

2. Otherwise it is inserted to the right of $\rho_{\eta_{d+1-l}}$.

For a similar procedure regarding a definition of the inversion representation in terms of different (non-)inversion counts, see, e.g., Keller et al. [44]. In fact, this procedure is based on the authors' ideas.

Example 3.1.5. Consider the representation $\eta = (3, 1, 0, 1, 0)$. Since $d = 5$, it holds $\pi^{(1)} = (5)$. Due to $\eta_4 = \eta_{d+1-2} = 1$, '4' has to be inserted to the right of $\rho_{\eta_3} = \rho_1 = 5$ such that $\pi^{(2)} = (5, 4)$. It holds $\eta_3 = 0$, which yields $\pi^{(3)} = (3, 5, 4)$. Then, because of $\eta_2 = 1$ it follows that ‘2’ has to be inserted to the right of $\rho_1 = 3$, so $\pi^{(4)} = (3, 2, 5, 4)$. Finally, it holds $\eta_1 = 3$, hence ‘1’ has to be inserted to the right of $\rho_4 = 4$ such that $\pi = \pi^{(5)} = (3, 2, 5, 4, 1)$. Note that this is precisely the permutation representation which has been considered in Fig. 3.1, which complements the representations mentioned so far for $\mathbf{x} = (9, 5, 4, 10, 8)$.

3.2 Digital Implementation

With regard to digital implementation, in order to derive fast algorithms for the determination of ordinal patterns stemming from a time series, it is necessary to keep computational and memory costs (in the sense of storing the obtained ordinal patterns) as low as possible. Therefore, the naive solution of storing ordinal patterns as d-dimensional arrays is disadvantageous, since, among other reasons, testing a pair of ordinal patterns for equality would require up to d comparisons, and storing arrays in general results in a way larger memory footprint compared to simple integers [14]. Hence, an ordinal pattern representation in terms of a single (non-negative) integer is preferable.

Definition 3.2.1. For $d \geq 2$, let unique non-negative integers $n \in \mathbb{N}_0$ be assigned to the ordinal patterns of $\mathbf{x}$ according to some bijective map. We call this a *number representation* of ordinal patterns.

Note that the map is not further specified, i.e., any bijective map can generate a number representation of ordinal patterns, though for digital implementation it is advantageous if the number representation can be directly computed from any other ordinal pattern representation mentioned before instead of implementing some sort of lookup table [14]. In this context, approaches/solutions using the inversion representation are already available: Let $\eta = (\eta_1, ..., \eta_d) \in \mathcal{I}_d$ be the inversion representation of an ordinal pattern. Keller et al. [44] proposed a numerical encoding defined by the relation

$$n_{KSE} = \sum_{j=1}^{d} \eta_j \cdot \frac{d!}{(d-j+1)!} \in \{0, 1, ..., d! - 1\}, \tag{3.5}$$

where ‘KSE’ refers to the name of the authors. Another approach based on the Lehmer code is proposed by Berger et al. [14]. There, ordinal patterns are enumerated by

$$n_{LC} = \sum_{j=1}^{d} \eta_j \cdot (d-j)! \in \{0, 1, ..., d! - 1\}. \tag{3.6}$$

Note that both maps (3.5) and (3.6) are bijective [14, 44]. The second encoding preserves the lexicographic sorting order with regard to a deviation from an increasing pattern $(1, 2, \ldots, d)$ [14]. Note that with regard to the inversion representation this is equivalent to $(\eta_1, ..., \eta_d) \preccurlyeq (\eta_1^*, ...\eta_d^*)$ if and only if

$$(\eta_1, ..., \eta_d) = (\eta_1^*, ..., \eta_d^*) \quad \text{or} \quad \eta_1 = \eta_1^*, ..., \eta_{k-1} = \eta_{k-1}^*, \eta_k \leq \eta_k^*$$

for some $k \in \mathbb{N}$ with $k \leq d - 1$ [44]. For a better understanding consider Table 3.1, where both numerical encodings for all possible ordinal patterns of length $d = 3$ are listed. Note that even though [44] claimed their numerical encoding to be lexicographic, here it is not as, e.g., Table 3.1 shows. Rather it follows a different order. This is due to the different

Table 3.1: Numerical encodings for ordinal patterns of length $d = 3$.

Rank Representation	Inversion Representation	n_{KSE}	n_{LC}
(1, 2, 3)	(0, 0, 0)	0	0
(1, 3, 2)	(0, 1, 0)	3	1
(2, 1, 3)	(1, 0, 0)	1	2
(2, 3, 1)	(1, 1, 0)	4	3
(3, 1, 2)	(2, 0, 0)	2	4
(3, 2, 1)	(2, 1, 0)	5	5

definitions of the vector $\mathbf{x}$ under consideration: The vector $\mathbf{x}$ considered by [44] is of the form $\mathbf{x} = (x_{-1}, x_{-2}, ..., x_{-d}) = (x^*_d, x^*_{d-1}, ..., x^*_1)$. Therefore, one could argue that the inversion representation the authors use is with respect to non-inversion counts rather than (right) inversion counts. In fact, consideration of variants of inversion counts, as, e.g., left inversion counts or non-inversion counts, does not result in a different type of encoding, but changes the produced enumeration order.

As the encodings presented above only vary with regard to the weights, at first glance one may say that mentioning both encodings is redundant. However, the approach based on the Lehmer code leads to a remarkably simple algorithm for extracting and storing ordinal patterns in computer memory, which results in a reduction of computational complexity [14].

Algorithms for extracting ordinal patterns from time series data using the encodings presented above have been proposed by Berger et al. [14], Keller et al. [44] and Unakafova and Keller [85]. For a thorough discussion of these algorithms in terms of strengths and weaknesses we especially refer to [14].

In summary, the inversion representation is very advantageous with regard to digital implementation of ordinal patterns, since it allows for a direct computation of number representations. What about the permutation and rank representations?

The rank representation is disadvantageous for digital implementation, since the identification of the rank of an entry needs d comparisons, which sums up to d^2 comparisons for the entire representation. In contrast, the identification of the j-th inversion number requires $d - j$ comparisons, which results in a total of $\sum_{j=1}^{d} d - j = (d^2 - d)/2$. This can make a major difference regarding the computational time with respect to extracting ordinal patterns from a large data set. Furthermore, (at least to our knowledge) it is not possible to find weights $w_1, ..., w_d \in \mathbb{N}$ such that the encoding $\mathrm{en} : S_d \to \{0, 1, ..., d! - 1\}$ defined by $\sum_{j=1}^{d} r_j \cdot w_j$ constitutes a bijective map. An alternative (injective) encoding, e.g., may be given by $\mathrm{en}' : S_d \to \mathbb{N}$ with

$$\mathrm{en}'(r) = \sum_{j=1}^{d} r_j \cdot d^{j-1}. \tag{3.7}$$

Even though this encoding is still very natural, the disadvantages in terms of computational time remain. Furthermore, noting that the image set is different, i.e., it contains larger values than $d! - 1$ in particular, and recalling that large integers demand more memory space if compared to smaller integers, the enumeration with regard to the inversion representation is still more advantageous.

The permutation representation is unsuitable, too, since even the extraction of the ordinal pattern of a vector makes the additional step of rank identification necessary.

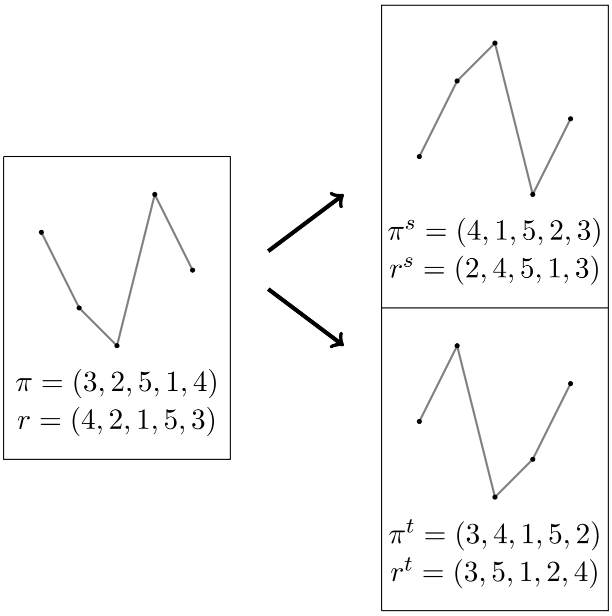


Figure 3.2: Ordinal patterns of $\mathbf{x} = (9, 5, 4, 10, 8)$ (left) inversed in space (top right) and time (bottom right).

3.3 Inverse Ordinal Patterns

For a vector $\mathbf{x} = (x_1, ..., x_d)$ there are two possible ordinal inversions, namely an inversion in time obtained by considering the entries of $\mathbf{x}$ in a reversed order, i.e. $(x_d, ..., x_1)$, and an inversion in space given by the reflected vector $(-x_1, ..., -x_d)$. Figuratively speaking, an inversion in time means reflecting on a vertical line, while a reflection on a horizontal line yields an inversion in space. These result in altered ordinal patterns and hence, they have different effects on the proposed representations, which we want to discuss in this section.

The occurrence of a certain pattern in another data set along with its inversion in space can be interpreted as antimonotone behavior, that is, if a certain value at one point in time is high, then it is likely that the corresponding value in the other data set is low. This is a phenomenon which can be observed, e.g., in finance and has been analyzed for the S&P 500 and the corresponding volatility index VIX in Schnurr [70] and Schnurr and Dehling [71].

First, let us consider representations of ordinal patterns inversed in space. For such an inversion, the ordinal pattern is reflected on a horizontal line such that the largest value becomes the least, the second largest becomes the second least, and so on. For the permutation representation $\pi = (\pi_1, ..., \pi_d)$ this means that it has to be read from right to left, that is, the permutation representation of the space-inversed ordinal pattern is given by $\pi^s = (\pi_d, ..., \pi_1)$ (or $\pi^s_j = \pi_{d+1-j}$ for all $j \in \{1, ..., d\}$). In contrast, the space-inversed rank representation $r^s = (r^s_1, ..., r^s_d)$ results from an inversion of the type $r^s_j = d + 1 - r_j$, $j \in \{1, ..., d\}$. For time-inversed ordinal patterns, the entries of $\mathbf{x}$ are considered from right to left instead of from left to right, so the time-inversed rank representation is given by the original rank representation read from right to left, i.e., $r^t = (r_d, ..., r_1)$, while the time-inversed permutation representation is obtained via $\pi^t = d + 1 - \pi_j$ for $j \in \{1, ..., d\}$. Note that the rank representations follow the respective transformations of the vector $\mathbf{x}$ under consideration, while the 'opposite' transformation (in terms of horizontal and vertical) yields the respective permutation representations.

Finally, let us consider inversion representations of ordinal patterns inversed in space and time. From (3.4), for the inversion counts η_j^s of the ordinal pattern inversed in space it follows

$$\eta_j^s = \sum_{k=j+1}^{d} \mathbb{1}_{\{r_j^s > r_k^s\}} = \sum_{k=j+1}^{d} \mathbb{1}_{\{d+1-r_j > d+1-r_k\}} = \sum_{k=j+1}^{d} \mathbb{1}_{\{r_j < r_k\}} = d - j - \eta_j$$

for $j \in \{1, ..., d\}$, since there are $d - j$ values to compare with r_j and we already know that η_j of these comparisons constitute inversions. Obviously, this can be easily computed from the inversion counts. On the other hand, for ordinal patterns inversed in time it holds

$$\eta_j^t = \sum_{k=j+1}^{d} \mathbb{1}_{\{r_j^t > r_k^t\}} = \sum_{k=j+1}^{d} \mathbb{1}_{\{r_{d+1-j} > r_{d+1-k}\}} = \sum_{l=1}^{d+1-j-1} \mathbb{1}_{\{r_{d+1-j} > r_l\}}$$

for $j \in \{1, ..., d\}$. These constitute the left non-inversion counts, which we cannot deduce from η_j in general.

3.4 Ties

Up to this point, we assumed that ties are not present in $\mathbf{x}$. This is an assumption often made in the literature. With regard to time series models, this matches the case that the probability of coincident values equals zero. With regard to data, however, one might still encounter ties. In practice, three approaches are common [72] (recall Section 1.1.3): In the first approach, the respective data points are skipped, i.e., the vectors containing ties are omitted. This is the approach used, e.g., in [7, 13]. In this case one might lose a lot of information, especially considering categorical data sets with a small number of categories. The second approach is randomization, e.g., by adding a small noise to the data in order to avoid ties, which has been done, e.g., by [10, 11]. This has the drawback of possibly underestimating co-movement between data sets or disregarding constant patterns in one data set.

The last of those approaches is an alteration of the respective ordinal pattern representations using tuples in terms of adding a supplementary condition, that is, then the permutation representation of an ordinal pattern of $\mathbf{x}$ is defined as the permutation $\pi = (\pi_1, ..., \pi_d)$ satisfying

$$x_{\pi_1} \leq ... \leq x_{\pi_d} \quad \text{and} \quad \pi_{j-1} < \pi_j \text{ if } x_{\pi_{j-1}} = x_{\pi_j} \tag{3.8}$$

for $j \in \{2, ..., d\}$. This representation is used, e.g., in [3, 17, 87], and it is equivalent to the adjusted rank representation $r = (r_1, ..., r_d)$ defined by

$$r_j < r_k \iff x_j < x_k \quad \text{or} \quad (x_j = x_k \text{ and } j < k) \tag{3.9}$$

by the same argument as in the case of distinct values considered in Section 3.1, and which in turn is used, e.g., in [87, 90]. This means in particular that both representations are used in [87]. Note the additional conditions in comparison to (3.1) and (3.2).

Since the increasing order is retained in case of ties, the inversion representation does not need to be adjusted further, and equivalence of all three representations using tuples is guaranteed. However, this approach maps the vectors $(1, 1, 1)$ and $(1, 10, 100)$ onto the same pattern $\pi = r = (1, 2, 3)$, thus, they are considered to exhibit the same up and down

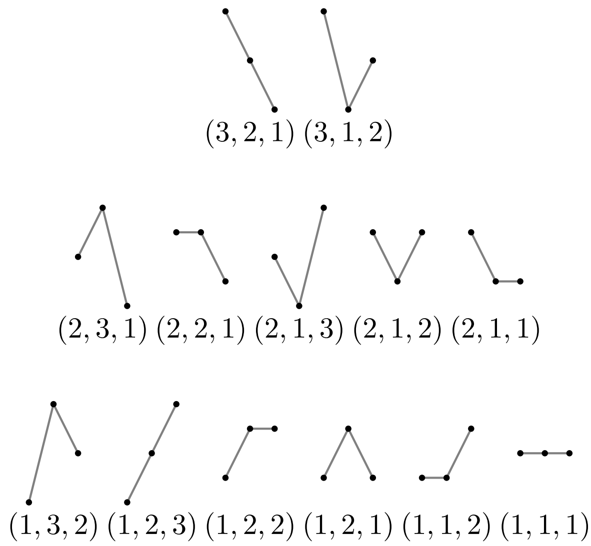


Figure 3.3: Generalized rank representations for ordinal patterns of length $d = 3$.

movement, which clearly is not the case here. Hence, with this approach valuable information is possibly lost. Due to this reason, Schnurr and Fischer [72] propose so-called *generalized ordinal patterns* explicitly allowing for ties by referring to a larger set of possible patterns. The authors propose the following representation, which we will simply refer to as the generalized rank representation:

Definition 3.4.1 (Schnurr and Fischer [72])**.** Suppose that the values $(y_1, ..., y_m)$, which are already ordered by the condition $y_1 < y_2 < ... < y_m$, are attained in the vector $\mathbf{x} = (x_1, ..., x_d)$. There, $m \in \{1, ..., d\}$ is the number of different values. The *generalized rank representation* of the ordinal pattern of $\mathbf{x}$ is defined as the vector $\psi = (\psi_1, ..., \psi_d) \in \mathbb{N}^d$ satisfying

$$\psi_j = k \iff x_j = y_k.$$

By this definition, the vector $(1, 5, 4, 3)$ yields the generalized rank representation $(1, 4, 3, 2)$, which coincides with the rank representation as defined in Def. 3.1.3, while the vector $(1, 1, 4, 3)$ has the generalized rank representation $(1, 1, 3, 2)$ in contrast to $(1, 2, 4, 3)$. Hence, this representation is in fact a generalization of the rank representation with respect to ties. All generalized ordinal patterns of length $d = 3$ with their respective generalized rank representations are depicted in Fig. 3.3 (in co-lexicographic order). Note that there are already 13 generalized ordinal patterns of length $d = 3$ (and 3 for $d = 2$). Denoting the set of all generalized rank representation patterns by T_d, the cardinal numbers $|T_d|, d \geq 2$, are given by the ordered Bell number of order d.

Schnurr and Fischer [72] applied generalized rank representations to hydrological data sets consisting of five flood classes (plus ‘absence of flood’) in terms of measuring the association within a data set by ordinal pattern dependence. As it is a categorical data set with only six categories, the occurrence of many ties is expected, so the authors compare the proposed generalized ordinal patterns to two classical approaches, namely randomization and the altered definition of sorting from beneath by the first appearance (see (3.8) and (3.9)). Overall, they demonstrate that the classical approaches tend to underestimate the dependence present in the data due to the changes of pattern structure, while their proposed approach of generalized patterns overcomes this. Nevertheless, a new drawback arises: Depending on the length of the ordinal pattern, a lot more patterns need to be considered, which can possibly result, e.g., in a greater computational cost with regard to digital implementation. Therefore, the recommendation is to use this approach when dealing with categorical time series, as it is especially designed for them, while the classical approaches are more advantageous for time series for which the probability of coincident values is very small.

The question now arises to what extent it is possible to find generalizations of the permutation and inversion representation, respectively, which are equivalent to the generalized rank representation as defined in Def. 3.4.1. As a matter of fact, there is a natural way in terms of preimages (in the first case), though notationally disadvantageous: The idea of permutation representations is to sort the indices of the entries of $\mathbf{x}$ according to their ordinal pattern. However, in case of ties, the indices cannot be sorted in a unique way nor can ties be read directly without further information from the arrangement of the indices in a tuple. This changes when we consider a tuple consisting of sets. Let

$$\tilde{\sigma} : \{1, ..., d\} \to \{1, ..., m\}$$

be the map that assigns the ranks to the respective indices, where m denotes the previously determined number of different values in $\mathbf{x}$. Clearly, $\tilde{\sigma}$ is surjective. Hence, for fixed $l \in \{1, ..., m\}$, the preimage $\tilde{\sigma}^{-1}(l) = \{\psi_1, ..., \psi_{l_n}\}$ is defined, and a generalized permutation representation can be given by

$$(\tilde{\sigma}^{-1}(1), ..., \tilde{\sigma}^{-1}(m)).$$

Consider the following illustrative example: For $\mathbf{x} = (4, 4, 6)$, the generalized rank representation is given by $\psi = (1, 1, 2)$, so the first two entries obtain rank 1, while the last entry has rank 2. Then, the generalized permutation representation as proposed above is given by $(\{1, 2\}, \{3\})$. This has the advantage that $\{1, 2\} = \{2, 1\}$, so no order is suggested for ties. Nevertheless, the notation is a bit cumbersome compared to sorting the indices themselves as it is done in the classical approach. Obviously, the generalized permutation representation and generalized rank representation are equivalent.

In case of ties, an inversion representation does not make sense as it cannot be defined in a unique way: Considering, e.g., the vectors $(4, 4, 5)$ and $(4, 5, 6)$, under the definition in (3.3) both are mapped to the same inversion representation $(0, 0, 0)$. The naive idea of an adjustment of (3.3) in terms of $\geq$, that is, $\eta_j = \sum_{k=j+1}^{d} \mathbb{1}_{\{x_j \geq x_k\}}$, does not solve the problem, since then $(4, 4, 5)$ and $(4, 3, 5)$ are both mapped to $(1, 0, 0)$. The same problem arises when considering other variants of inversion counts. As a consequence, a number representation consisting of $d!$ consecutive integers can be directly computed neither from a generalized inversion representation (as it does not exist) nor the generalized rank representation (due to the same reasons as for the classical rank representation). However, an enumeration according to the encoding $\mathrm{en}'(r) = \sum_{j=1}^{d} \psi_j \cdot d^{j-1}$ (see Section 3.2) is advantageous with regard to digital handling of generalized ordinal patterns so that working with vectors can be avoided. An even more advantageous method is considering $\sum_{j=1}^{d} (\psi_j - 1) \cdot d^{j-1}$ in order to keep the values of the image set as low as possible. Note that the obtained number representations do not consist of consecutive integers here. (Compare with the R-package "`ordinalpattern`", where the function "countingpatterns" has been added in Version 0.2.5, which computes the empirical (generalized) ordinal pattern distribution according to this approach.)

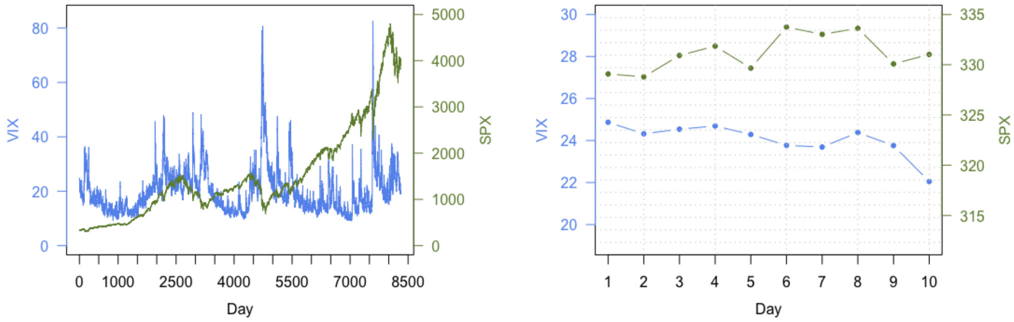


Figure 3.4: 'Open prices' of VIX and SPX, respectively, in the time period 01/02/1990 to 31/01/2023 corresponding to $n = 8313$ data points (left). First 10 data points corresponding to the time period 01/02/1990 to 14/02/1990 (right).

3.5 Case Study

In what follows, we consider the S&P 500 (SPX) and the corresponding volatility index (VIX) as a real-world example in order to illustrate the use of the aforementioned ordinal pattern representations as well as their performance with regard to digital efficiency. We performed our analysis in GNU R (Version 4.4.0) on a MacBook Pro (Apple M1).

We have analyzed daily data in the time period 01/02/1990 to 31/01/2023 resulting in $n = 8313$ data points, and which is available as open source historical data on finance.yahoo.com. We have restricted ourselves to the 'open prices' and, if not mentioned otherwise, we have considered ordinal patterns of length $d = 3$.

The first 10 data points are illustrated at the r.h.s. of Fig. 3.4 from which the respective ordinal patterns can be mapped directly to the rank or permutation representation. If one is interested in the probability of certain patterns, e.g., the upward movement $\pi = r = (1, 2, 3)$ as it might be connected to economic growth (in the case of SPX), determining the relative frequencies for estimation is crucial. Here, we make use of the number representation n_{LC} (see Eq. (3.6)), which utilizes indirectly the inversion representation $i = (i_1, .., i_d)$, and use the 'Plain Algorithm' as proposed by Berger et al. [14]. In summary, this yields the relative frequencies stated in Table 3.2. Using the R-package "microbenchmark", we have obtained the results in far less than one second, respectively, that is, out of 1000 repetitions, an average of about 32 milliseconds was needed for each of the two data sets. Note that the data exhibits a negligible number of ties, hence, here we have used the altered definition of ordinal patterns which retains the increasing order in case of ties (see Eq. (3.8), (3.9)), so the inversion representation does not need to be adjusted further.

In addition, we have investigated the computational cost with regard to ordinal pattern

Table 3.2: Relative frequencies of ordinal patterns n_{LC} of length $d = 3$ with regard to VIX and SPX rounded to the third digit.

n_{LC}	0	1	2	3	4	5
VIX	0.210	0.135	0.137	0.134	0.131	0.253
SPX	0.295	0.130	0.127	0.118	0.121	0.209

Table 3.3: Ordinal pattern dependence (OPD) for different lengths d with regard to VIX and SPX rounded to the third digit.

d	2	3	4	5	6	7
OPD	-0.365	-0.251	-0.146	-0.076	-0.036	-0.017

dependence, which can be conveniently computed via the function "patterndependence" from the R-package "ordinalpattern". This function uses an approach for assigning number representations to the respective patterns which is not based on the inversion representation, but works along the lines of Eq. (3.7). Initially, an additional randomization was required in case of ties, so slightly different results can be obtained in each run. (Note that this was the case when the paper on which this chapter is based on was published.) Since Version 0.2.5, however, there is the possibility to choose whether the function uses randomization or the method in favor of increasing patterns mentioned before. We have summarized our results with regard to the latter for different pattern lengths d in Table 3.3. All in all, the time needed for the computation of ordinal pattern dependence with $d = 3$ is about one quarter of the time needed with regard to the pattern probabilities computed before (8.5 milliseconds on average out of 1000 repetitions), even though these probabilities must also be determined for ordinal pattern dependence. Moreover, the computation of ordinal pattern dependence with $d = 7$ takes still far less than one second, but twice the time with regard to the aforementioned computations (about 62 milliseconds on average). This is probably the case, since parts of the computation in "patterndependence" are outsourced to C shortening the computation time.

3.6 Interim Conclusion

In probability theory and statistics in the context of ordinal pattern analysis, most of the time the distribution of ordinal patterns is of interest. In order to determine or estimate probabilities of certain ordinal patterns, the pattern representation itself is not relevant, since the choice of the representation usually does not influence the obtained results directly. However, it is still advantageous to opt for a representation for which the ordinal pattern can be read directly, namely the permutation or rank representation. Here, we prefer the second one, since the concept of ranks is more intuitive and the ups and downs of the ordinal pattern can be read directly. Furthermore, a vector like $\mathbf{x} = (2, 3, 1, 4)$ is mapped to $r = (2, 3, 1, 4)$ (instead of $\pi = (3, 1, 2, 4)$) which is very natural.

When considering inverse patterns, permutation and rank representation are both reasonable, since they behave in a similar, but 'opposite' way regarding the transformations needed for the respective inversions (see Section 3.3). On the other hand, the inversion representation, even though counterintuitive, is advantageous with respect to ordinal patterns inversed in space due to its closed form in terms of the original inversion counts. But with regard to inversions in time, it is inconvenient, because then one has to compute non-inversion counts which cannot be obtained from the (right) inversion counts directly. However, the inversion representation is very practical when it comes to digital implementation, since it leads to remarkably simple algorithms for extracting and storing ordinal patterns in computer memory, that keep computational and memory costs very low.

In the context of ties being present in the data, the use of the three classical approaches

Table 3.4: Ordinal patterns of length $d = 2$ in (i), $d = 3$ in (ii), and $d = 4$ in (iii). Unless explicitly indicated otherwise, from now on classical ordinal patterns are referred to by the letter π.

(i)

$d=2$	
k	$\pi^{(k)}$
1	(1, 2)
2	(2, 1)

(ii)

$d=3$			
k	$\pi^{(k)}$	k	$\pi^{(k)}$
1	(1, 2, 3)	2	(1, 3, 2)
3	(2, 1, 3)	4	(2, 3, 1)
5	(3, 1, 2)	6	(3, 2, 1)

(iii)

$d=4$					
k	$\pi^{(k)}$	k	$\pi^{(k)}$	k	$\pi^{(k)}$
1	(1, 2, 3, 4)	2	(1, 2, 4, 3)	3	(1, 3, 2, 4)
4	(1, 3, 4, 2)	5	(1, 4, 2, 3)	6	(1, 4, 3, 2)
	...		...		...
22	(4, 2, 3, 1)	23	(4, 3, 1, 2)	24	(4, 3, 2, 1)

using tuples is recommended for data sets for which many ties are not expected, e.g., data stemming from real-valued time series. Then, even for the third approach, where the definitions of the permutation and rank representation are altered, respectively, the previously discussed advantages and disadvantages remain valid. However, if, e.g., a categorical data set with a small number of categories is considered, many ties are to be expected. Hence, the classical approaches would lead to a distortion of the underlying distribution and/or loss of valuable information. Therefore, in this case the generalized rank representation is very beneficial, even though the number of possible patterns increases even more rapid for length d than the set of permutations considered in the case of vectors containing pairwise distinct elements. The equivalent generalized permutation representation, which we proposed, is notationally more cumbersome, hence, we still recommend the use of the generalized rank representation.

All in all, if not stated otherwise, in the remaining we refer to classical ordinal patterns by rank representations denoted by π (instead of r) in order to keep the notation simple and intuitive. We sort them according to the number representations $n_{LC} + 1 \in \{1, \ldots, d!\}$ as proposed by Berger et al. [14], since these correspond to a lexicographical sorting order of the patterns. A brief overview on the actual numberings in combination with the respective rank representations for $d \in \{2, 3, 4\}$ is shown in Table 3.4.

4 Limit Theorems for the Symbolic Correlation Integral and Rényi-2 Entropy for Short-range Dependent Time Series

Originally, ordinal patterns have been introduced by Bandt and Pompe [10] as a tool to define permutation entropy which is a measure for quantifying complexity or uncertainty within a given time series (or data stemming from it). As discussed before, ordinal patterns have many desirable properties like invariance under monotone transformations, robustness with respect to small noise and simplicity in computation. These properties are directly transferred to ordinal pattern based measures, e.g., permutation entropy.

Since permutation entropy is defined as the Shannon entropy of the ordinal pattern distribution (see Definition 2.2.2), it is a natural idea to also consider other variants of complexity measures based on ordinal patterns (for a discussion of different variants of permutation entropy, see Keller et al. [46]). Both the Rényi and the Tsallis entropy converge to the Shannon entropy for $q \to 1$, so it is especially natural to generalize permutation entropy to variants of these (cf. Liang et al. [49], Zunino et al. [92])

Here, we will particularly consider a variant based on Rényi-2 entropy. Even though both variants seem to behave not too differently from the practical perspective (see Keller et al. [46] for a discussion from the viewpoint of their application in EEG analysis), the Rényi-2 entropy has some interesting properties and advantages. First of all, it is strongly related to the symbolic correlation integral recently proposed by Caballero-Pintado et al. [21], which is inspired by the widely used classical correlation integral defined by Grassberger and Procaccia [31] and can be interpreted as the degree of recurrence of ordinal patterns in the time series (cf. [46, p. 8], [21, p. 537]). In fact, from a practical viewpoint the symbolic correlation integral and Rényi-2 entropy can be used more or less interchangeably as we will see later. Another advantage of the Rényi-2 entropy or symbolic correlation integral is a strong relation to U-statistics, which will prove useful in the context of establishing limit theorems.

In this chapter, our goal is to derive the limit distribution of the symbolic correlation integral (and hence also the Rényi-2 entropy) for a broad class of short-range dependent processes, namely 1-approximating functionals. Therefore, we complement the results by Caballero-Pintado et al. [21] who only considered the i.i.d. case. Our contributions will prove to be useful in a variety of classification tasks, that is, they allow us to distinguish time series or data stemming from time series based on the degree of complexity present in each one of them. For example, to a certain extent we are able to distinguish, e.g., ARMA-models that differ only in the choice of their parameters. Furthermore, our approach also allows for testing whether two time series follow the same underlying model in the sense of a distinction between, e.g., an AR- and an MA-model. Possible practical applications include, e.g., the distinction between healthy patients and patients with specific diseases for which the collected data exhibits a different degree of complexity if compared to the data collected from healthy patients. Well-known examples for such a diseases are epilepsy as well as different kinds of heart diseases (see Amigó et al. [3] and the references mentioned therein). Classification of sleep states might be another possible application.

This chapter is organized as follows: First we give a brief introduction into the main quantity of interest, namely the symbolic correlation integral (Section 4.1). Then we derive its limit distribution for the class of 1-approximating functionals (Section 4.2) and estimate the limit variance (Section 4.3), followed by a short simulation study in Section 4.4 to support our theoretical findings. An interim conclusion rounds off this chapter.

The development of hypothesis tests as well as their practical applications are not addressed in this work. Instead this can be found in the joint work [75] with A. Schnurr and M. R. Marín, on which this chapter (as well as the introduction above) is based on in most parts. In the course of this chapter, we will clearly point out the results to which we have not contributed significantly, but which we nevertheless need for the rigor of our mathematical conclusions, by referring to the joint work accordingly. The included simulations (with the exception of Proposition 4.4.1) were conducted independently of [75] as a result of our own work.

4.1 Rényi-2 Permutation Entropy and the Symbolic Correlation Integral

In this chapter, we consider the stationary one-dimensional time series $(X_t)_{t\in\mathbb{Z}}$ whose finite dimensional distributions we assume to be continuous. Since we are interested in the relationship between d consecutive data points, it is convenient to consider the modified time series $(\overline{X}_t)_{t\in\mathbb{Z}}$ defined by the components

$$\overline{X}_t := (X_t, ..., X_{t+d-1})$$

obtained by a sliding window approach. Note that if the time series $(X_t)_{t\in\mathbb{Z}}$ is stationary, then $(\overline{X}_t)_{t\in\mathbb{Z}}$ is also stationary, since considering d consecutive data points is only a simple functional used on the original time series. Let F denote the cumulative distribution function

$$F(y) = F_{\overline{X}_0}(y) = F_{\overline{X}_t}(y) = \mathbb{P}(\overline{X}_t \le y)$$

for $y \in \mathbb{R}^d$. Given some $t \in \mathbb{Z}$ and recalling that $\Pi : \mathbb{R}^d \to S_d$ denotes the function which assigns each d-dimensional vector its ordinal pattern, the main quantity under consideration in this chapter is

$$\begin{aligned}\sum_{\pi\in S_d} \mathbb{P}(\mathbb{1}_{\{\Pi(\overline{X}_t)=\pi\}})^2 &= \sum_{\pi\in S_d} \left(\int_{\mathbb{R}^d} \mathbb{1}_{\{\Pi(x)=\pi\}}\, dF(x)\right)^2 \\ &= \sum_{\pi\in S_d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{1}_{\{\Pi(x)=\pi\}}\mathbb{1}_{\{\Pi(y)=\pi\}}\, dF(x)dF(y) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{1}_{\{\Pi(x)=\Pi(y)\}}\, dF(x)dF(y) =: S^d, \end{aligned} \tag{4.1}$$

which does not depend on $t \in \mathbb{Z}$ by stationarity. Note that we have applied Fubini's theorem in the second equation. The r.h.s. of (4.1) is the symbolic correlation integral (SCI) as recently proposed by Caballero-Pintado et al. [21]. As the l.h.s. shows, it is nothing more than the exponential function applied to the sign-reversed Rényi-2 entropy

$$R_2 = R_2((p_\pi)_{\pi\in S_d}) = -\log \sum_{\pi\in S_d} p_\pi^2$$

of the ordinal pattern distribution $(p_\pi)_{\pi\in S_d} = \big(\mathbb{P}(\Pi(\overline{X}_t) = \pi)\big)_{\pi\in S_d}$ (see Chapter 2.2).

The definition of the SCI is highly influenced by the widespread and well-known classical correlation integral by Grassberger and Procaccia [31], which is given by the probability of two arbitrary points on the orbit of the state space that are within a distance of ε for some $\varepsilon > 0$ fixed beforehand [21]. In other words, the correlation integral depends highly on the a priori selection of ε. The SCI, on the other hand, avoids this ε-dependence and, as Caballero-Pintado et al. [21] have shown, it can be used to construct a new test for serial dependence and causality, that is consistent, nuisance-parameter-free, and computationally efficient.

Eq. (4.1) shows that the SCI can be easily computed if the ordinal pattern probabilities are known. For example, in case of a uniform distribution, which is, e.g., the case for i.i.d. series, it holds

$$S^d = \sum_{\pi\in S_d} p_\pi^2 = d! \cdot \left(\frac{1}{d!}\right)^2 = \frac{1}{d!}.$$

For most of the time series models, nothing is known yet about the obtained ordinal pattern distributions. Exceptions are Gaussian as well as ARMA-processes, that is at least under some additional restrictions. The corresponding theorems can be found in Sections 5 and 6 of Bandt and Shiha [11].

Note that the SCI attains its minimum in case of uniformly distributed ordinal patterns, that is $1/d!$, and its maximum of 1 in case of a one-point distribution. Recall in this regard Remark 2.2.4. For the Rényi-2 permutation entropy $R_2 = -\log(S^d)$ maximum and minimum are reversed.

We consider the U-statistic

$$S_n^d := \frac{2}{n(n-1)} \sum_{1\le j<k\le n} \mathbb{1}_{\{\Pi(\overline{X}_j)=\Pi(\overline{X}_k)\}}$$

as an estimator of S^d. Note the connection to the recurrence of ordinal patterns by definition. By now, the limit results for this statistic are limited to the i.i.d. case [21], which is a serious drawback, since several applications one has in mind exhibit serial dependence. We close this gap in the following.

4.2 Limit Theorems for the Symbolic Correlation Integral

The key idea for giving more general statements for a broad class of short-range dependent processes, namely 1-approximating functionals (see Chapter 2.4), is to use the theorems of Borovkova et al. [18]. In order to make use of them, amongst others we have to show that our kernel

$$h(x,y) := \mathbb{1}_{\{\Pi(x)=\Pi(y)\}}$$

satisfies the technical condition of p-continuity (Def. 2.4.15) with regard to our time series $(\overline{X}_t)_{t\in\mathbb{Z}}$.

Proposition 4.2.1 (Schnurr et al. [75, Proposition 4.2])**.** *Let $p \ge 1$. The kernel $h : (x,y) \mapsto \mathbb{1}_{\{\Pi(x)=\Pi(y)\}}$ is p-continuous with respect to $(\overline{X}_t)_{t\in\mathbb{Z}}$.*

For the proof we refer the reader to [75]. There, we make use of the concept of *minimal spread* of a vector. To our knowledge this concept is new, at least in the context of ordinal pattern analysis.

Definition 4.2.2 (Schnurr et al. [75, Definition 2.8])**.** Let $\overline{X}_1 = (X_1, ..., X_d) \in \mathbb{R}^d$. The *minimal spread* of the vector is

$$\mathrm{ms}(\overline{X}_1) := \min\{|X_j - X_k| : 1 \leq j < k \leq d\}.$$

In particular, we show in our proof that the function $\phi :]0, \infty[\to]0, \infty[$ with regard to the p-continuity condition can be explicitly given for our kernel, namely by $\phi(\varepsilon) = 2^p \mathbb{P}(\mathrm{ms}(U) \leq 2\varepsilon)$. Compare in this context Borovkova et al. [18]. They work with the Grassberger-Procaccia dimension estimator and hence, use the family of kernels $h_t(x, y) = \mathbb{1}_{\{|x-y| \leq t\}}$. In order to guarantee that these satisfy the 1-continuity assumption, the family of distribution functions of $|X_j - X_k|$ has to be *equicontinuous* in t. This might be very difficult to check in practice. Using a different kernel and other method of proof, our limit results work without these equicontinuity assumptions [75]. Hence, this kernel is a significant advantage in favor of the entropy concept we are using here.

In order to employ the theorems of Borovkova et al. [18], it is even more intriguing to consider how the r-approximating condition of $(X_t)_{t\in\mathbb{Z}}$ is transferred to $(\overline{X}_t)_{t\in\mathbb{Z}}$. The following lemma shows that it is, at least to some degree, inherited by $(\overline{X}_t)_{t\in\mathbb{Z}}$.

Lemma 4.2.3. *Let $r \geq 1$. If the time series $(X_t)_{t\in\mathbb{Z}}$ can be expressed as an r-approximating functional $X_t = f((Z_{t+k})_{k\in\mathbb{Z}})$ of the stationary time series $(Z_t)_{t\in\mathbb{Z}}$ with approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of size $-\lambda$, then $\overline{X}_t = g((Z_{t+k})_{k\in\mathbb{Z}})$ itself satisfies the r-approximating condition for $k \geq d$ with approximating constants of the same size.*

Proof. Let $(X_t)_{t\in\mathbb{Z}}$ be an r-approximating functional of $(Z_t)_{t\in\mathbb{Z}}$. It holds

$$\begin{aligned}\mathbb{E}\left\|\overline{X}_0 - \mathbb{E}(\overline{X}_0|\mathcal{A}_{-k}^k)\right\|_r^r &= \mathbb{E}\left(\left|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\right|^r + ... + \left|X_{d-1} - \mathbb{E}(X_{d-1}|\mathcal{A}_{-k}^k)\right|^r\right)\\ &= \mathbb{E}\left|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-k}^k)\right|^r + ... + \mathbb{E}\left|X_{d-1} - \mathbb{E}(X_{d-1}|\mathcal{A}_{-k}^k)\right|^r.\end{aligned}$$

Here, we have used the r-norm, which can be done, since every norm on $\mathbb{R}^d$ is equivalent. For $k \geq d$, $\mathcal{A}_{i-(k-i)}^{i+(k-i)} = \mathcal{A}_{-k+2i}^k \subset \mathcal{A}_{-k}^k$ constitutes a sub-σ-algebra of $\mathcal{A}_{-k}^k$ for all $0 \leq i \leq d-1$, so, e.g., [24, Theorem 10.28] yields

$$\mathbb{E}\left|X_i - \mathbb{E}(X_i|\mathcal{A}_{-k}^k)\right|^r \leq 2\mathbb{E}\left|X_i - \mathbb{E}(X_i|\mathcal{A}_{-k+2i}^k)\right|^r$$

for all $0 \leq i \leq d-1$. Hence, using Lemma 2.4.9 it follows

$$\begin{aligned}\mathbb{E}\left\|\overline{X}_0 - \mathbb{E}(\overline{X}_0|\mathcal{A}_{-k}^k)\right\|_r^r &\leq 2\sum_{i=0}^{d-1}\mathbb{E}\left|X_i - \mathbb{E}(X_i|\mathcal{A}_{-k+2i}^k)\right|^r\\ &= 2\sum_{i=0}^{d-1}\mathbb{E}\left|X_0 - \mathbb{E}(X_0|\mathcal{A}_{-(k-i)}^{k-i})\right|^r\\ &\leq 2\sum_{i=0}^{d-1} a_{k-i}.\end{aligned}$$

□

Unfortunately, without further assumptions we cannot validate the r-approximating condition of $(\overline{X}_t)_{t\in\mathbb{Z}}$ for $0 \le k < d$, that is, the boundedness of

$$\mathbb{E}\left\|\overline{X}_0 - \mathbb{E}(\overline{X}_0|\mathcal{A}_{-k}^k)\right\|_r^r \qquad \text{for } 0 \le k \le d-1.$$

In fact, this makes it impossible to use the limit theorems for U-statistics of r-approximating functionals established by Borovkova et al. [18] (cf. the discussion in Section 2.4). For this reason we make use of generalizations which are suitable for our endeavor and which we have derived in Sections 2.4.2–2.4.4. Then, we obtain the following law of large numbers for our estimator S_n^d.

Theorem 4.2.4. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional of a stationary and absolutely regular time series with summable approximating constants $(a_k)_{k\in\mathbb{N}_0}$. Then, it holds*

$$S_n^d \xrightarrow{\mathbb{P}} S^d,$$

as $n \to \infty$.

Note that we are still dealing with a two-sided functional, which is common practice with regard to time series analysis. But since we are considering data in order to estimate the SCI, for simplicity we omit observations indexed by $t \le 0$.

Proof. The statement follows from Theorem 2.4.16, since the time series $(X_t)_{t\in\mathbb{N}}$ is stationary and itself a 1-approximating functional of an absolutely regular time series which approximating constants are summable disregarding the first d constants (Lemma 4.2.3). Furthermore, we have seen in Proposition 4.2.1 that our kernel satisfies the 1-Lipschitz condition. Finally, the family of random variables $(\mathbb{1}_{\{\Pi(\overline{X}_j)=\Pi(\overline{X}_k)\}})_{j,k\in\mathbb{N}}$ is uniformly integrable, since it is bounded by 1, which is obviously integrable w.r.t. a probability measure. □

For a kernel $h : \mathbb{R}^{2d} \to \mathbb{R}$, recall that we write

$$h_1(x) := \int_{\mathbb{R}^d} h(x, y)\ dF(y)$$

(see Section 2.3). Now we can prove a central limit theorem. Since our kernel h is bounded and 1-continuous, the following central limit theorem follows by Theorem 2.4.17.

Theorem 4.2.5. *Let $(X_t)_{t\in\mathbb{N}}$ be a 1-approximating functional of an absolutely regular time series with mixing coefficient $(\beta_k)_{k\in\mathbb{N}_0}$, and let h and h_1 be as above. Suppose that the sequences $(\beta_k)_{k\in\mathbb{N}_0}$, $(a_k)_{k\in\mathbb{N}_0}$ and $(\phi(a_k))_{k\in\mathbb{N}_0}$ satisfy the following summability condition:*

$$\sum_{k=1}^{\infty} k^2(\beta_k + a_k + \phi(a_k)) < +\infty.$$

Then the series

$$\sigma^2 = \operatorname{Var}(h_1(\overline{X}_1)) + 2\sum_{k=2}^{\infty} \operatorname{Cov}(h_1(\overline{X}_1), h_1(\overline{X}_k)) \tag{4.2}$$

converges absolutely and, as $n \to \infty$,

$$\sqrt{n}(S_n^d - S^d) \xrightarrow{D} N(0, 4\sigma^2).$$

Since in our case we know the function ϕ explicitly, we can derive the following:

$$\phi(\varepsilon) = 2\mathbb{P}(\mathrm{ms}(\overline{X}_k) \leq 2\varepsilon) \leq 2 \sum_{k \leq i,j \leq k+d-1} \mathbb{P}(X_i - X_j \leq 2\varepsilon).$$

This means that if the distribution functions $F_{X_i - X_j}$ (for $0 \leq i, j, \leq d-1$) are all Lipschitz continuous, then the summability condition

$$\sum_{k=1}^{\infty} k^2(\beta_k + a_k) < +\infty$$

is sufficient [75].

Due to $S^d > 0$ for all ordinal pattern distributions $(p_\pi)_{\pi\in\Pi}$, by application of the "Delta method" (see [77, Theorem 3.1.A]) we finally obtain the following corollary with regard to the estimation of the Rényi-2 permutation entropy $R_2 = -\log(S^d)$:

Corollary 4.2.6. *Under the assumptions of Theorem 4.2.5, it holds*

$$\sqrt{n}(\log(S_n^d) - \mathrm{PE}_2) \xrightarrow{D} N(0, 4\sigma^2/(S^d)^2)$$

as $n \to \infty$.

Recall that $1/d! \leq S^d \leq 1$. Hence, the limit variance w.r.t. the Rényi-2 permutation entropy becomes larger by application of the "Delta method" if compared to the limit variance of the SCI.

4.3 Estimating the Limit Variance

If we can find an estimator $\hat{\sigma}_n^2$ such that $\hat{\sigma}_n^2 \xrightarrow{D} \sigma^2$, then Slutsky's theorem yields

$$\frac{\sqrt{n}(S_n^d - S^d)}{2\hat{\sigma}_n} \xrightarrow{D} \mathrm{N}(0,1),$$

so, e.g., asymptotic confidence intervals can be determined. De Jong and Davidson [25] have proposed consistent estimators based on kernels for the series on the r.h.s. of σ^2, provided some technical conditions are satisfied. As we are going to employ their results in the following, first we state the necessary assumptions:

Assumption 4.3.1. *Let* $\kappa(\cdot) : \mathbb{R} \to [-1, 1]$ *be a kernel function satisfying the following conditions:*

- $\kappa(0) = 1$,
- $\kappa(\cdot)$ *is symmetric in the sense that* $\kappa(x) = \kappa(-x)$ *for all* $x \in \mathbb{R}$,
- $\kappa(\cdot)$ *is continuous at 0 and at all but a finite number of points,*
- $\int_{-\infty}^{\infty} |\kappa(x)| dx < \infty$, *and*

∘ $\int_{-\infty}^{\infty} |\psi(\xi)| d\xi < \infty$ *where*

$$\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} \kappa(x) e^{i\xi x} dx.$$

The authors stated the next assumption originally in terms of L_r-near epoch dependent triangular arrays. However, for our purpose consideration of time series is enough. Note that considering a time series $(Y_t)_{t\in\mathbb{N}}$ one can always define a triangular array $(Y_{nt})_{n,t\in\mathbb{N}}$ by $Y_{nt} = n^{-1/2} Y_t$. Then, using the triangular constant array c_{nt} defined by $c_{nt} = n^{-1/2}$ yields Assumption 4.3.2 as derived below.

Assumption 4.3.2. *$(Y_t)_{t\in\mathbb{N}}$ is a 2-approximating functional of size -1 on a strongly mixing (resp. uniformly mixing) time series $(Z_t)_{t\in\mathbb{Z}}$ of size $-p/(p-2)$ (resp. $-p/(2p-2)$), and it holds*

$$\sup_{t\geq 1} \|Y_t\|_p < \infty \tag{4.3}$$

for some $p > 2$, where $\|Y\|_p := (\mathbb{E}\,\|Y\|^p)^{1/p}$ denotes the L_p-norm of the random variable Y. Under uniform mixing, $p = 2$ is also permitted if $|Y_t|^2$ is uniformly integrable.

We will not go into "strong mixing" and "uniform mixing" in detail here. For us, it is sufficient to be familiar with their relation to absolute regularity, which is illustrated in Fig. 2.2.

Assumption 4.3.3. *For a bandwidth sequence $(b_n)_{n\in\mathbb{N}}$, it holds*

$$\lim_{n\to\infty} (b_n^{-1} + b_n \cdot n^{-1}) = 0.$$

There are several possible choices for the kernel function $\kappa(\cdot)$ and the bandwidth sequence $(b_n)_{n\geq 1}$ such that Assumptions 4.3.1 and 4.3.3 are fulfilled. One possible choice is the Bartlett-kernel $\kappa(x) = (1 - |x|) \cdot \mathbb{1}_{[-1,1]}(x)$ and $b_n = \log(n)$, $n \in \mathbb{N}$. We will use these in Section 4.4.

We define

$$\sigma_n^2 := \mathrm{Var}(h_1(\overline{X}_1)) + 2\sum_{k=2}^{n} \mathrm{Cov}(h_1(\overline{X}_1), h_1(\overline{X}_k)) = \frac{1}{n} \sum_{i,j=1}^{n} \mathrm{Cov}(h_1(\overline{X}_i), h_1(\overline{X}_j))$$

and its estimator

$$\hat{\sigma}_n^2 := \frac{1}{n} \sum_{i,j=1}^{n} \kappa((j-i)/b_n) \cdot (h_1(\overline{X}_i) - S_n^d)(h_1(\overline{X}_j) - S_n^d).$$

Note that the equality holds due to stationarity of $(\overline{X}_t)_{t\in\mathbb{N}}$. The subsequent theorem shows consistency of the proposed estimator $\hat{\sigma}_n^2$.

Theorem 4.3.4. *Let $(X_t)_{t\in\mathbb{N}}$ be a 2-approximating functional of a stationary absolute regular time series $(Z_t)_{t\in\mathbb{Z}}$ with constants $(a_k)_{k\in\mathbb{N}_0}$ and mixing coefficients $(\beta_k)_{k\in\mathbb{N}_0}$, and let h and h_1 be as above. Suppose that the sequences $(\beta_k)_{k\in\mathbb{N}_0}$, $(a_k)_{k\in\mathbb{N}_0}$ and $(\phi(a_k))_{k\in\mathbb{N}_0}$ satisfy the following summability condition:*

$$\sum_{k=1}^{\infty} k^4 (\beta_k + a_k + \phi(a_k)) < +\infty.$$

Moreover, suppose $\kappa(x)$ and $(b_n)_{n\in\mathbb{N}}$ satisfy Assumptions 4.3.1 and 4.3.3. Then

$$\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$.

The proof is based on De Jong and Davidson [25, Theorem 2.1]. Hence, we need to verify Assumption 4.3.2. In this regard, we need to show that the immediate application of h_1 to $(\overline{X}_t)_{t\in\mathbb{N}}$ yields a 2-approximating functional with approximating constants of the required size. For this we utilize of Lemma 4.3.6 which makes a statement about the immediate application of an r-continuous function onto an r-approximating functional, and which is a generalization of [18, Proposition 2.11] with regard to r-approximating functionals in general for $r \geq 1$. For convenience, first we complement the definition of p-continuous functions with just one argument.

Definition 4.3.5 ([18, Definition 2.10])**.** Let F be some distribution on $\mathbb{R}^d$. A measurable function $g : \mathbb{R}^d \to \mathbb{R}^d$ is called *p-continuous* with respect to F if there exists a function $\phi :]0,\infty[\to]0,\infty[$ with $\phi(\varepsilon) = o(1)$ as $\varepsilon \to 0$ such that

$$\mathbb{E}\left(\left|g(Y) - g(Y')\right|^p \mathbb{1}_{\{|Y-Y'|\leq\varepsilon\}}\right) \leq \phi(\varepsilon) \tag{4.4}$$

holds for all random vectors Y and Y' with distribution F. If the underlying distribution is clearly understood, we simply say that g is p-continuous.

Lemma 4.3.6. *For $r \geq 1$, let $(Y_t)_{t\in\mathbb{N}}$ be an r-approximating functional of $(Z_t)_{t\in\mathbb{N}}$ with constants $(a_k)_{k\in\mathbb{N}_0}$ of size $-\lambda$, and let $g : \mathbb{R}^d \to \mathbb{R}^d$ be r-continuous with respect to the distribution of Y_0. Furthermore, suppose that $\|g(Y_0)^r\|_{2+\delta} < \infty$ for some $\delta > 0$. Then $(g(Y_t))_{t\in\mathbb{N}}$ is also an r-approximating functional of $(Z_t)_{t\in\mathbb{N}}$ with constants*

$$b_k = \phi(\sqrt{2}a_k^{1/2r}) + 2^r \left\|g(Y_0)^r\right\|_{2+\delta} \cdot (\sqrt{2}a_k^{1/2r})^{(1+\delta)/(2+\delta)}. \tag{4.5}$$

If g is bounded, then the same holds with approximating constants

$$b_k = \phi(\sqrt{2}a_k^{1/2r}) + 2^r \left\|g(Y_0)^r\right\|_{\infty} \cdot \sqrt{2}a_k^{1/2r} \tag{4.6}$$

instead of (4.5)*.*

The proof proceeds analogously to the one given in Borovkova et al. [18] for the restricted case of 1-approximating functionals.

Proof. Let $(Z'_t)_{t\in\mathbb{N}}$ be a copy of $(Z_t)_{t\in\mathbb{N}}$ with $Z_t = Z'_t$ for $-k \leq t \leq k$, and let $(Y_t)_{t\in\mathbb{N}}$ and $(Y'_t)_{t\in\mathbb{N}}$ denote the respective functionals. Defining $B := \{|Y_0 - Y'_0| > \varepsilon\}$ it holds

$$\mathbb{E}\left|g(Y_0) - g(Y'_0)\right|^r = \mathbb{E}\left(\left|g(Y_0) - g(Y'_0)\right|^r \cdot \mathbb{1}_{B^c}\right) + \mathbb{E}\left(\left|g(Y_0) - g(Y'_0)\right|^r \cdot \mathbb{1}_{B}\right).$$

While we can use $\phi(\varepsilon)$ as a bound for the first summand due to the r-continuity of g, for the second summand it holds

$$\begin{aligned}\mathbb{E}\left(\left|g(Y_0) - g(Y'_0)\right|^r \cdot \mathbb{1}_{B}\right) &\leq \left\|\left|g(Y_0) - g(Y'_0)\right|^r\right\|_{2+\delta} \cdot \left\|\mathbb{1}_B\right\|_{(2+\delta)/(1+\delta)} \\ &\leq \left\|2^{r-1}\left(\left|g(Y_0)\right|^r + \left|g(Y'_0)\right|^r\right)\right\|_{2+\delta} \cdot \mathbb{P}(B)^{(1+\delta)/(2+\delta)}\end{aligned}$$

$$\leq 2^{r-1}\left(\|g(Y_0)^r\|_{2+\delta} + \left\|g(Y_0')^r\right\|_{2+\delta}\right) \cdot \mathbb{P}(B)^{(1+\delta)/(2+\delta)}$$
$$= 2^r \|g(Y_0)^r\|_{2+\delta} \cdot \mathbb{P}(B)^{(1+\delta)/(2+\delta)},$$

where we have consecutively applied the Hölder-, the c_r- and the Minkowski-inequality, respectively. (Recall the c_r-inequality from Example 2.4.8.)
By the r-approximating condition and Lemma 2.7 (i) of [18] it holds

$$\mathbb{E}\left|Y_0 - Y_0'\right|^r \leq 2^r \cdot a_k.$$

Then, using the Markov-inequality (as well as the Hölder-inequality) yields

$$\mathbb{P}(B) \leq \frac{\mathbb{E}\left|Y_0 - Y_0'\right|}{\varepsilon} \leq \frac{(\mathbb{E}\left|Y_0 - Y_0'\right|^r)^{1/r}}{\varepsilon} \leq \frac{2a_k^{1/r}}{\varepsilon}.$$

Choosing $\varepsilon = \sqrt{2}a_k^{1/2r}$ and summing everything up, it follows

$$\mathbb{E}\left|g(Y_0) - g(Y_0')\right|^r \leq \phi(\sqrt{2}a_k^{1/2r}) + 2^r \|g(Y_0)^r\|_{2+\delta} \cdot (\sqrt{2}a_k^{1/2r})^{(1+\delta)/(2+\delta)},$$

which already implies the r-approximating condition for $g(Y_0)$ by Lemma 2.7 (ii) of [18]. □

Remark 4.3.7. The attentive reader may have noticed that by Lemma 4.3.6 it is implicitly required that the approximating constants $(a_k)_{k\in\mathbb{N}_0}$ of $(Y_t)_{t\in\mathbb{N}}$ are finite, since the function ϕ is defined on the open set $]0,\infty[$. However, if there is a finite number of approximating constants such that $a_k = \infty$, then without loss of generality, we may set $b_k = \infty$ for all k such that $a_k = \infty$, as the approximating property is then still inherited in the same way, especially with regard to the convergence rate of the approximating constants.

Nevertheless, there exists a series of finite constants b_k' such that the approximating condition is satisfied: By assumption it holds $\|g(Y_0)^r\|_{2+\delta} < \infty$, so $\|g(Y_0)\|_r < \infty$ in particular. Example 2.4.8 then yields that $\mathbb{E}\left\|g(Y_0) - \mathbb{E}(g(Y_0)|\mathcal{A}_{-k}^k)\right\|^r$ is bounded by some (finite) constant for all $k \in \mathbb{N}_0$. Let us again emphasize that the number of constants which are affected by this is finite, thus they still do not have an influence on the convergence rate as obtained in Lemma 4.3.6. Therefore, even if some $a_k = \infty$, under the assumptions of Lemma 4.3.6 it follows that there is a series of (finite) approximating constants $(b_k')_{k\in\mathbb{N}_0}$ with respect to $g(Y_0)$ with the convergence rate implied by the above lemma.

Proof of Theorem 4.3.4. As already mentioned, the proof is based on Theorem 2.1 of [25], so we need to verify Assumption 4.3.2. Without loss of generality, here it is sufficient to only consider $Y_t := h_1(\overline{X}_t)$, $t \in \mathbb{N}$, since adding a constant does not affect our results. By Lemma 4.2.3 it holds that $(\overline{X}_t)_{t\in\mathbb{N}}$ is a 2-approximating functional of the same size as $(X_t)_{t\in\mathbb{N}}$, though $a_0, \ldots, a_{d-1}$ are possibly equal to infinity. Furthermore, by Proposition 4.2.1 it follows that our kernel h as defined before is 2-continuous in particular, so Lemma 2.15 of [18] yields that h_1 is also 2-continuous. The 2-approximating condition is preserved when 2-continuous functions are applied (see Lemma 4.3.6), therefore $(Y_t)_{t\in\mathbb{N}}$ remains to be 2-approximating, and even though the size of the approximating constants changes, now all approximating constants are finite (cf. Remark 4.3.7).
Using the supposed summability conditions yields that $k^4\beta_k \to 0$ as $k \to \infty$. Consequently, $(Z_t)_{t\in\mathbb{Z}}$ is absolutely regular of size -4. Note that absolute regularity implies strong mixing of at least same size: For absolute regularity coefficients $(\beta_k)_{k\in\mathbb{N}_0}$ of size $-\lambda_0$ it holds

$$k^{\lambda_0}\alpha_k \leq k^{\lambda_0}\beta_k \to 0$$

as $k \to \infty$, where the series $(\alpha_k)_{k\in\mathbb{N}_0}$ denotes the strong mixing coefficients. (For further information on mixing, see, e.g., [24, Chapter 14].) Therefore, the inequality $-\frac{p}{p-2} \geq -4$ needs to be fulfilled, which is the case for $p \geq \frac{8}{3}$. Equation (4.3) is satisfied due to the assumed stationarity: It holds

$$\sup_{t\geq 1} \|Y_t\|_3 = \left\|h_1(\overline{X}_1)\right\|_3 \leq 1 < \infty,$$

since $h_1(x) = \mathbb{E}(\mathbb{1}_{\pi(x)=\pi(\overline{X}_0)}) \leq 1$ for all x by monotonicity.
Furthermore, in an analogous way it follows that the approximating constants $(a_k)_{k\geq 0}$ of $(X_t)_{t\in\mathbb{N}}$ are of size -4. Denoting the approximating constants of $(\overline{X}_t)_{t\in\mathbb{N}}$ by $(\overline{a}_k)_{k\in\mathbb{N}_0}$, by Lemma 4.3.6 $(\overline{a}_k)_{k\in\mathbb{N}_0}$ is of size -1. □

4.4 Simulations

In this section, we complement our theoretical results with simulations, where we analyze the behavior of our estimators on the basis of three examples. However, we do not consider specific applications such as hypothesis tests here. These are discussed in the joint work [75] with A. Schnurr and M. R. Marín.

Throughout this section, we set $N = 10000$ as the number of repetitions. Furthermore, we use the Bartlett-kernel $\kappa(x) = (1 - |x|) \cdot \mathbb{1}_{[-1,1]}(x)$ and $b_n = \log(n)$, $n \in \mathbb{N}$.

4.4.1 I.I.D. Random Variables

First of all, we consider the most trivial example of an r-approximating functional, namely an i.i.d. series where the functional is given by the identity. To be precise, we consider $n \in \{10000, 30000, 50000\}$ ordinal patterns of length $d = 3$ stemming from data simulated according to a series consisting of $T = n + d - 1$ independent standard normally distributed random variables. Recall that it holds $S^d = 1/d!$ in the i.i.d. case (independent of the choice of the underlying marginal distribution).

The l.h.s. of Fig. 4.1 shows the empirical distribution of the respective estimators S_n^3, while the r.h.s. shows the empirical distribution of its standardization $\sqrt{n}(S_n^3 - 1/6)/2\hat{\sigma}_n$. First of all, observe that the estimator seems to converge from the left to the theoretical value $S_3 = \frac{1}{6}$. In particular, this means that there is an empirical bias to the left. This can be explained by the fact that even though S_d obtains its minimum of $1/d!$ in the i.i.d case, the lower bound of S_n^d lies even below in case of uniformly distributed ordinal patterns and only converges to S_d as n tends to infinity. However, taking a closer look onto the x-axis (on the l.h.s.), we observe that not only the deviation in the mean is already very small for the case $n = 10000$, but also the empirical variance. Furthermore, no significant changes can be found in the empirical densities of the standardized estimator (r.h.s.) for $n > 10000$. Hence, it seems that it has reached its limit distribution. However, it deviates from the asymptotic standard normal distribution proposed in Theorem 4.2.5.

In this regard, let us consider the empirical densities for the standardized versions of S_n^d for $d \in \{2, 3, 4, 5\}$ and $N = n = 10000$ and the respective QQ-Plots in Fig. 4.2. For $d = 2$, a rather strong deviation from the standard normal distribution becomes apparent, which seems to disappear for larger d. How can this be in spite of our theoretical results?

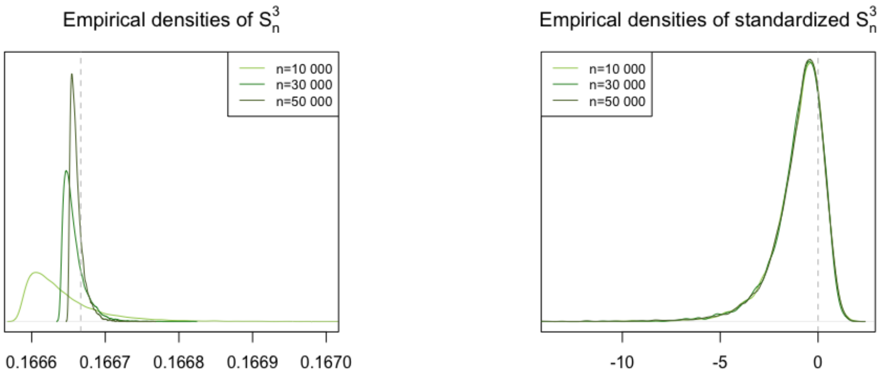


Figure 4.1: Empirical densities of S_n^3 (left) and $\sqrt{n}(S_n^3 - 1/6)/2\hat{\sigma}_n$ (right), where the dashed lines denote the respective theoretical means $S_d = 1/6$ and 0.

Proposition 4.4.1. *Let h and h_1 be as above. If the ordinal patterns of length d are uniformly distributed, then it follows* $\mathrm{Var}(h_1(\overline{X}_1)) = 0$ *and hence, the limit variance σ^2 defined by* (4.2) *equals zero.*

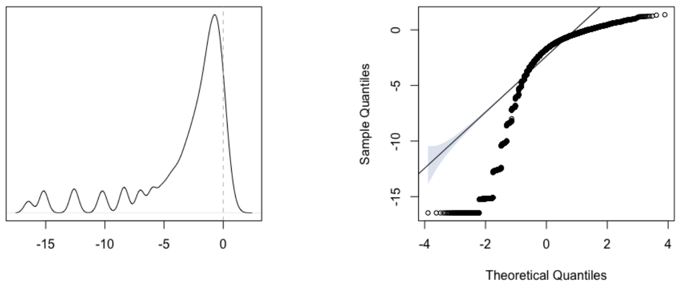
Proof. Since $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$ for any random variable $X \in L_2 := \{Y : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ random variable $: \mathbb{E}|Y|^2 < \infty\}$, it suffices to show that

$$\mathbb{E}(h_1(\overline{X}_1)^2) = (S^d)^2 = \frac{1}{(d!)^2}.$$

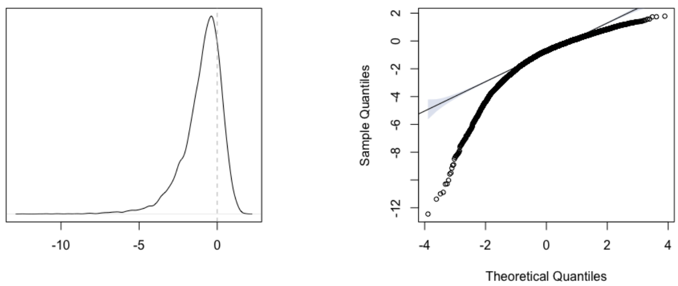
The second claim then follows by definition of the limit variance (4.2) and the Cauchy-Schwarz inequality.

To simplify notation, for the duration of this proof we set $\overline{X} = \overline{X}_1$. Now, let us consider the random variable $h_1(\overline{X})^2$. Using Fubini's theorem twice yields

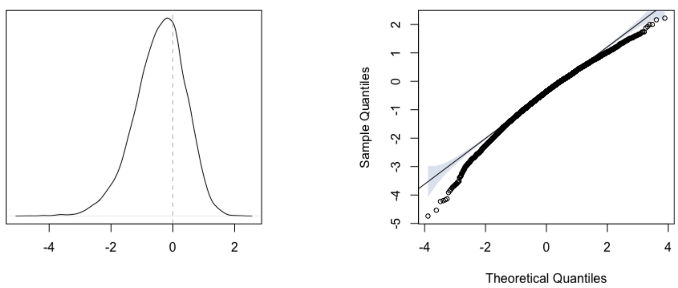
$$\begin{aligned}
h_1(\overline{X})^2 &= \left(\int h(\overline{X}, y) dF(y) \right)^2 \\
&= \int \int h(\overline{X}, y_1) h(\overline{X}, y_2) dF(y_1) dF(y_2) \\
&= \int \int \mathbb{1}_{\{\Pi(\overline{X}) = \Pi(y_1) = \Pi(y_2)\}} dF(y_1) dF(y_2) \\
&= \int \int \sum_{\pi \in S_d} \mathbb{1}_{\{\Pi(\overline{X}) = \pi\}} \mathbb{1}_{\{\Pi(y_1) = \Pi(y_2) = \pi\}} dF(y_1) dF(y_2) \\
&= \sum_{\pi \in S_d} \mathbb{1}_{\{\Pi(\overline{X}) = \pi\}} \int \int \mathbb{1}_{\{\Pi(y_1) = \pi\}} \mathbb{1}_{\{\Pi(y_2) = \pi\}} dF(y_1) dF(y_2) \\
&= \sum_{\pi \in S_d} \mathbb{1}_{\{\Pi(\overline{X}) = \pi\}} \left(\int \mathbb{1}_{\{\Pi(y) = \pi\}} dF(y) \right)^2 \\
&= \sum_{\pi \in S_d} \mathbb{1}_{\{\Pi(\overline{X}) = \pi\}} p_\pi^2,
\end{aligned}$$



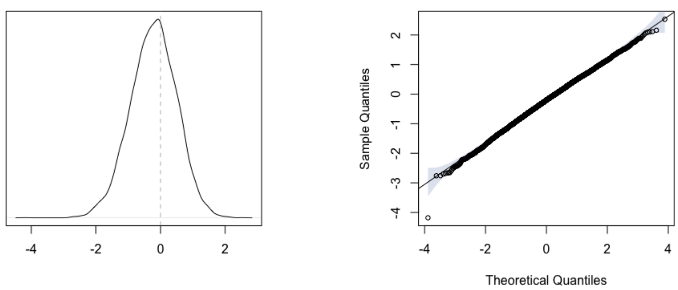
(i)



(ii)



(iii)



(iv)

Figure 4.2: Empirical densities of $\sqrt{n}(S_n^3 - S^d)/2\hat{\sigma}_n$ and the respective QQ-Plots, where $N = n = 10000$ and (i) $d = 2$, (ii) $d = 3$, (iii) $d = 4$, (iv) $d = 5$. The dashed line denotes the theoretical mean 0 and the confidence bands are computed for a nominal level of 5%.

where p_π denotes the probability of the ordinal pattern $\pi \in S_d$. By application of the expected value it then follows

$$\mathbb{E}(h_1(\overline{X})^2) = \sum_{\pi \in S_d} p_\pi^2 \cdot \mathbb{E}\left(\mathbb{1}_{\{\Pi(\overline{X})=\pi\}}\right) = \sum_{\pi \in S_d} p_\pi^3.$$

Note that up to this point, we have not used our assumption, but if we do so now, then we obtain

$$\mathbb{E}(h_1(\overline{X})^2) = \sum_{\pi \in S_d} p_\pi^3 = d! \cdot \left(\frac{1}{d!}\right)^3 = \frac{1}{(d!)^2}.$$

□

Hence, in the i.i.d. case the limit variance equals 0 and thus, we end up in the theoretical limit distribution N$(0, 0)$ which denotes the point mass at the origin (see Theorem 2.4.17). Consequently, the form of the distribution for finite n does not matter. To be precise, it may be of any form here as it only has to converge to a point mass.

The observation above is interesting for two reasons: Firstly, this special case of a limit distribution is not often discussed in the literature for explicit estimators - or at least we have not encountered such a case yet. It is, therefore, crucial to cover this case here, especially for the sake of practitioners. Secondly, this shows the value of the results of Caballero-Pintado et al. [21] who derived limit theorems under the assumption of i.i.d. with regard to a different convergence rate (hence, it does not lead to a one-point distribution). Therefore, our contributions can be seen as a complement rather than a generalization of the results of [21].

Unfortunately, their results do not cover other models where the obtained ordinal pattern distribution is uniform. Furthermore, Proposition 4.4.1 makes no statement about the other direction, that is, whether there exist other ordinal pattern distributions such that the limit variance σ^2 equals zero. Based on the proof, conclusions can only be drawn with regard to $\mathrm{Var}(h_1(\overline{X}_1))$: It is in fact zero if and only if the ordinal pattern distribution is uniform. This holds, since $\sum_{\pi \in S_d} p_\pi^3$ obtains its minimum only in this case. However, if $\mathrm{Var}(h_1(\overline{X}_1))$ is strictly positive, then it still may be canceled out by the autocovariances $\mathrm{Cov}(h_1(\overline{X}_1), h_1(\overline{X}_k))$, $k \geq 2$. Conditions for potential existence of such processes are still an open question.

4.4.2 MA(1)-Process

Now, let us consider two examples of simple r-approximating functionals which exhibit some form of short-range dependence. The examples are chosen in such a way that we can determine the theoretical values of the SCI using the results of Bandt and Shiha [11]. The SCI for most of the other, especially the more complex, time series cannot be computed as nothing is known yet about the ordinal pattern distributions stemming from those.

The first example is given by an MA(1)-process on a Gaussian white noise $(Z_t)_{t\in\mathbb{N}} \sim$ WN$(0, 1)$ with parameter $\theta = 1$, that is,

$$X_t = Z_{t-1} + Z_t.$$

For $d = 2$, the ordinal pattern distribution is uniform (cf. [11, Proposition 7]), therefore we obtain similar results to Fig. 4.2 (i), that is, a one-point distribution. For $d = 3$, however,

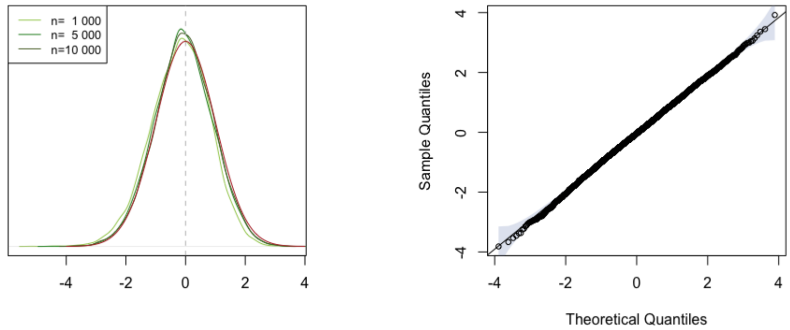


Figure 4.3: Empirical densities of $\sqrt{n}(S_n^3 - S^3)/2\hat{\sigma}_n$ for varying n (left), where $(X_t)_{t\in\mathbb{N}}$ is an MA(1)-process. The dashed line indicates the theoretical mean 0 and the standard normal density is depicted in red. The r.h.s. shows the QQ-plot for $n = 10000$. There the confidence bands are computed for a nominal level of 5%.

each of the monotone patterns (1, 2, 3) and (3, 2, 1) has probability 1/4, while all the other patterns have probability 1/8 each (cf. again [11, Proposition 7]), which results in the SCI given by $S^3 = 3/16$. Our simulation results for $d = 3$ and $N = 10000$ are depicted in Fig. 4.3. The standard normal distribution in the limit can be clearly detected already for $n \geq 1000$.

4.4.3 AR(1)-Process

Finally, let us consider an AR(1)-process on a Gaussian white noise $(Z_t)_{t\in\mathbb{N}} \sim \mathrm{WN}(0, 1)$ with autoregressive parameter $\phi = 0.5$, i.e.,

$$X_t = 0.5X_{t-1} + Z_t.$$

Note that this process is stationary. Let $p^{(1)}$ and $p^{(6)}$ denote the probability of the increasing and decreasing ordinal pattern (1, 2, 3) and (3, 2, 1), respectively. By [11, Corollary 1] it holds

$$p^{(1)} = \frac{1}{\pi} \arcsin\left(\frac{1}{2}\sqrt{1+\phi}\right).$$

Furthermore, [11, Proposition 4] states that $p^{(1)} = p^{(6)}$ and all the remaining ordinal pattern probabilities are given by $(1 - 2p^{(1)})/4$ each. This leads to

$$S^3 = 2 \cdot (p^{(1)})^2 + \frac{(1 - 2p^{(1)})^2}{4} \approx 0.1722$$

for $\phi = 0.5$. Our simulation results are summarized in Fig. 4.4 and Table 4.1. The latter shows that the estimations of the ordinal pattern probabilities are very close to their theoretical values. Compare in this regard also the simulation results of Bandt and Shiha [11, Table 2]. Again, the convergence to a standard normal distribution is clearly there.

4.5 Interim Conclusion

In this chapter, we have complemented the work of Caballero-Pintado et al. [21] by establishing limit theorems for the SCI, and hence the Rényi-2 permutation entropy, for the broad class

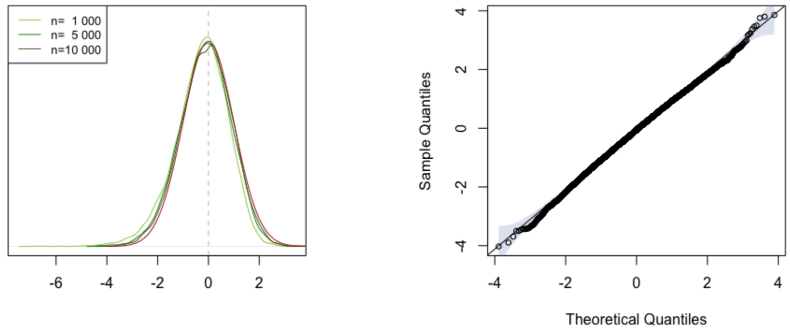


Figure 4.4: Empirical densities of $\sqrt{n}(S_n^3 - S^3)/2\hat{\sigma}_n$ for varying n (left), where $(X_t)_{t\in\mathbb{N}}$ is an AR(1)-process. The dashed line indicates the theoretical mean 0 and the standard normal density is depicted in red. The r.h.s. shows the QQ-plot for $n = 10000$. There the confidence bands are computed for a nominal level of 5%.

Table 4.1: Theoretical ordinal pattern probabilities and their estimations (relative frequencies) obtained from the first run of our simulation with $n = 10000$.

j	1	2	3	4	5	6
$p^{(j)}$	0.210	0.145	0.145	0.145	0.145	0.210
$\hat{p}^{(j)}$	0.212	0.145	0.146	0.143	0.142	0.212

of short-range dependent processes of approximating functionals. We have shown asymptotic normality and estimated the limit variance. Furthermore, we have supported our theoretical results by simulations. In this process, we have seen that the limit variance with regard to our estimator equals 0 in case of a uniform ordinal pattern distribution, for which we also provided a theoretical proof. In this case, the estimator does not converge to a normal distribution, but the special case of a one-point distribution. We recommend practitioners to use the results of Caballero-Pintado et al. [21] in case of an i.i.d. series as an estimator with different convergence rate is discussed there. However, the constraints of our theory remain in case of, e.g., an MA(1)-process and ordinal patterns of length $d = 2$, since there the probabilities of both ordinal patterns coincide resulting in a uniform distribution. Although we do recommend the use of ordinal patterns of length $d = 3$, as more information is retained and the required time for computation is still tractable here, there is still a bitter flavor to the above: It is not known for many, if not most, time series models what the resulting ordinal pattern distributions look like. Hence, neither do we know if it corresponds to a uniform distribution nor can we compute the theoretical value of the SCI. Hence, it makes sense to close the existing gap on ordinal pattern distributions in the literature as good as possible as a part of future research. This is accompanied by the open question with regard to the potential existence of time series models which do not imply a uniform ordinal pattern distribution, but nevertheless a zero limit variance.

However, classification or discrimination problems can still make sense here. In the joint work [75] we have developed hypothesis tests on whether two time series exhibit the same degree of complexity/structure that work without the information on the theoretical SCI. In this regard, we have conducted a simulation study where we have considered different models including AR-, MA- and ARCH-processes with varying parameters. Furthermore, we have illustrated the applicability of our test with a real-world data example given by the BonnEEG-database. In a nutshell, the Bonn EEG-data base is publicly available and consists of 5 data sets. These are given by EEG-recordings of healthy (open and closed eyes) and epileptic patients, whereby the data is recorded either when no epileptic seizure was present (either in the hemisphere of the epileptogenic zone or in the opposite hemisphere) or during an epileptic seizure. For a more detailed description of the dataset, we refer the reader to [6]. We have tested all pairwise possibilities of the 5 EEG data series described above at a significance level of 0.05. In all cases the null hypothesis that the data series come from the same data generating process is rejected. For more details, see [75].

Another direction for future research is to develop the above tests into tests for time-reversibility or even Gaussianity: If a stationary time series is time-reversible, then ordinal patterns would have the same probability as their inversions in time. In case of time-reversibility, in addition to the classic method there is also the possibility of estimating ordinal pattern probabilities by taking the average between the relative frequency of the pattern itself and the relative frequency of its inversion. The difference of the SCI obtained in the classical way and in this way is non-negative. Recall that SCI obtains its minimum for uniform distributions and note that the distribution at hand is artificially brought closer to a uniform distribution by averaging. A large difference would be in favor of time-irreversibility.

In a similar way one can obtain a test for Gaussianity, since for stationary Gaussian stochastic processes both inversions in time and in space have the same probability. Hence, in this case one would average between 4 values. Both tests mentioned fit the more general situation where the set of ordinal patterns is divided into non-empty subsets and the null that inside each subset ordinal pattern probabilities are the same is tested. An analysis with regard to the performance of such tests (possibly compared to the performance of classical tests for Gaussianity) would be particularly interesting.

5 Ordinal Pattern Dependence and Multivariate Measures of Dependence

With regard to dependence *between* time series, ordinal pattern dependence, which has been proposed by Schnurr [70] and Schnurr and Dehling [71], captures how strong the co-movement between two data sets or two time series is. Applications of ordinal pattern dependence include finance [70], manufacturing [82] and hydrology [28, 72, 73].

Grothe et al. [32] have suggested an axiomatic framework for multivariate measures of dependence between random vectors of the same dimension in which Betken et al. [17] aimed to fit ordinal pattern dependence. However, there is an error in the proof of the fifth axiom [79, p. 89ff]. In general, this axiom cannot be verified for ordinal pattern dependence. In this regard, we provide several suitable counterexamples in Section 5.2. In order to make it easier for the reader to understand the idea, these are arranged in such a way that they increase in difficulty and lead to the most common requirements in practice, namely stationarity of a time series. In Section 5.3 we give a proof making different assumptions with regard to the (conditional) cumulative distribution functions and (conditional) survival functions, i.e., we make an alteration to the assumptions on concordance ordering.

While the first four axioms in [32] are (more or less) canonical, the fifth one is strongly inspired by Schmid et al. [69], who deal with dependence *within* one random vector. In particular, the authors deal with copula-based measures of association. With regard to these it is quite natural to define dependence or association in terms of joint cumulative distribution functions and survival functions, but this might not be the case for measures which are not based on copulas, so arguably one could have formulated this axiom differently. This is discussed in Section 5.4. We conclude this chapter with some further thoughts on measuring dependence in Section 5.5.

Sections 5.1–5.3 are largely based on the joint work [80] with A. Schnurr supplemented with further examples.

5.1 Mathematical Background

Recalling that $\Pi : \mathbb{R}^d \to S_d$ denotes the map that assigns to any vector $\mathbf{x}$ its ordinal pattern, first we give the definition of the main quantity under consideration.

Definition 5.1.1. The *ordinal pattern dependence (OPD)* between two random vectors $\mathbf{X} := (X_1, \ldots, X_d)$ and $\mathbf{Y} := (Y_1, \ldots, Y_d)$ is defined by

$$\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y}) := \frac{\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y})) - \sum_{\pi \in S_d} \mathbb{P}(\Pi(\mathbf{X}) = \pi)\mathbb{P}(\Pi(\mathbf{Y}) = \pi)}{1 - \sum_{\pi \in S_d} \mathbb{P}(\Pi(\mathbf{X}) = \pi)\mathbb{P}(\Pi(\mathbf{Y}) = \pi)}.$$

Intuitively speaking, one compares in the numerator the probability of coincident patterns with the hypothetical case of independence, before the value is standardized by the denominator.

Remark 5.1.2. The above definition focuses on positive dependence in a sense, namely the co-occurrence of the same ordinal patterns. One may also consider a kind of negative dependence, namely the co-movement of $\mathbf{X}$ and $-\mathbf{Y} = (-Y_1, \dots, -Y_d)$ instead [82]. A natural way to consider both dependencies at the same time is the generalization to the so-called *standardized OPD* given by

$$\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y})^+ - \mathrm{OPD}_d(\mathbf{X}, -\mathbf{Y})^+,$$

where $a^+ := \max\{0, a\}$ for all $a \in \mathbb{R}$ [17, 71]. In what follows, however, (for simplicity) we only consider (positive) OPD in the sense of Definition 5.1.1.

For any random vector $\mathbf{X} = (X_1, \dots, X_d)$, let

$$F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

and

$$\overline{F}_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \geq \mathbf{x}) = \mathbb{P}(X_1 \geq x_1, \dots, X_d \geq x_d)$$

denote the joint cumulative distribution function (cdf) and survival function, respectively. Note that for $d \geq 2$ in general $\overline{F}_{\mathbf{X}}(\mathbf{x}) \neq 1 - F_{\mathbf{X}}(\mathbf{x})$.

With the following definition, Grothe et al. [32] proposed an axiomatic theory for multivariate measures of dependence between random vectors of the same dimension. We state it in the wording of Betken et al. [17]:

Definition 5.1.3 (Betken et al. [17, Definition 2])**.** Let L_0 denote the space of random vectors with values in $\mathbb{R}^d$ on the common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A function $\mu : L_0 \times L_0 \to \mathbb{R}$ is called a *d-dimensional measure of dependence* if

1. it takes values in $[-1, 1]$,
2. it is invariant with respect to simultaneous permutations of the components within two random vectors $\mathbf{X}, \mathbf{Y}$,
3. it is invariant with respect to increasing transformations of the components within two random vectors $\mathbf{X}, \mathbf{Y}$,
4. it is zero for two independent random vectors $\mathbf{X}, \mathbf{Y}$,
5. it respects *concordance ordering*, i.e., for two pairs of random vectors $\mathbf{X}, \mathbf{Y}$ and $\mathbf{X}^*, \mathbf{Y}^*$ which satisfy $\mathbf{X} \stackrel{D}{=} \mathbf{X}^*$ and $\mathbf{Y} \stackrel{D}{=} \mathbf{Y}^*$, it holds that

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \preccurlyeq_C \begin{pmatrix} \mathbf{X}^* \\ \mathbf{Y}^* \end{pmatrix} \Rightarrow \mu(\mathbf{X}, \mathbf{Y}) \leq \mu(\mathbf{X}^*, \mathbf{Y}^*).$$

Here, $\preccurlyeq_C$ denotes concordance ordering, i.e.,

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \preccurlyeq_C \begin{pmatrix} \mathbf{X}^* \\ \mathbf{Y}^* \end{pmatrix} \text{ if and only if } F_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}} \text{ and } \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}},$$

where $\leq$ is meant pointwise.

In Theorem 2.3 of [17] it has been claimed that OPD is a d-dimensional measure of dependence. While the proof of the first four axioms is correct, there is an error regarding the treatment of the conditional probabilities and concordance ordering in the proof of the fifth axiom (see Silbernagel [79, p. 89ff]).

5.2 Counterexamples

We now give counterexamples showing that OPD cannot satisfy the fifth axiom and thus does not fit into the axiomatic framework of multivariate measures of dependence as proposed by Grothe et al. [32]. As stated by Betken et al. [17], for the verification of the last axiom it is sufficient to restrict our considerations to the probability $\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y}))$, since the remaining terms of $\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y})$ only relate to the distributions of $\mathbf{X}$ and $\mathbf{Y}$ separately. (Note that $\mathbf{X} \overset{D}{=} \mathbf{X}^*$ and $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$.) For $d = 2$, there are only two patterns, namely the upward pattern (1, 2) and the downward pattern (2, 1). We have to deal with vectors of length $2d$ when considering dependence, so even in case of $d = 2$, the problem is 4-dimensional.

At first, consider the following illustrative example for a discrete state space:

Example 5.2.1. Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}^*, \mathbf{Y}^*)$ be random vectors on a common probability space with state space $\mathbb{N}^4$ defined by the probabilities

$$\mathbb{P}((\mathbf{X}, \mathbf{Y}) = (1, 2, 3, 4)) = \mathbb{P}((\mathbf{X}, \mathbf{Y}) = (3, 2, 3, 2)) = \frac{1}{2}$$
$$\mathbb{P}((\mathbf{X}^*, \mathbf{Y}^*) = (1, 2, 3, 2)) = \mathbb{P}((\mathbf{X}^*, \mathbf{Y}^*) = (3, 2, 3, 4)) = \frac{1}{2},$$

where $\mathbf{X} = (X_1, X_2), \mathbf{Y} = (Y_1, Y_2), \mathbf{X}^* = (X_1^*, X_2^*)$ and $\mathbf{Y}^* = (Y_1^*, Y_2^*)$ denote the two-dimensional marginals, respectively. It is easy to see that the marginals satisfy the desired equalities in distribution, i.e., $\mathbf{X} \overset{D}{=} \mathbf{X}^*$ and $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$. (For illustration of the respective distributions, see Fig. 5.1).

The cdfs as well as the survival functions of $(\mathbf{X}, \mathbf{Y})$ and $(\mathbf{X}^*, \mathbf{Y}^*)$, respectively, are outlined in Table 5.1. Note that here it is sufficient to state the respective functions at the mass points, since they are constant hereinafter up to the next mass point. Denoting the respective cdfs by $F_{\binom{\mathbf{X}}{\mathbf{Y}}}$ and $F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ as well as the respective survival functions by $\overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}}$ and $\overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$, we observe $F_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ and $\overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ pointwise. Thus, the defined random vectors satisfy the required conditions.

However, regarding the probabilities with respect to the increasing pattern it holds

$$\begin{aligned}\mathbb{P}(X_1 \leq X_2, Y_1 \leq Y_2) &= \mathbb{P}((\mathbf{X}, \mathbf{Y}) = (1, 2, 3, 4)) \\ &= \frac{1}{2} > 0 = \mathbb{P}(X_1^* \leq X_2^*, Y_1^* \leq Y_2^*).\end{aligned}$$

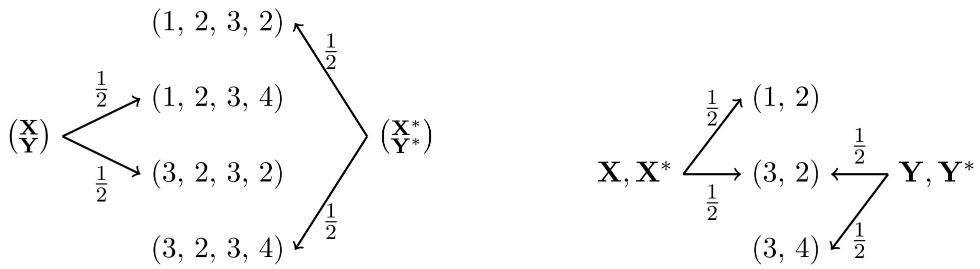
Analogously, for the decreasing pattern it holds

$$\begin{aligned}\mathbb{P}(X_1 \geq X_2, Y_1 \geq Y_2) &= \mathbb{P}((\mathbf{X}, \mathbf{Y}) = (3, 2, 3, 2)) \\ &= \frac{1}{2} > 0 = \mathbb{P}(X_1^* \geq X_2^*, Y_1^* \geq Y_2^*).\end{aligned}$$

Consequently, this yields

$$\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y})) = 1 > 0 = \mathbb{P}(\Pi(\mathbf{X}^*) = \Pi(\mathbf{Y}^*)),$$

and hence, $\mathrm{OPD}_2(\mathbf{X}, \mathbf{Y}) > \mathrm{OPD}_2(\mathbf{X}^*, \mathbf{Y}^*)$. Thus, we have found two vectors which satisfy the property of concordance ordering showing the opposite behavior in terms of OPD.



(i) (ii)

Figure 5.1: Illustration of the respective pmfs: joint pmfs in (i) and pmfs of the respective marginals in (ii). Here, the arrows imply that the random vector of interest attains the respective outcome with the indicated probability (of 1/2).

Table 5.1: Cdfs and survival functions of $(\mathbf{X}, \mathbf{Y})$ and $(\mathbf{X}^*, \mathbf{Y}^*)$, respectively.

(x_1, x_2, y_1, y_2)	$F_{\binom{\mathbf{X}}{\mathbf{Y}}}$	$F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$	$\overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}}$	$\overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$
(1, 2, 3, 2)	0	$\frac{1}{2}$	1	1
(1, 2, 3, 4)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(3, 2, 3, 2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(3, 2, 3, 4)	1	1	0	0

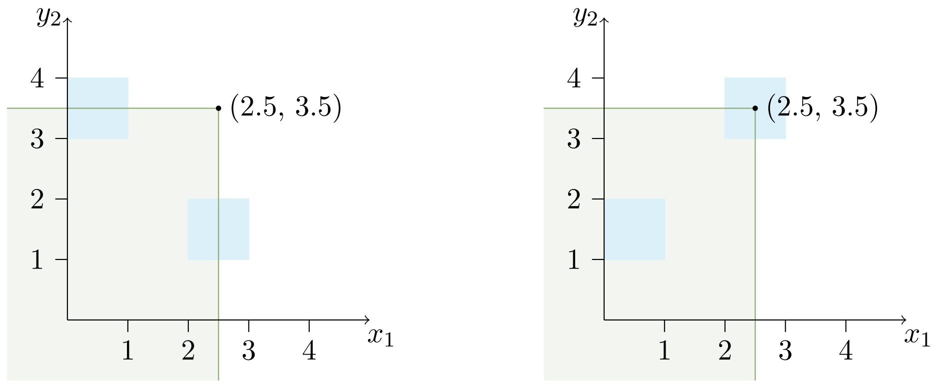
Even though the consideration of ties is avoided in this example, it is still discrete-valued. Thus, we extend it to the $\mathbb{R}^4$-valued case by proposing a continuous uniform distribution based on the aforementioned points:

Example 5.2.2. Define the probability densities $f_{\binom{\mathbf{X}}{\mathbf{Y}}}$ and $f_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ by

$$
\begin{aligned}
f_{\binom{\mathbf{X}}{\mathbf{Y}}}(x_1, x_2, y_1, y_2) = \frac{1}{2} \cdot \mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) \cdot \mathbb{1}_{]2,3[}(y_1) \cdot \mathbb{1}_{]3,4[}(y_2) \\
+ \frac{1}{2} \cdot \mathbb{1}_{]2,3[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) \cdot \mathbb{1}_{]2,3[}(y_1) \cdot \mathbb{1}_{]1,2[}(y_2) \\
f_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}(x_1, x_2, y_1, y_2) = \frac{1}{2} \cdot \mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) \cdot \mathbb{1}_{]2,3[}(y_1) \cdot \mathbb{1}_{]1,2[}(y_2) \\
+ \frac{1}{2} \cdot \mathbb{1}_{]2,3[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) \cdot \mathbb{1}_{]2,3[}(y_1) \cdot \mathbb{1}_{]3,4[}(y_2),
\end{aligned}
$$

where $\mathbb{1}_{]a,b[}(z)$ denotes the indicator function on the open interval $]a, b[$ for $a, b \in \mathbb{R}, a < b$. Concerning the marginal densities it holds

$$
\begin{aligned}
f_{\mathbf{X}}(x_1, x_2) &= \int f_{\binom{\mathbf{X}}{\mathbf{Y}}}(x_1, x_2, y_1, y_2)\, d(y_1, y_2) \\
&= \frac{1}{2} \cdot \mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) + \frac{1}{2} \cdot \mathbb{1}_{]2,3[}(x_1) \cdot \mathbb{1}_{]1,2[}(x_2) \\
&= \int f_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}(x_1, x_2, y_1, y_2)\, d(y_1, y_2) \\
&= f_{\mathbf{X}^*}(x_1, x_2),
\end{aligned}
$$



(i) **(ii)**

Figure 5.2: Illustration of probability mass of the auxiliary functions h (left) and h^* (right): The blue squares denote the places where the respective functions hold mass 0.5. Otherwise they are zero. Considering the rectangle $]-\infty, 2.5] \times]-\infty, 3.5]$ (green) as an illustrative example, we observe that h has less cumulative mass than h^*. This holds for any point (x_1, y_2).

and hence, $\mathbf{X} \overset{D}{=} \mathbf{X}^*$. Analogously, it follows $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$.

Regarding the verification of the condition on the cdfs, it suffices to take a closer look at the respective density functions. We observe that the term $\mathbb{1}_{]1,2[}(x_2) \cdot \mathbb{1}_{]2,3[}(y_1)$ appears in each summand of both of the density functions. Therefore, it is sufficient to fix $x_2 > 1$ and $y_1 > 2$, and to consider subsequently the two-dimensional auxiliary functions

$$h(x_1, y_2) := \frac{1}{2} \cdot \mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]3,4[}(y_2) + \frac{1}{2} \cdot \mathbb{1}_{]2,3[}(x_1) \cdot \mathbb{1}_{]1,2[}(y_2)$$
$$h^*(x_1, y_2) := \frac{1}{2} \cdot \mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]1,2[}(y_2) + \frac{1}{2} \cdot \mathbb{1}_{]2,3[}(x_1) \cdot \mathbb{1}_{]3,4[}(y_2),$$

which determine where the mass of the probability distribution lies. For any point $(x_1, y_2) \in \mathbb{R}^2$ we consider the resulting rectangle $]-\infty, x_1] \times]-\infty, y_2]$. For any of these, h has less cumulative mass than h^*, which shows that it holds $F_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ pointwise (see Fig. 5.2). Fixing $x_2 < 1$ and $y_1 < 2$, and considering rectangles of the form $[x_1, \infty[\times [y_2, \infty[$ for $x_1, y_2 \in \mathbb{R}$, the respective result for the survival functions, i.e., $\overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ pointwise, follows analogously.

Nevertheless, with a close look at the probability densities it follows

$$\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y})) = 1 > 0 = \mathbb{P}(\Pi(\mathbf{X}^*) = \Pi(\mathbf{Y}^*)),$$

which completes this counterexample.

In practice, however, ordinal patterns, and OPD in particular, are often considered in the context of stationary time series. This can be broken down to the additional condition of $X_1 \overset{D}{=} X_2$ and $Y_1 \overset{D}{=} Y_2$ in our case. Obviously, the counterexamples previously presented do not satisfy these additional restrictions. However, such counterexamples still exist, though these then become more complicated. This is due to the fact that the components of $\mathbf{X}$ and $\mathbf{Y}$, respectively, have to be defined on the same state space. If we want to proceed as before, that is, if we want to give the idea by considering the discrete case first, then this would lead us to not being able to make a statement on the joint cdfs and survival functions in the respective mass points as, e.g., there is no order for $(1,2,1,2)$ and $(2,1,1,2)$ in the 4-dimensional space. However, this does not mean that discrete stationary examples do not exist - it is only potentially very tendious to determine the cdfs and survival functions for every relevant point by hand. Therefore, we directly consider the following example with regard to the continuous case:

Example 5.2.3. We consider the probability densities $f = f_{\binom{\mathbf{X}}{\mathbf{Y}}}$ and $f^* = f_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ defined by

$$\begin{aligned} f(x_1,x_2,y_1,y_2) &= \mathbb{1}_{\{0\le x_1\le x_2\le 1\}}\cdot \mathbb{1}_{\{1<y_1\le y_2\le 2\}} + \mathbb{1}_{\{0\le x_2< x_1\le 1\}}\cdot \mathbb{1}_{\{1<y_2<y_1\le 2\}} \\ &\quad + \mathbb{1}_{\{1<x_1\le x_2\le 2\}}\cdot \mathbb{1}_{\{0\le y_1\le y_2\le 1\}} + \mathbb{1}_{\{1<x_2<x_1\le 2\}}\cdot \mathbb{1}_{\{0\le y_2<y_1\le 1\}} \\ f^*(x_1,x_2,y_1,y_2) &= \mathbb{1}_{\{0\le x_1\le x_2\le 1\}}\cdot \mathbb{1}_{\{0\le y_2<y_1\le 1\}} + \mathbb{1}_{\{0\le x_2< x_1\le 1\}}\cdot \mathbb{1}_{\{1<y_2<y_1\le 2\}} \\ &\quad + \mathbb{1}_{\{1<x_1\le x_2\le 2\}}\cdot \mathbb{1}_{\{0\le y_1\le y_2\le 1\}} + \mathbb{1}_{\{1<x_2<x_1\le 2\}}\cdot \mathbb{1}_{\{1<y_1\le y_2\le 2\}}. \end{aligned}$$

Obviously it holds

$$\begin{aligned} f_{\mathbf{X}}(x_1,x_2) &= \frac{1}{2}\left(\mathbb{1}_{\{0\le x_1\le x_2\le 1\}} + \mathbb{1}_{\{0\le x_2<x_1\le 1\}}\right) + \frac{1}{2}\left(\mathbb{1}_{\{1<x_1\le x_2\le 2\}} + \mathbb{1}_{\{1<x_2<x_1\le 2\}}\right) \\ &= \frac{1}{2}\left(\mathbb{1}_{\{0\le x_1,x_2\le 1\}} + \mathbb{1}_{\{1<x_1,x_2\le 2\}}\right) \\ &= f_{\mathbf{X}^*}(x_1,x_2) \end{aligned}$$

as well as

$$f_{X_1}(x) = \frac{1}{2}\left(\mathbb{1}_{\{0\le x\le 1\}} + \mathbb{1}_{\{1<x\le 2\}}\right) = \frac{1}{2}\mathbb{1}_{\{0\le x\le 2\}} = f_{X_2}(x)$$

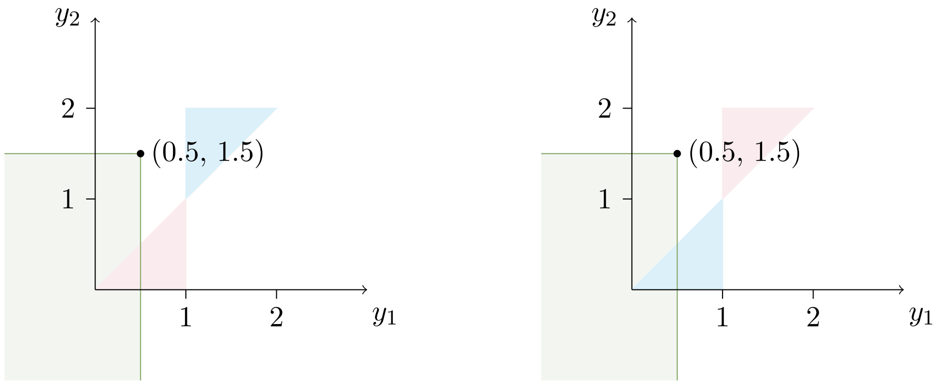
such that $\mathbf{X} \overset{D}{=} \mathbf{X}^*$ and $X_1 \overset{D}{=} X_2$. Analogously it follows $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$ and $Y_1 \overset{D}{=} Y_2$.

In order to prove the conditions on the joint cdfs and survival functions, first we take a closer look at the respective density functions and observe that the summands $\mathbb{1}_{\{0\le x_2<x_1\le 1\}}\cdot \mathbb{1}_{\{1<y_2<y_1\le 2\}}$ and $\mathbb{1}_{\{1<x_1\le x_2\le 2\}}\cdot \mathbb{1}_{\{0\le y_1\le y_2\le 1\}}$ appear in both f and f^*. Therefore, for our purpose it is sufficient to only consider the auxiliary functions h and h^* defined by

$$\begin{aligned} h(x_1,x_2,y_1,y_2) &:= \mathbb{1}_{\{0\le x_1\le x_2\le 1\}}\cdot \mathbb{1}_{\{1<y_1\le y_2\le 2\}} + \mathbb{1}_{\{1<x_2<x_1\le 2\}}\cdot \mathbb{1}_{\{0\le y_2<y_1\le 1\}} \\ h^*(x_1,x_2,y_1,y_2) &:= \mathbb{1}_{\{0\le x_1\le x_2\le 1\}}\cdot \mathbb{1}_{\{0\le y_2<y_1\le 1\}} + \mathbb{1}_{\{1<x_2<x_1\le 2\}}\cdot \mathbb{1}_{\{1<y_1\le y_2\le 2\}}. \end{aligned}$$

Now, let us consider Fig. 5.3. There, it is depicted where the respective auxiliary functions h and h^* hold mass, where blue denotes the case $0 \le x_1 \le x_2 \le 1$, while red denotes the case $1 < x_2 < x_1 \le 2$. Note that each auxiliary function contains both cases, and only these two cases in particular. Hence, it is sufficient to limit our considerations with regard to the comparison of the respective cdfs and survival functions to these two scenarios.

With regard to the cdfs, in case of $0 \le x_1 \le 1$ **or** $0 \le x_2 \le 1$ we only consider the mass denoted by the blue triangles, while we need to consider the triangles of both colors in case of $x_1, x_2 > 1$. Then, for both scenarios rectangles of the form $]-\infty, y_1] \times]-\infty, y_2]$ contain more mass with regard to h^* if compared to h, so $F_{\binom{\mathbf{X}}{\mathbf{Y}}} \le F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ pointwise. In contrast, with regard to survival functions we need to proceed 'the other way around': For $0 \le x_1 \le 1$ **or** $0 \le x_2 \le 1$ we only consider the red triangles. In case of $x_1, x_2 > 1$ it is both colors again. Hence, for rectangles of the form $[y_1, \infty[\times [y_2, \infty[$, h^* holds more mass, and therefore $\overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}} \le \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}$ pointwise, which shows that all conditions are satisfied.



(i) **(ii)**

Figure 5.3: Illustration of the mass of h (left) and h^* (right): The functions attain the value 1 for $0 \leq x_1 \leq x_2 \leq 1$ in the blue area and for $1 \leq x_2 \leq x_1 \leq 2$ in the red area. Otherwise they are zero. Consider the rectangle $]-\infty, 0.5] \times]-\infty, 1.5]$ (green) as an illustrative example with regard to the respective cdfs.

Now, we return to considering the density functions f and f^*. By construction of f, an increasing pattern $x_1 \leq x_2$ in the first component occurs if and only if it also occurs in the second component, that is $y_1 \leq y_2$. The same holds true for decreasing patterns. However, the situation is different with regard to f^*: There, coincident patterns are only predetermined by two of the four summands. The remaining two summands only allow for non-coincident patterns. Then, due to

$$\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y})) = 1 > \frac{1}{2} = \mathbb{P}(\Pi(\mathbf{X}^*) = \Pi(\mathbf{Y}^*))$$

it follows $\mathrm{OPD}_2(\mathbf{X}, \mathbf{Y}) > \mathrm{OPD}_2(\mathbf{X}^*, \mathbf{Y}^*)$.

5.3 Proof under Slightly Different Assumptions

Although OPD does not satisfy the fifth axiom and hence does not fit into the axiomatic framework of multivariate measures of dependence as proposed by Grothe et al. [32], we can prove a similar result under stronger/different assumptions.

In what follows, we consider OPD with respect to random vectors which stem from stationary bivariate time series $(X_t, Y_t)_{t\in\mathbb{N}}$, $(X_t^*, Y_t^*)_{t\in\mathbb{N}}$ with continuous marginal distributions. Then, due to the assumed stationarity, it is sufficient to consider the first components of the time series, i.e., without loss of generality we consider

$$\begin{aligned}\mathbf{X} &= (X_1, \ldots, X_d), \mathbf{Y} = (Y_1, \ldots, Y_d),\\ \mathbf{X}^* &= (X_1^*, \ldots, X_d^*), \mathbf{Y}^* = (Y_1^*, \ldots, Y_d^*).\end{aligned}$$

Theorem 5.3.1. *Let $\mathbf{X}, \mathbf{X}^*, \mathbf{Y}$ and $\mathbf{Y}^*$ be d-dimensional random vectors with $\mathbf{X} \overset{D}{=} \mathbf{X}^*$ and $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$. Let $J = \{1, \ldots, d\}$ denote an index set.*
(a) Suppose for all $I \subset J$, $I^c := J \setminus I$, with $I, I^c \neq \emptyset$ it holds

$$F_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \quad \text{and} \quad \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}}$$

as well as

$$F_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \quad \text{and} \quad \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}$$

pointwise, where $\binom{\mathbf{X}}{\mathbf{Y}}^{I} = (\mathbf{X}^I, \mathbf{Y}^I)$ *denotes the subvector of variables with indices in* I*. Then, it holds that*

$$\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y}) \leq \mathrm{OPD}_d(\mathbf{X}^*, \mathbf{Y}^*). \tag{5.1}$$

(b) Alternatively, the condition

$$F_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \leq F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \quad \text{and} \quad \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \leq \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \tag{5.2}$$

for all $I \subset J$ *with* $I \neq J$ *is sufficient for* (5.1) *to hold. In particular, with* $I = \emptyset$ *we make a statement about the (unconditional) cdf and survival function, meaning that this case corresponds to concordance ordering.*

Note that Theorem 2.5.3 ensures existence and uniqueness of the regular conditional distributions used above, such that the (conditional) cdfs and survival functions are defined for almost all $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2\cdot|I|}$ by definition, respectively. This statement on regular conditional distributions is important as in general the consideration of conditional probabilities where the event that is conditioned on has probability mass zero can be problematic. The proof of Theorem 5.3.1 follows the ideas of Betken et al. [17] and hence, makes use of the so-called disintegration theorem. Basically, disintegration denotes the representation of a (conditional) expectation as an integral in terms of a (regular conditional) probability distribution (for a brief overview, see Chapter 2.5).

Proof of Theorem 5.3.1. We give a proof for $d = 2$ and $d = 3$. Though the difficulties of the proof are not revealed for $d = 2$, it gives an intuition how to proceed in case $d = 3$. The proof for $d > 3$ works analogously to the case $d = 3$, but is notationally more complicated.

Since the remaining terms of $\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y})$ only relate to the distributions of $\mathbf{X}$ and $\mathbf{Y}$ separately, we restrict our considerations to

$$\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y})) = \sum_{\pi \in S_d} \mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y}) = \pi).$$

Now, without loss of generality, it is sufficient to only consider the probability

$$\mathbb{P}(\Pi(\mathbf{X}) = \Pi(\mathbf{Y}) = \pi)$$

with regard to **one** arbitrary pattern $\pi \in S_d$, as for any other pattern $\pi' \neq \pi$ there exists a permutation $\sigma : \{1, \ldots, d\} \to \{1, \ldots, d\}$ such that

$$\mathbb{P}(\Pi(X_1, \ldots, X_d) = \Pi(Y_1, \ldots, Y_d) = \pi') = \mathbb{P}(\Pi(X_{\sigma(1)}, \ldots, X_{\sigma(d)}) = \Pi(Y_{\sigma(1)}, \ldots, Y_{\sigma(d)}) = \pi).$$

Hence, everything can be reduced to a re-indexing such that considering these probabilities for the respective patterns is analogous. Here, we choose the increasing patterns (1, 2) and (1, 2, 3), respectively.

We denote the conditional cdfs and survival functions by

$$\begin{aligned}
\overline{F}^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}}(\mathbf{x},\mathbf{y}) &:= \overline{F}^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}\Big|\binom{\mathbf{X}}{\mathbf{Y}}^{I}}(\mathbf{x},\mathbf{y}) \\
&:= \mathbb{P}\left((\mathbf{X},\mathbf{Y})^{I^c} \geq (\mathbf{x},\mathbf{y}) | X_{i_j} = x_{i_j}, Y_{i_j} = y_{i_j} \forall j \in \{1,\dots,k\}\right) \\
\overline{F}^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}}(\mathbf{x},\mathbf{y}) &:= \overline{F}^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}\Big|\binom{\mathbf{X}}{\mathbf{Y}}^{I}}(\mathbf{x},\mathbf{y}) \\
&:= \mathbb{P}\left((\mathbf{X}^*,\mathbf{Y}^*)^{I^c} \geq (\mathbf{x},\mathbf{y}) | X_{i_j} = x_{i_j}, Y_{i_j} = y_{i_j} \forall j \in \{1,\dots,k\}\right) \\
F^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}\Big|\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}(\mathbf{x},\mathbf{y}) &:= \mathbb{P}\left((\mathbf{X},\mathbf{Y})^{I^c} \leq (\mathbf{x},\mathbf{y}) | X^*_{i_j} = x_{i_j}, Y^*_{i_j} = y_{i_j} \forall j \in \{1,\dots,k\}\right) \\
F^{x_{i_1},\dots,x_{i_k},y_{i_1},\dots,y_{i_k}}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}\Big|\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}(\mathbf{x},\mathbf{y}) &:= \mathbb{P}\left((\mathbf{X}^*,\mathbf{Y}^*)^{I^c} \leq (\mathbf{x},\mathbf{y}) | X^*_{i_j} = x_{i_j}, Y^*_{i_j} = y_{i_j} \forall j \in \{1,\dots,k\}\right)
\end{aligned}$$

for all $I = \{i_1,\dots,i_k\} \subset \{1,\dots,d\}$ with $1 \leq k \leq d-1$, $I^c := \{1,\dots,d\} \setminus I$ and $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^{2(d-k)}$.

Suppose $d = 2$. We begin with the proof of (a) in this setting. Using disintegration as stated in Theorem 2.5.4 (or alternatively the law of total expectation), it holds

$$\begin{aligned}
\mathbb{P}(X_1 \leq X_2, Y_1 \leq Y_2) &= \int_{\mathbb{R}^2} \mathbb{P}(X_1 \leq X_2, Y_1 \leq Y_2 | X_1 = x_1, Y_1 = y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1) \\
&= \int_{\mathbb{R}^2} \mathbb{P}(x_1 \leq X_2, y_1 \leq Y_2 | X_1 = x_1, Y_1 = y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1) \\
&= \int_{\mathbb{R}^2} \overline{F}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1}{X_2}}(x_1,y_1).
\end{aligned}$$

Using our assumption and reversing the previous steps yields

$$\begin{aligned}
\int_{\mathbb{R}^2} \overline{F}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1) &\leq \int_{\mathbb{R}^2} \overline{F}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1) \\
&= \mathbb{P}(X_1 \leq X_2^*, Y_1 \leq Y_2^*).
\end{aligned}$$

In an analogous way we obtain

$$\begin{aligned}
\mathbb{P}(X_1 \leq X_2^*, Y_1 \leq Y_2^*) &= \int_{\mathbb{R}^2} F^{x_2,y_2}_{\binom{X_1}{Y_1}\Big|\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2) \\
&\leq \int_{\mathbb{R}^2} F^{x_2,y_2}_{\binom{X_1^*}{Y_1^*}\Big|\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2) \\
&= \mathbb{P}(X_1^* \leq X_2^*, Y_1^* \leq Y_2^*).
\end{aligned}$$

Under the assumptions in (b) the desired inequality follows due to the fact that

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \preccurlyeq_C \begin{pmatrix}\mathbf{X}^*\\ \mathbf{Y}^*\end{pmatrix} \quad \text{implies} \quad \begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}^{I'} \preccurlyeq_C \begin{pmatrix}\mathbf{X}^*\\ \mathbf{Y}^*\end{pmatrix}^{I'}$$

for all subvectors of variables with indices in $I' \subset J$, i.e., removing dimensions does not influence which scenario has the larger dependence measure [32]. Note that $\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \preccurlyeq_C \begin{pmatrix}\mathbf{X}^*\\ \mathbf{Y}^*\end{pmatrix}$

holds due to Equation (5.2) with $I = \emptyset$. Hence, we deduce that

$$\int \mathbb{1}_{[a,\infty[\times[b,\infty[}(x,y)\, d\mathbb{P}_{\binom{X_i}{Y_i}}(x,y) \le \int \mathbb{1}_{[a,\infty[\times[b,\infty[}(x,y)\, d\mathbb{P}_{\binom{X_i^*}{Y_i^*}}(x,y) \tag{5.3}$$

for all $a, b \in \mathbb{R}$ and $i \in \{1, \ldots, d\}$. Survival functions can be approximated by sums of indicator functions, i.e., $\overline{F}(x,y) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{1}_{\{[a_k,\infty[\times[b_k,\infty[\}}(x,y)$ for constants $a_k, b_k \in \mathbb{R}$, so the bounded convergence theorem yields

$$\begin{aligned}\int_{\mathbb{R}^2} \overline{F}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1) &\le \int_{\mathbb{R}^2} \overline{F}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1^*}{Y_1^*}}(x_1,y_1)\\ &= \mathbb{P}(X_1^* \le X_2^*, Y_1^* \le Y_2^*).\end{aligned}$$

Now, suppose $d = 3$. Defining

$$\mathbb{P}^{x_1,y_1}(A) := \mathbb{P}(A | X_1 = x_1, Y_1 = y_1)$$

for any event A, and using disintegration (Theorem 2.5.4) twice yields

$$\begin{aligned}&\mathbb{P}(X_1 \le X_2 \le X_3, Y_1 \le Y_2 \le Y_3)\\ &\quad= \int_{\mathbb{R}^2} \mathbb{P}^{x_1,y_1}(x_1 \le X_2 \le X_3, y_1 \le Y_2 \le Y_3)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)\\ &\quad= \int_{\mathbb{R}^2} \int_{[x_1,\infty[\times[y_1,\infty[} \mathbb{P}^{x_1,y_1}(x_2 \le X_3, y_2 \le Y_3 | X_2 = x_2, Y_2 = y_2)\\ &\qquad\qquad d\mathbb{P}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1).\end{aligned}$$

Since

$$\begin{aligned}&\mathbb{P}^{x_1,y_1}(x_2 \le X_3, y_2 \le Y_3 | X_2 = x_2, Y_2 = y_2)\\ &\quad= \mathbb{P}(x_2 \le X_3, y_2 \le Y_3 | X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2)\\ &\quad= \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3}{Y_3}}(x_2,y_2),\end{aligned}$$

due to our assumptions in (a) it follows

$$\begin{aligned}&\mathbb{P}(X_1 \le X_2 \le X_3, Y_1 \le Y_2 \le Y_3)\\ &\quad= \int_{\mathbb{R}^2} \int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3}{Y_3}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)\\ &\quad\le \int_{\mathbb{R}^2} \int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3^*}{Y_3^*}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1).\end{aligned}$$

Furthermore,

$$\overline{F}^{x_1,y_1}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}}(x_2,x_3,y_2,y_3) \le \overline{F}^{x_1,y_1}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}}(x_2,x_3,y_2,y_3)$$

with $I^c = \{2, 3\}$ yields

$$\overline{F}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2) = \overline{F}^{x_1,y_1}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}}(x_2,-\infty,y_2,-\infty)$$
$$\leq \overline{F}^{x_1,y_1}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}}(x_2,-\infty,y_2,-\infty)$$
$$= \overline{F}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)$$

such that

$$\int \mathbb{1}_{[a,\infty[\times[b,\infty[}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2) \leq \int \mathbb{1}_{[a,\infty[\times[b,\infty[}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)$$

for all $a,b \in \mathbb{R}$. Conditional survival functions are indeed survival functions, which can be approximated by sums of indicator functions of the form above. Hence, by the bounded convergence theorem it holds

$$\int_{\mathbb{R}^2}\int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3^*}{Y_3^*}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2}{Y_2}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)$$
$$\leq \int_{\mathbb{R}^2}\int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3^*}{Y_3^*}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)$$
$$= \mathbb{P}(X_1 \leq X_2^* \leq X_3^*, Y_1 \leq Y_2^* \leq Y_3^*).$$

By defining

$$\widetilde{\mathbb{P}}^{x_3,y_3}(A) := \mathbb{P}(A \mid X_3^* = x_3, Y_3^* = y_3)$$

for any event A, in an analogous way it holds

$$\mathbb{P}(X_1 \leq X_2^* \leq X_3^*, Y_1 \leq Y_2^* \leq Y_3^*)$$
$$= \int_{\mathbb{R}^2} \widetilde{\mathbb{P}}^{x_3,y_3}(X_1 \leq X_2^* \leq x_3, Y_1 \leq Y_2^* \leq y_3)\, d\mathbb{P}_{\binom{X_3^*}{Y_3^*}}(x_3,y_3)$$
$$= \int_{\mathbb{R}^2}\int_{]-\infty,x_3]\times]-\infty,y_3]} \widetilde{\mathbb{P}}^{x_3,y_3}(X_1 \leq x_2, Y_1 \leq y_2 \mid X_2^* = x_2, Y_2^* = y_2)$$
$$d\widetilde{\mathbb{P}}^{x_3,y_3}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_3^*}{Y_3^*}}(x_3,y_3)$$
$$= \int_{\mathbb{R}^2}\int_{]-\infty,x_3]\times]-\infty,y_3]} F^{x_2,x_3,y_2,y_3}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c}\Big|\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}(x_2,y_2)\, d\widetilde{\mathbb{P}}^{x_3,y_3}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_3^*}{Y_3^*}}(x_3,y_3)$$
$$\leq \int_{\mathbb{R}^2}\int_{]-\infty,x_3]\times]-\infty,y_3]} F^{x_2,x_3,y_2,y_3}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c}\Big|\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}(x_2,y_2)\, d\widetilde{\mathbb{P}}^{x_3,y_3}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_3^*}{Y_3^*}}(x_3,y_3)$$
$$= \mathbb{P}(X_1^* \leq X_2^* \leq X_3^*, Y_1^* \leq Y_2^* \leq Y_3^*)$$

with $I = \{2,3\}$ and $I^c = \{1\}$, accordingly. Under the alternative assumption (b), the function

$$\overline{H}(x_1,y_1) := \int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3^*}{Y_3^*}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)$$

can be, at least up to scaling, considered as a survival function, so by using an approximation by sums of indicator functions as above and Inequality (5.3) it follows

$$\begin{aligned}
&\mathbb{P}(X_1 \le X_2 \le X_3, Y_1 \le Y_2 \le Y_3)\\
&\qquad \le \int_{\mathbb{R}^2} \int_{[x_1,\infty[\times[y_1,\infty[} \overline{F}^{x_1,x_2,y_1,y_2}_{\binom{X_3^*}{Y_3^*}}(x_2,y_2)\, d\mathbb{P}^{x_1,y_1}_{\binom{X_2^*}{Y_2^*}}(x_2,y_2)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)\\
&\qquad = \int_{\mathbb{R}^2} \overline{H}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1}{Y_1}}(x_1,y_1)\\
&\qquad \le \int_{\mathbb{R}^2} \overline{H}(x_1,y_1)\, d\mathbb{P}_{\binom{X_1^*}{Y_1^*}}(x_1,y_1)\\
&\qquad = \mathbb{P}(X_1^* \le X_2^* \le X_3^*, Y_1^* \le Y_2^* \le Y_3^*).
\end{aligned}$$

□

The assumptions on the conditional cdfs and conditional survival functions seem quite restrictive. Nevertheless, there are still natural examples of random vectors which satisfy these as we illustrate in the following.

Example 5.3.2. Let $(\mathbf{X},\mathbf{Y})$ and $(\mathbf{X}^*,\mathbf{Y}^*)$ be 4-dimensional random vectors defined on a common probability space with

$$\mathbf{X} = (X_1, X_2), \quad \mathbf{X}^* = (X_1^*, X_2), \quad \mathbf{Y} = (Y_1, Y_2), \quad \mathbf{Y}^* = (Y_1^*, Y_2),$$

and $\mathbf{X} \overset{D}{=} \mathbf{X}^*$ and $\mathbf{Y} \overset{D}{=} \mathbf{Y}^*$, i.e., in particular it holds $X_1 \overset{D}{=} X_1^*$ and $Y_1 \overset{D}{=} Y_1^*$. Furthermore, it is enough to suppose that (X_1, Y_1) and (X_1^*, Y_1^*) are both independent of (X_2, Y_2), respectively. Therefore, it is sufficient to only consider the distributions of (X_1, Y_1) and (X_1^*, Y_1^*), respectively.

Again, we start with a construction of a discrete example in order to extend it to the continuous case afterwards in a natural way. Suppose the probability distributions of (X_1, Y_1) and (X_1^*, Y_1^*) are given by

$$\begin{aligned}
\mathbb{P}((X_1,Y_1) = (1,3)) = \mathbb{P}((X_1,Y_1) = (2,2)) &= \frac{1}{2}\\
\mathbb{P}((X_1^*,Y_1^*) = (1,2)) = \mathbb{P}((X_1^*,Y_1^*) = (2,3)) &= \frac{1}{2}.
\end{aligned}$$

In particular, we observe that $X_1 \overset{D}{=} X_1^*$ and $Y_1 \overset{D}{=} Y_1^*$. (For an overview, see Fig. 5.4.) The cdfs as well as the survival functions are outlined in Tab. 5.2. We observe

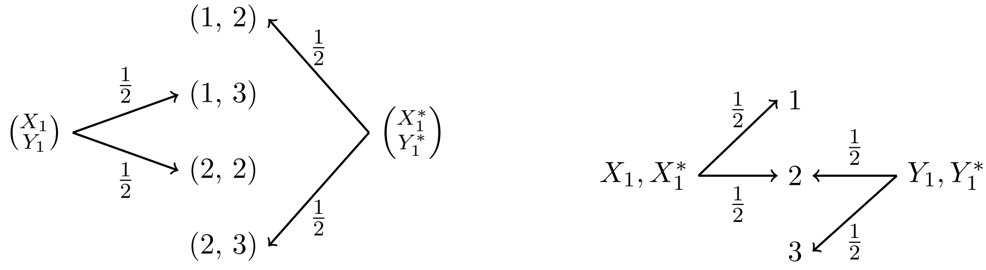
$$F_{\binom{X_1}{Y_1}} \le F_{\binom{X_1^*}{Y_1^*}} \quad \text{and} \quad \overline{F}_{\binom{X_1}{Y_1}} \le \overline{F}_{\binom{X_1^*}{Y_1^*}}$$

which implies

$$F_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \le F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \quad \text{and} \quad \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}} \le \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}}{\mathbf{Y}}^{I}}$$

as well as

$$F_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \le F_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \quad \text{and} \quad \overline{F}_{\binom{\mathbf{X}}{\mathbf{Y}}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}} \le \overline{F}_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I^c} \big| \binom{\mathbf{X}^*}{\mathbf{Y}^*}^{I}}$$



(i) **(ii)**

Figure 5.4: Illustration of the respective pmfs: joint pmfs in (i) and pmfs of the respective marginals in (ii). Here, the arrows imply that the random vector of interest attains the respective outcome with the indicated probability (of 1/2).

Table 5.2: Cdfs and survival functions of (X_1, Y_1) and (X_1^*, Y_1^*), respectively.

(x_1, y_1)	$F_{\binom{X_1}{Y_1}}$	$F_{\binom{X_1^*}{Y_1^*}}$	$\overline{F}_{\binom{X_1}{Y_1}}$	$\overline{F}_{\binom{X_1^*}{Y_1^*}}$
(1, 2)	0	$\frac{1}{2}$	1	1
(1, 3)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 3)	1	1	0	$\frac{1}{2}$

for all $I \subsetneq \{1, 2\}$ and $I^c = \{1, 2\} \setminus I$ due to the assumed independence, therefore, both sets of assumptions in Theorem 5.3.1 are satisfied.

All that is left is to extend the random vectors to the continuous case. For this, let $f_{\binom{X_2}{Y_2}}$ denote the probability density function of (X_2, Y_2), and define the densities f and f^* of (X_1, Y_1) and (X_1^*, Y_1^*), respectively, by

$$f(x_1, y_1) := f_{\binom{X_1}{Y_1}}(x_1, y_1) = \frac{1}{2}\left(\mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]2,3[}(y_1) + \mathbb{1}_{]1,2[}(x_1) \cdot \mathbb{1}_{]1,2[}(y_1)\right)$$

$$f^*(x_1, y_1) := f_{\binom{X_1^*}{Y_1^*}}(x_1, y_1) = \frac{1}{2}\left(\mathbb{1}_{]0,1[}(x_1) \cdot \mathbb{1}_{]1,2[}(y_1) + \mathbb{1}_{]1,2[}(x_1) \cdot \mathbb{1}_{]2,3[}(y_1)\right).$$

Let (X_1, Y_1) and (X_1^*, Y_1^*) be independent of (X_2, Y_2), respectively, such that the densities of $(\mathbf{X}, \mathbf{Y})$ and $(\mathbf{X}^*, \mathbf{Y}^*)$ are given by

$$f_{\binom{\mathbf{X}}{\mathbf{Y}}}(x_1, x_2, y_1, y_2) = f_{\binom{X_1}{Y_1}}(x_1, y_1) \cdot f_{\binom{X_2}{Y_2}}(x_2, y_2)$$

$$f_{\binom{\mathbf{X}^*}{\mathbf{Y}^*}}(x_1, x_2, y_1, y_2) = f_{\binom{X_1^*}{Y_1^*}}(x_1, y_1) \cdot f_{\binom{X_2}{Y_2}}(x_2, y_2).$$

Obviously, it holds $X_1 \overset{D}{=} X_1^*$ and $Y_1 \overset{D}{=} Y_1^*$. Moreover, it holds $F_{\binom{X_1}{Y_1}} \leq F_{\binom{X_1^*}{Y_1^*}}$ and $\overline{F}_{\binom{X_1}{Y_1}} \leq \overline{F}_{\binom{X_1^*}{Y_1^*}}$ (see Fig. 5.5). Therefore, again both sets of assumptions in Theorem 5.3.1 are satisfied.

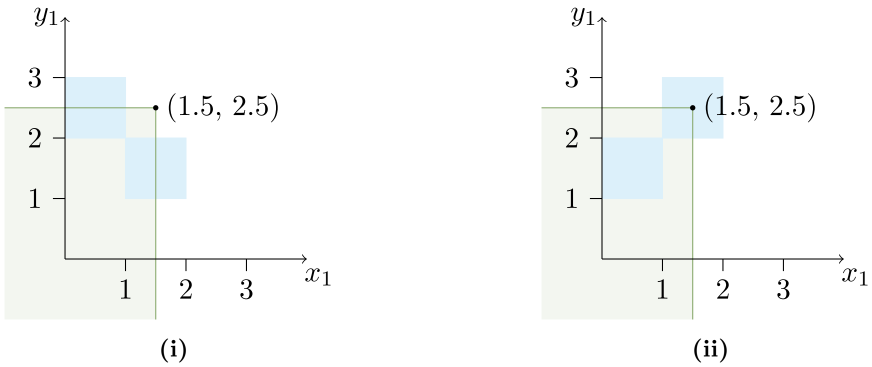


Figure 5.5: Examplary illustration of the density functions f and f^*: The blue squares denote the places where the respective functions hold mass. Considering the exemplary rectangle $]-\infty, 1.5] \times]-\infty, 2.5]$ (green) we observe that f has less cumulative mass than f^*. This holds for any point $(x_1, y_1) \in \mathbb{R}^2$ as well as for rectangles of the form $[x_1, \infty[\times [y_1, \infty[\subset \mathbb{R}^2$.

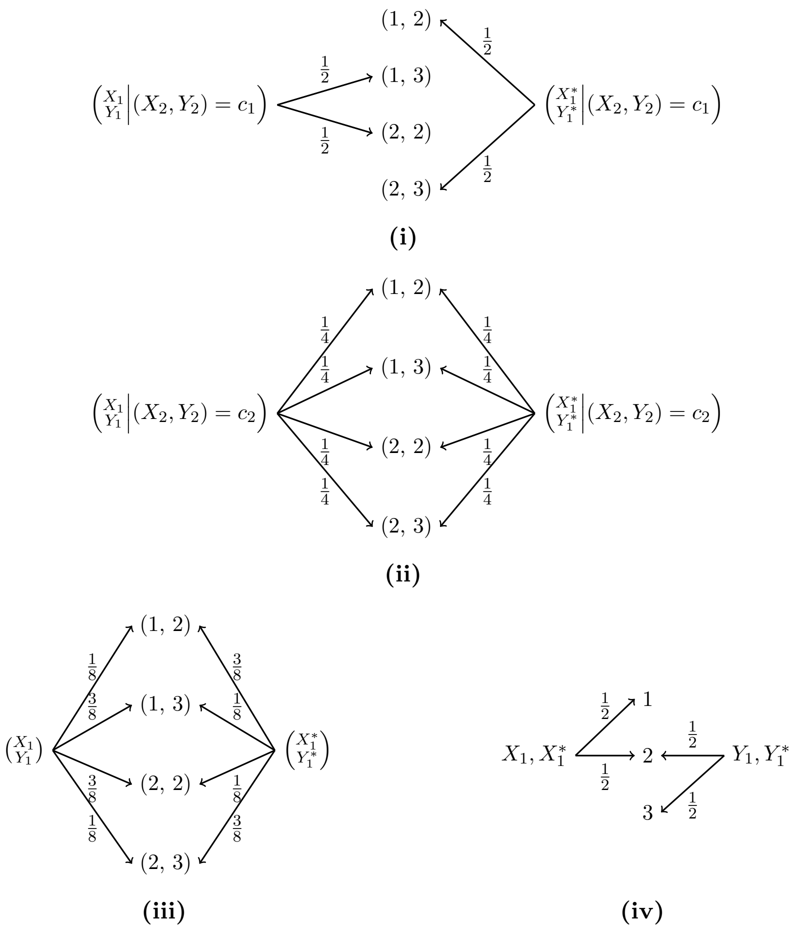


Figure 5.6: Illustration of the respective pmfs: conditional pmfs in (i) and (ii), joint pmfs in (iii) and pmfs of the respective marginals in (iv). The arrows imply that the random vector of interest attains the respective outcome with the indicated probability.

Example 5.3.3. In the previous example we assumed independence of (X_1, Y_1) and (X_1^*, Y_1^*) with regard to (X_2, Y_2). Let us also present an example with dependent random vectors. For this, assume that $(X_2, Y_2) \in \{c_1, c_2\}$ and

$$\mathbb{P}((X_2, Y_2) = c_1) = \frac{1}{2} = \mathbb{P}((X_2, Y_2) = c_2).$$

We consider the dependence structure as illustrated in Fig. 5.6 (i) and (ii). The respective conditional cdfs and conditional survival functions are outlined in Tables 5.3 (i) and (ii). We observe that the conditions on the conditional cdfs and survival functions required by the first set of assumptions are satisfied. Moreover, even the conditions on the joint distributions which are necessary for the alternative set of assumptions are fulfilled, as Fig. 5.6 (iii) and Table 5.3 (iii) indicate. Extending this to the continuous case in an analogous manner as we have done before concludes this example.

Table 5.3: (Conditional) cdfs and survival functions of (X_1, Y_1) and (X_1^*, Y_1^*), respectively. Here, $F^c_{\binom{X_1}{Y_1}}$ denotes the cdf of $(X_1, Y_1 | (X_2, Y_2) = c)$. Note that all of the desired inequalities are satisfied.

(x_1, y_1)	$F^{c_1}_{\binom{X_1}{Y_1}}$	$F^{c_1}_{\binom{X_1^*}{Y_1^*}}$	$\overline{F}^{c_1}_{\binom{X_1}{Y_1}}$	$\overline{F}^{c_1}_{\binom{X_1^*}{Y_1^*}}$
(1, 2)	0	$\frac{1}{2}$	1	1
(1, 3)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 3)	1	1	0	$\frac{1}{2}$

(i)

(x_1, y_1)	$F^{c_2}_{\binom{X_1}{Y_1}}$	$F^{c_2}_{\binom{X_1^*}{Y_1^*}}$	$\overline{F}^{c_2}_{\binom{X_1}{Y_1}}$	$\overline{F}^{c_2}_{\binom{X_1^*}{Y_1^*}}$
(1, 2)	$\frac{1}{4}$	$\frac{1}{4}$	1	1
(1, 3)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 3)	1	1	$\frac{1}{4}$	$\frac{1}{4}$

(ii)

(x_1, y_1)	$F_{\binom{X_1}{Y_1}}$	$F_{\binom{X_1^*}{Y_1^*}}$	$\overline{F}_{\binom{X_1}{Y_1}}$	$\overline{F}_{\binom{X_1^*}{Y_1^*}}$
(1, 2)	$\frac{1}{8}$	$\frac{3}{8}$	1	1
(1, 3)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
(2, 3)	1	1	$\frac{1}{8}$	$\frac{3}{8}$

(iii)

5.4 Other Stochastic Order Relations

By means of their fifth axiom, Grothe et al. [32] attempted to make a statement of the form: 'If one pair of random vectors is more interdependent than another pair, then the measure of dependence should reflect this'. Inspired by Schmid et al. [69], they used multivariate concordance ordering for this, which is defined by the ordering of the respective multivariate cdfs and survival functions. Intuitively this means that random vectors are concordant if their components tend to be either "all large together or all small together" [38, p. 1]. The question remains whether this is the most natural characterization of dependence. In fact, one might argue that it is more natural to consider such a behavior in a bivariate sense instead of in all dimensions simultaneously, that is, two random vectors are more dependent if they show the above tendency just componentwise. This leads us to so-called supermodular ordering.

Definition 5.4.1 (Müller and Scarsini [60])**.** A function $g : \mathbb{R}^d \to \mathbb{R}$ is called *supermodular* if

$$g(\mathbf{x} \wedge \mathbf{y}) + g(\mathbf{x} \vee \mathbf{y}) \geq g(\mathbf{x}) + g(\mathbf{y}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$$

where the lattice operators $\wedge$ and $\vee$ are defined as

$$\mathbf{x} \wedge \mathbf{y} := (\min\{x_1, y_1\}, \ldots, \min\{x_d, y_d\})$$

and

$$\mathbf{x} \vee \mathbf{y} := (\max\{x_1, y_1\}, \ldots, \max\{x_d, y_d\}).$$

For an equivalent way to define supermodular functions in terms of difference operators, we refer to Müller and Scarsini [60].

Definition 5.4.2 (Müller and Scarsini [60, Definition 2.3])**.** A random vector $\mathbf{V} = (V_1, \ldots, V_d)$ is said to be smaller than a random vector $\mathbf{W} = (W_1, \ldots, W_d)$ in the *supermodular order* (or *order by L-superadditive functions*), written $\mathbf{V} \preccurlyeq_{sm} \mathbf{W}$, if $\mathbb{E}g(\mathbf{V}) \leq \mathbb{E}g(\mathbf{W})$ for all supermodular functions g such that the expectations exist.

Both supermodular and concordance ordering fulfill the nine axioms for multivariate positive dependence orders as proposed by Joe [39, p. 38ff] (cf. Müller and Scarsini [60]). It is well-known that supermodular order implies concordance order [60]. Let us remark in this context that concordance order has been proposed by Joe [38] recognizing that it is a generalization of supermodular order. The author argues for the use of the weaker property as it can be checked more easily in practice. Equivalence of these stochastic order relations holds only in case of $d = 2$, while supermodular ordering is strictly stronger for the cases $d \geq 3$ (see Joe [38, Example A.1.3] for $d > 3$, and Müller and Scarsini [60, Theorem 2.6], who have complemented it with the case $d = 3$).

Therefore, it might be interesting to check whether OPD respects supermodular ordering. The following example shows that it does not:

Example 5.4.3. We consider again the discrete distributions from Example 5.2.1, which were illustrated in Fig. 5.1, and in order to simplify the notation we write $\mathbf{V} = (\mathbf{X}, \mathbf{Y})$ and $\mathbf{W} = (\mathbf{X}^*, \mathbf{Y}^*)$. Let g denote a supermodular function. Then it holds

$$\mathbb{E}g(\mathbf{V}) = \frac{1}{2} \cdot \left(g(\underbrace{1,2,3,4}_{=\mathbf{a}}) + g(\underbrace{3,2,3,2}_{=\mathbf{b}}) \right) \leq \frac{1}{2} \cdot \left(g(\underbrace{1,2,3,2}_{=\mathbf{a}\wedge\mathbf{b}}) + g(\underbrace{3,2,3,4}_{=\mathbf{a}\vee\mathbf{b}}) \right) = \mathbb{E}g(\mathbf{W}),$$

so $\mathbf{V} \preccurlyeq_{sm} \mathbf{W}$. However, according to Example 5.2.1 it holds $\mathrm{OPD}_2(\mathbf{X}, \mathbf{Y}) > \mathrm{OPD}_2(\mathbf{X}^*, \mathbf{Y}^*)$, so OPD does not respect supermodular ordering.

In an analogous way, the condition on supermodular ordering can be verified for the refined Counterexamples 5.2.2–5.2.3. We only show the proof of the latter, since this is the more involved case.

Example 5.4.4. We reconsider the probability distribution from Example 5.2.3. Again we simplify the notation by $\mathbf{V} = (\mathbf{X}, \mathbf{Y})$ and $\mathbf{W} = (\mathbf{X}^*, \mathbf{Y}^*)$. Let g denote a supermodular function. Then it holds

$$\begin{aligned}
\mathbb{E}g(\mathbf{V}) = \int g(x_1, x_2, y_1, y_2) \cdot \Big(&\mathbb{1}_{\{0 \le x_1 \le x_2 \le 1\}} \cdot \mathbb{1}_{\{1 < y_1 \le y_2 \le 2\}} \\
&+ \mathbb{1}_{\{1 < x_1 \le x_2 \le 2\}} \cdot \mathbb{1}_{\{0 \le y_1 \le y_2 \le 1\}}\Big)\, d(x_1, x_2, y_1, y_2) \\
+ \int g(x_1, x_2, y_1, y_2) \cdot \Big(&\mathbb{1}_{\{0 \le x_2 < x_1 \le 1\}} \cdot \mathbb{1}_{\{1 < y_2 < y_1 \le 2\}} \\
&+ \mathbb{1}_{\{1 < x_2 < x_1 \le 2\}} \cdot \mathbb{1}_{\{0 \le y_2 < y_1 \le 1\}}\Big)\, d(x_1, x_2, y_1, y_2) \\
\le \int g(x_1, x_2, y_1, y_2) \cdot \Big(&\mathbb{1}_{\{0 \le x_1 \le x_2 \le 1\}} \cdot \mathbb{1}_{\{0 \le y_1 \le y_2 \le 1\}} \\
&+ \mathbb{1}_{\{1 < x_1 \le x_2 \le 2\}} \cdot \mathbb{1}_{\{1 < y_1 \le y_2 \le 2\}}\Big)\, d(x_1, x_2, y_1, y_2) \\
+ \int g(x_1, x_2, y_1, y_2) \cdot \Big(&\mathbb{1}_{\{0 \le x_2 < x_1 \le 1\}} \cdot \mathbb{1}_{\{0 \le y_2 < y_1 \le 1\}} \\
&+ \mathbb{1}_{\{1 < x_2 < x_1 \le 2\}} \cdot \mathbb{1}_{\{1 < y_2 < y_1 \le 2\}}\Big)\, d(x_1, x_2, y_1, y_2) \\
= \mathbb{E}g(\mathbf{W}),&
\end{aligned}$$

where we used the supermodularity of g twice. Note that for all

$$\begin{aligned}
\mathbf{a} &\in \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : 0 \le x_1 \le x_2 \le 1,\ 1 < y_1 \le y_2 \le 2\} \\
\mathbf{b} &\in \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : 1 \le x_1 \le x_2 \le 2,\ 0 \le y_1 \le y_2 \le 1\},
\end{aligned}$$

it follows

$$\begin{aligned}
\mathbf{a} \wedge \mathbf{b} &\in \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : 0 \le x_1 \le x_2 \le 1,\ 0 \le y_1 \le y_2 \le 1\} \\
\mathbf{a} \vee \mathbf{b} &\in \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : 1 \le x_1 \le x_2 \le 2,\ 1 < y_1 \le y_2 \le 2\}.
\end{aligned}$$

Similar applies to the indicators of the second integral. Hence, $\mathbf{V}$ is smaller than $\mathbf{W}$ in the supermodular order, but Example 5.2.3 yields $\mathrm{OPD}_2(\mathbf{X}, \mathbf{Y}) > \mathrm{OPD}_2(\mathbf{X}^*, \mathbf{Y}^*)$.

Thus, there are stationary continuously distributed examples such that OPD does not respect supermodular order either. Although humbling, these results make sense at second glance. Concordance order as well as supermodular order are defined on the level of the concrete values, while OPD operates on the level of ordinal patterns, that is, on the level of the ordinal relations between those values. Even if values have a greater tendency to be large or small together in one pair of random vectors if compared to another vector (either dimensionwise or all together), a destruction of the prevalent ordinal relations cannot be ruled out.

5.5 Quantifying Dependence: A Critical Look at Axiomatic Approaches

People have been interested in relevant criteria for distinguishing measures of dependence. In this regard, Grothe et al. [32] proposed an axiomatic framework for multivariate measures of association between random vectors (see Def. 5.1.3), to which Betken et al. [17] referred under the notion of multivariate measures of dependence. The ideas of Grothe et al. [32] were based on the work of Schmid et al. [69], who gave an overview on existing criteria for copula-based multivariate measures of association. Multivariate measures of concordance were also discussed by them. Those are particularly characterized by the fact that they satisfy concordance ordering (Axiom 5). Intuitively speaking this means that the more the random variables tend to be large or small together, the larger the value of the measure at hand should be (cf., e.g., [38]). Hence, this is a statement of the form 'the more dependent a pair of random vectors $(\mathbf{X}, \mathbf{Y})$ is, the larger the value of the measure applied to this pair should be', where dependence is characterized by concordance in this case.

Obviously, the purpose of measures of dependence is to actually measure dependence. However, in order to do so, first the term 'dependence' needs to be defined. Even though dependence is such a central concept in statistics, it is rarely defined formally, and if it is, then it is equated to non-independence (see Geenens [30] and the references mentioned therein). This would make dependence a binary concept: Either random variables are dependent or not, without room for nuance [30, p. 2]. On the other hand, various measures (e.g., (multivariate) Kendall's τ, Spearman's ρ_S and Pearson's correlation, OPD, etc.) have been and are still being developed in order to somehow quantify dependence. Hence, dependence seems to be quantifiable, which proves the binary definition to be inadequate (cf. [30, p. 2]).

Geenens [30] has recently proposed a general definition of dependence between two random variables defined on the same probability space. The author defines dependence between such random variables as the information which is necessary and sufficient to unambiguously identify their joint distribution in its Fréchet class, that is, the class of distributions with given marginals. Furthermore, Geenens argues that most of the proposed dependence measures, or at least the classical ones, do not actually measure dependence. The author writes as follows [30, p. 28]:

> Most of statistical theory developed from the initial works of Galton and Pearson on the Gaussian distribution (Stigler, 2002). As a result, much of the statistical jargon in use today retains a strong Gaussian connotation. For example, the fact that the dependence reduces down to the parameter ρ in a bivariate Gaussian distribution [...] largely explains why the correlation coefficient has been referred to as a 'dependence measure' universally, including outside the Gaussian model. [...] This has caused, through a similar process of conflation, most of those relations to be indiscriminately labelled as 'dependence'.

And furthermore with a special focus on concordance [30, p. 29]:

> That the two concepts [correlation and concordance] coincide in the Gaussian case may explain why concordance is commonly understood as 'some form' of dependence. For example, the term 'monotone dependence' ('positive'/'negative', for concordance/discordance) is often used, while Spearman's ρ_S and Kendall's

τ are routinely referred to as 'dependence measures', in spite of specifically accounting for concordance.

It becomes clear that Geenens has a rigorous understanding of dependence and dependence measures in particular. The author states that a dependence measure should attain the value 0 if and only if the random variables under consideration are independent. (Note that Axiom 4 of Definition 5.1.3 constitutes a less strict version, that is, independence $\Longrightarrow$ value 0.) As a matter of fact, e.g., concordance/discordance indicates an influence of one variable on another. But neither does the absence of concordance/discordance imply independence, nor do different levels of concordance implicate gradations in the degree of dependence (cf. [30, Section 5]). Therefore, the author takes the view that "there is in general no link between concordance and dependence" [30, p. 30].

We take a different, less strict perspective. That is, we think of different types of dependence, e.g., linear dependence, concordance/discordance, association, ordinal dependence, to name but a few; cf. Section 3.10 and Fig. 3.2 of Müller and Stoyan [61] for an overview on (the relations between) various concepts of dependence. In accordance, the proposed measures focus on the respective qualities, e.g., Pearson's correlation (as well as its multivariate generalizations) measures *linear* dependence. It becomes clear that each of those has advantages and disadvantages compared to the others with regard to different contexts. Moreover, it is not a contradiction if a measure of concordance is zero for non-concordant, but dependent random variables. However, we agree with the author that these terms should be used more precisely in the literature in general. Note that there *are*, of course, communities of experts in this field who already do so.

Concordance is clearly not the type of dependence that is measured by OPD, so Axiom 5 is not well-suited in this case. That raises the question whether it is a good idea at all to propose *one* axiom which has to be satisfied by *all* dependence concepts. Instead it might be the better choice to directly distinguish between the respective concepts. Therefore, a direction for future research might be to develop a stochastic order relation on the ordinal level suited for measures based on ordinal patterns. It might be interesting to see whether this can be done in such a way that it fulfills the properties of a multivariate dependence order as suggested by Joe [39, Section 2.2.3] (or at least most of them as, e.g., bivariate concordance arguably does not make so much sense in terms of ordinal patterns). To our knowledge, this has not been done yet.

At this point, we would also like to draw the reader's critical eye to another issue: OPD is based on the comparison of the ordinal relations within both vectors. In fact, it is a special comparison, since only identical and, depending on the chosen variant of OPD, opposite (= space-inversed) patterns are considered. In general, it is also possible to propose similar measures by considering other types of ordinal pattern relations. Because what does dependence between two time series mean in terms of their ordinal patterns? It means that the more frequently specific ordinal patterns in one time series are coupled with specific (possibly different) ordinal patterns in the other, the more dependent are these time series. It should be noted, however, that we obtain ordinal patterns in a sliding window approach. So do bijective maps $\sigma : S_d \to S_d$ exist which are not the identity or the spatial inversion such that this overlapping structure is preserved? The answer is yes, at least for $d \geq 3$. (Recall that there are only two patterns for $d = 2$ such that the identity and spatial inversion are the only possibilities in this case.)

For illustration consider the following example for $d = 3$: Suppose the increasing pattern (1, 2, 3) is directly followed by the pattern (1, 3, 2), that is, they have an overlap of length 2. Then mapping $(1, 2, 3) \mapsto (3, 2, 1)$ and $(1, 3, 2) \mapsto (2, 3, 1)$ does not respect the overlapping structure, since (3, 2, 1) cannot be directly followed by (2, 3, 1). On the contrary, the mapping $(1, 2, 3) \mapsto (3, 2, 1)$ and $(1, 3, 2) \mapsto (2, 1, 3)$ works with the sliding window approach. Note that when constructing such a bijection, it is crucial to consider all possible combinations of subsequent overlapping patterns. Hence, the above just serves as a simplified example.

Even though such a bijection between two time series indicates perfect ordinal dependence, OPD still might be equal to zero or at least close to zero (again depending on the considered version of OPD). On one hand it is a serious drawback that OPD does not account for such dependencies. On the other hand, they seem to be kind of artificial, that is, in practice this is not the kind of dependence one would intuitively think of. Hence, another direction for future research might be to propose variants or generalizations of OPD and perform a comparative analysis in terms of advantages and disadvantages of the respective measures.

6 Some Multivariate Extensions of Ordinal Patterns

Within the scope of this dissertation, so far we have only considered ordinal patterns stemming from a one-dimensional time series. With regard to those, we have discussed the estimation of the SCI, and hence also the estimation of the Rényi-2 permutation entropy. These allow for the quantification of the complexity in a time series, assuming that the time series under consideration exhibits some form of short-range dependence. Furthermore, we have investigated OPD in the context of the axiomatic framework for multivariate measures of dependence as proposed by Grothe et al. [32]. Note that OPD is still defined in terms of the ordinal patterns of the respective univariate time series.

In what follows, we consider two multivariate extensions of ordinal patterns, namely *multivariate ordinal patterns* and *spatial ordinal patterns*. These have been recently introduced by Mohr et al. [58] and Bandt and Wittfeld [12], respectively (recall the historical background in Chapter 1.1.5). We discuss different possibilities of representing those as well as their respective advantages and disadvantages with a special focus on digital implementation. In addition, the representations are also briefly considered in terms of the occurrence of ties.

This chapter builds vaguely on the preliminaries of the joint work [82] with C. H. Weiß and A. Schnurr. This applies in particular to the part regarding the partition of the set of spatial ordinal patterns. However, this chapter contains additional examples and further thoughts which complement our discussion of univariate ordinal patterns in Chapter 3.

The subsequent chapter continues with a comparative analysis of various methods for testing for dependence between two time series consisting of serially independent and identically distributed random variables based on multivariate extensions of ordinal patterns. In particular, OPD is placed in the context of multivariate ordinal patterns.

6.1 Multivariate Ordinal Patterns

Let $m \geq 2$ denote the dimension of a multivariate time series. Mohr et al. [58] defined multivariate ordinal patterns as the vector consisting of the univariate ordinal patterns with respect to each dimension:

Definition 6.1.1 (Mohr et al. [58, Definition 3])**.** Let $d \geq 2$. A matrix

$$\begin{pmatrix} x_{11} & \dots & x_{d1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \dots & x_{dm} \end{pmatrix} \in \mathbb{R}^{m \times d}$$

has the *multivariate ordinal pattern (MOP)* $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^\top \in S_d^m$, where π_i denotes the univariate ordinal patterns of $(x_{1i}, \dots, x_{di})$, $i = 1, \dots, m$.

As a consequence, $(d!)^m$ MOPs are possible.

Table 6.1: Enumeration of MOPs with $m = 2$.

$\pi_1 \backslash \pi_2$	0	1	2	$\dots$	$d! - 1$
0	0	1	2	$\dots$	$d! - 1$
1	$d!$	$d! + 1$	$d! + 2$	$\dots$	$2d! - 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$d! - 1$	$(d!)^2 - d!$	$(d!)^2 - (d! - 1)$	$(d!)^2 - (d! - 2)$	$\dots$	$(d!)^2 - 1$

Note that here columns are indicated before rows. This is due to the fact that we are going to consider such matrices stemming from bivariate time series $(\mathbf{X}_t)_t = (X_{t,1}, X_{t,2})_t$ where the time component t runs columnwise, i.e., from left to right, in particular.

As the attentive reader may have noticed, the above definition is independent of the chosen representation of univariate ordinal patterns presented in Chapter 3. This leads to MOPs being represented by matrices or, in case of number representations, vectors, as it is illustrated by the subsequent example:

Example 6.1.2. Consider the (2×3)-dimensional matrix $\mathbf{x} \in \mathbb{N}_0^{2\times 3}$ given by

$$\mathbf{x} = \begin{pmatrix} 2 & 5 & 4 \\ 1 & 7 & 3 \end{pmatrix}.$$

In order to determine its MOP, we need to obtain the univariate ordinal patterns with regard to each component. In this case, both components exhibit the same ordinal pattern given by $\pi = r = (1, 3, 2)$ or number representation $n_{LC} = 1$. Hence, the MOP of $\mathbf{x}$ is given by

$$\begin{pmatrix} 1 & 3 & 2 \\ 1 & 3 & 2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

depending on the choice of ordinal pattern representation.

As we have already discussed in Section 3.2, with regard to digital implementation of ordinal patterns, a representation in terms of one number is preferable to a vector (or even matrix). Therefore, the question arises whether this kind of representation is also possible for MOPs. We suggest the following way to do so for the bivariate case $m = 2$:

For MOPs of length d, there are $d!$ possible univariate ordinal patterns for both components, respectively. With regard to number representations, this leads to the integers $0, 1, 2, \dots, d! - 1$ as possible representations. So how can we number the $(d!)^2$ MOPs consecutively without the need for a look-up table? A simple way is to proceed row by row. For the pattern represented by '0' in the first component, the MOPs are numbered from 0 to $d! - 1$ depending on the ordinal pattern in the second component, while the MOPs range from $d!$ to $2d! - 1$ for the univariate ordinal pattern '1' in the first component, and so on. This yields the formula

$$\boldsymbol{\pi} = d! \cdot \pi_1 + \pi_2 \tag{6.1}$$

for computing the suggested number representation for MOPs where π_1 and π_2 denote the number representations of the respective univariate components. An overview of this enumeration is given in Tab. 6.1.

Example 6.1.3. Consider the matrix $\mathbf{x} \in \mathbb{N}^{2\times 3}$ from Example 6.1.2. We have seen that its MOP represented by univariate number representations is given by $(1,1)^\top$. By application of Eq. (6.1) we obtain the MOP number representation $\boldsymbol{\pi} = 3! \cdot 1 + 1 = 7$.

For $m \geq 2$, Eq. (6.1) can be generalized to

$$\boldsymbol{\pi} = (d!)^{m-1} \cdot \pi_1 + \cdots + d! \cdot \pi_{m-1} + \pi_m. \tag{6.2}$$

As MOPs consist of the rowwise univariate ordinal patterns, our discussion and guideline on (univariate) representations remains valid. This includes the consideration of ties.

6.2 Spatial Ordinal Patterns

Bandt and Wittfeld [12] proposed an extension of univariate ordinal patterns to the multivariate case by considering all entries of a matrix 'as a whole' and determining its 'ordinal pattern'.

Definition 6.2.1 (Bandt and Wittfeld [12])**.** Let $d \geq 2$. A matrix

$$\begin{pmatrix} x_{11} & \dots & x_{d1} \\ x_{12} & \dots & x_{d2} \end{pmatrix} = \begin{pmatrix} w_1 & \dots & w_d \\ w_{d+1} & \dots & w_{2d} \end{pmatrix} \in \mathbb{R}^{2\times d}$$

(potentially stemming from a bivariate time series) has *spatial ordinal pattern (SOP)*

$$\boldsymbol{\Pi} := \begin{pmatrix} r_1 & \dots & r_d \\ r_{d+1} & \dots & r_{2d} \end{pmatrix}$$

if and only if

$$r_j < r_k \iff w_j < w_k$$

for all $1 \leq j < k \leq 2d$.

Note that the ordinal information between the components is preserved here. This is in contrast to MOPs, where only the ordinal information within the respective components is kept. Furthermore, note that here, we have implicitly set $m = 2$. Generalization to $m \geq 2$ is straightforward, but it is not recommendable due to the large number of SOPs which have to be considered resulting in large computation times: For $m = 2$, there are already $(2 \cdot d)!$ SOPs to be considered. This already leads to 24 possible SOPs for $d = 2$ and 720 for $d = 3$. In general, $(m \cdot d)!$ patterns have to be considered. So far, it is even common to consider $m = d = 2$ in practice, see Bandt and Wittfeld [12] and Weiß and Kim [88].

The above definition of SOPs is given in terms of ranks, therefore it can be regarded as a rank representation. Vectorization of SOPs, that is, considering a SOP as the univariate ordinal pattern $(r_1, \dots, r_d, r_{d+1}, \dots, r_{2d})$ of $(x_{11}, \dots, x_{d1}, x_{12}, \dots, x_{d2})$, allows for the application of our previous results: Number representations as discussed in Chapter 3 are applicable here and we especially emphasize the use of these with regard to digital implementation of SOPs, since vectors of length $2d$ (or $(2 \times d)$-dimensional matrices) are much more time-consuming with regard to testing for equality. Similar holds for their respective memory footprints.

Example 6.2.2. Consider the matrices $\mathbf{x}$ and $\mathbf{x}^*$ given by

$$\mathbf{x} = \begin{pmatrix} 2 & 5 \\ 1 & 7 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^* = \begin{pmatrix} 5 & 4 \\ 7 & 3 \end{pmatrix}.$$

Their respective SOPs are given by

$$\mathbf{\Pi} = \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{\Pi}^* = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}.$$

By vectorization they can also be represented by inversions, that is, $\eta = (1,1,0,0)$ and $\eta^* = (2,1,1,0)$. Hence, we obtain, e.g, the number representations $n_{LC} = 1\cdot(4-1)!+1\cdot(4-2)! = 8$ and $n^*_{LC} = 2\cdot(4-1)!+1\cdot(4-2)!+1\cdot(4-3)! = 15$.

However, apart from digital implementation we do not recommend expressing SOPs as vectors. This notation should only be utilized when a number representation is derived as it may be confusing for the reader to refer to SOPs being matrices by some vector. Therefore, we also do not recommend the use of permutations or inversions for the representation of SOPs due to the matrix-structure. Instead, we think it makes sense to stick with the rank representation.

In case of ties, the definition of SOPs can be slightly adjusted, e.g., in favor of increasing patterns, that is,

$$r_j < r_k \iff w_j < w_k \quad \text{or} \quad (w_j = w_k \text{ and } j < k).$$

At this point, let us emphasize that ties can be not only encountered componentwise, but also between components here. To illustrate this, consider the matrix $\mathbf{x}$ given by

$$\mathbf{x} = \begin{pmatrix} 4 & 3 \\ 6 & 4 \end{pmatrix}.$$

With regard to MOPs, there are no ties present such that we obtain $\boldsymbol{\pi} = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}$. With regard to its SOP, however, it holds $w_1 = w_4$, which then leads to the pattern $\mathbf{\Pi} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$.

Another possibility in case of ties is of course again adding small noise to the data. If many ties are present in the data, then both approaches lead to a distortion of the ordinal structure or a loss of information, as the presence ties also contains information. An interesting approach may be the generalization of SOPs in terms of generalized ordinal patterns (see Section 3.4). However, this is a point for future research, though the number of generalized SOPs and the associated potential impracticality have to be kept in mind.

Now, let $\mathcal{S}$ denote the set consisting of the SOPs of length d. For $1 \leq K \leq (2d)!$, suppose there is a partition of $\mathcal{S}$ into K non-empty subsets $\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_K$. These may either contain exactly one SOP each, i.e., $K = (2d)!$, or be defined according to some rule. This allows for a reduction of ordinal pattern probabilities which have to be considered and hence, leads to a reduction of the length of the time series needed in order to obtain representative probabilities.

Focusing on $d = 2$, Bandt and Wittfeld [12] proposed a partition into so-called 'types' of SOPs. Types imply a partition of $\mathcal{S}$ into the following three subsets:

$$\begin{aligned}
\mathcal{S}_1 &= \Big\{ \left(\begin{smallmatrix} \mathbf{1} & 2 \\ 3 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} \mathbf{1} & 3 \\ 2 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & \mathbf{1} \\ 4 & 3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 4 \\ \mathbf{1} & 3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 3 & \mathbf{1} \\ 4 & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 3 & 4 \\ \mathbf{1} & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 2 \\ 3 & \mathbf{1} \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 3 \\ 2 & \mathbf{1} \end{smallmatrix}\right) \Big\}, \\
\mathcal{S}_2 &= \Big\{ \left(\begin{smallmatrix} 1 & \mathbf{2} \\ 4 & 3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 4 \\ \mathbf{2} & 3 \end{smallmatrix}\right), \left(\begin{smallmatrix} \mathbf{2} & 1 \\ 3 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} \mathbf{2} & 3 \\ 1 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 3 & \mathbf{2} \\ 4 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 3 & 4 \\ \mathbf{2} & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 1 \\ 3 & \mathbf{2} \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 3 \\ 1 & \mathbf{2} \end{smallmatrix}\right) \Big\}, \qquad (6.3) \\
\mathcal{S}_3 &= \Big\{ \left(\begin{smallmatrix} 1 & \mathbf{3} \\ 4 & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 4 \\ \mathbf{3} & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & \mathbf{3} \\ 4 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 4 \\ \mathbf{3} & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} \mathbf{3} & 1 \\ 2 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} \mathbf{3} & 2 \\ 1 & 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 1 \\ 2 & \mathbf{3} \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 2 \\ 1 & \mathbf{3} \end{smallmatrix}\right) \Big\}.
\end{aligned}$$

As noted by [12], the first type corresponds to a monotonic behavior in both rows and columns, while the second type matches a monotonic behavior in either rows or columns and, hence, an antimonotonic behavior in the respective other direction. By contrast, in case of the third type, both lowest and both highest ranks appear either on the diagonal or the antidiagonal, respectively. Note that a SOP's type is equal to that rank number which shares a diagonal with the rank 4, see the bold ranks in (6.3).

Bandt and Wittfeld [12] used this classification to describe and distinguish textures in images. Shortly thereafter, Weiß and Kim [88] have not only considered SOPs but also this classification to test for (spatial) dependence (cf. Section 1.1.5).

Bandt and Wittfeld [12] have provided the following pseudo-algorithm, which can be used to calculate the type directly, either from the SOP $\mathbf{\Pi} = (r_1, r_2, r_3, r_4)$ or the local data $w = (w_1, w_2, w_3, w_4)$. For the former, substitute w by $\mathbf{\Pi}$ in the algorithm.

Algorithm 1 The algorithm for computation of the type of a 2×2-matrix provided by Bandt and Wittfeld [12].

$a = (w_1 < w_2) + (w_3 < w_4)$
if $a = 2$ **then**
    $a = 0$
end if
$b = (w_1 < w_3) + (w_2 < w_4)$
if $b = 2$ **then**
    $b = 0$
end if
type $= a + b + 1$

This means that it is once again possible to represent the ordinal structure of a 2×2-matrix by some number type $\in \{1, 2, 3\}$, which can be easily computed. A generalization to 'extended types' for $d > 2$ (or at least $d = 3$) has not yet been proposed in the literature.

6.3 Interim Conclusion

All in all, our recommendation on the representation of choice for the considered multivariate extensions of ordinal patterns remains the same as in the univariate case. That is, for mathematical considerations it is advantageous to opt for the rank representation as there the pattern can be read directly. In the context of ties, the classical three approaches discussed before are reasonable regarding data sets for which not many ties are expected. If, on the other hand, many ties are expected, consideration of generalized ordinal patterns in the definition of MOPs is preferable, though the set of possible generalized MOPs becomes quite large then. Consideration of generalized SOPs remains future research as it is not in the scope of this work.

Alternatively, the multivariate extensions can be represented by numbers, which is again particularly practical in the context of digital implementation. At this point, let us remark another option with regard to the indexing of MOP probabilities: Instead of using the number representation $n_{LC} + 1$ of the multivariate pattern as an index for the respective pattern probability $p_{n_{LC}+1}$, it can be more natural to use the number representations of the respective univariate ordinal patterns to indicate a joint pattern probability. Considering the MOP

$$\begin{pmatrix} 1 & 3 & 2 \\ 1 & 3 & 2 \end{pmatrix}$$

from Examples 6.1.2 and 6.1.3, its probability can be either represented by $p_{n_{LC}+1} = p_8$ or $p_{2,2}$, since it holds $n_{LC} = 1$ for the pattern $(1, 3, 2)$ and we use number representations of the form $n_{LC} + 1$. Note that in the case of univariate number representations of the form $n_{LC} + 1$, number representations for MOPs can be derived by $d! \cdot (\pi_1 - 1) + \pi_2$. Compare this with Eq. (6.1). The first variant p_8 is more compact in its notation, but as there are $(d!)^2$ MOPs, without a look-up table it can get confusing quite fast. In comparison, $d!$ univariate patterns are more manageable. Furthermore, consideration of the joint univariate pattern probabilities is necessary anyway in order to derive the MOP probabilities.

Another interesting question is which extension is preferable in practice. This question cannot be answered in general as the answer may depend on the context. It has been shown that SOPs perform well in the context of spatial dependence (see [12]). In fact, it has to be noted that SOPs preserve more ordinal information than MOPs, which also shows in the number of possible patterns. This is probably why they are better suited for the spatial context (as their name may already suggest).

However, to our knowledge the only comparative study of these extensions has been conducted in the context of non-parametric tests for dependence between times series (see the subsequent chapter, which is based on the joint work Silbernagel et al. [82]). There the performance of the proposed tests is additionally compared to some (classical) competitors.

7 Non-parametric Tests for Dependence Between Time Series based on Multivariate Extensions of Ordinal Patterns

In this chapter, we consider multivariate random variables being observed sequentially in time, which we denote as $(\mathbf{X}_t)_{t\in\mathbb{N}}$. Our primary interest here is the mutual pairwise dependence between the individual components, so without loss of generality, we assume that $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ is a *bivariate* process. We are interested in whether the components $X_{t,1}$ and $X_{t,2}$ are dependent or not, that is, we want to test the null hypothesis $\mathbb{H}_0$ that $X_{t,1}$ and $X_{t,2}$ are independent of each other against the alternative $\mathbb{H}_1$ that they are not:

$$\mathbb{H}_0: \ X_{t,1} \perp\!\!\!\perp X_{t,2} \qquad \text{vs.} \qquad \mathbb{H}_1: \ X_{t,1} \not\perp\!\!\!\perp X_{t,2}.$$

Moreover, we assume that $(\mathbf{X}_t)_{t\in\mathbb{N}}$ is an i.i.d. sequence of two-dimensional random vectors. Such kind of data do not only occur in multivariate time series analysis, see [51] for a comprehensive discussion, but also when monitoring multiple-stream processes. The latter refers to simultaneous measurements from several individual sources and is particularly relevant in manufacturing applications (e.g., parallel production lines or machines with several heads), see [59, Section 10.3] for details and references. Later in Section 7.4, for example, we shall be concerned with cascade process data and with product wafer data from a photolithographic process.

An obvious first idea for a corresponding hypothesis test is to rely on the classical Pearson correlation coefficient. This Pearson test, however, suffers from several drawbacks: It solely focuses on linear dependence, it is of parametric nature and actually assumes normally distributed data, and it is not robust against outliers, just to mention a few. Thus, the rank-based Spearman correlation might appear more attractive in practice, although this test is not an all-encompassing solution either. For example, it is only able to uncover monotone forms of dependence. For illustration, let us look at the three examples shown in Figure 7.1. Neither Pearson nor Spearman lead to a rejection of the null of independence, although we are faced with strong forms of cross-dependence in either case: quadratic dependence in (i), piecewise-linear dependence in (ii), and a quite appealing co-movement of the pairs being plotted in part (iii). The last example was derived from the R-package "datasauRus" as described in https://stats.stackexchange.com/q/595761.

In view of these apparent drawbacks of the classical Pearson and Spearman correlation coefficients, we propose a new approach of testing for cross-dependence in $(\mathbf{X}_t)_{t\in\mathbb{N}}$. Inspired by Weiß [87], we utilize appropriate kinds of (multivariate extensions of) ordinal patterns. Being rank-based statistics as well, ordinal patterns are known to be robust against outliers. In addition, if the considered process $(\mathbf{X}_t)_{t\in\mathbb{N}}$ is continuously distributed, then corresponding tests of the i.i.d.-null are non-parametric (distribution-free), which is highly attractive for applications in practice. In the case of discrete integer-valued data, however, the non-parametric nature can be preserved by adding uniform noise prior to the ordinal pattern analysis, see Section 7.3.1 for details.

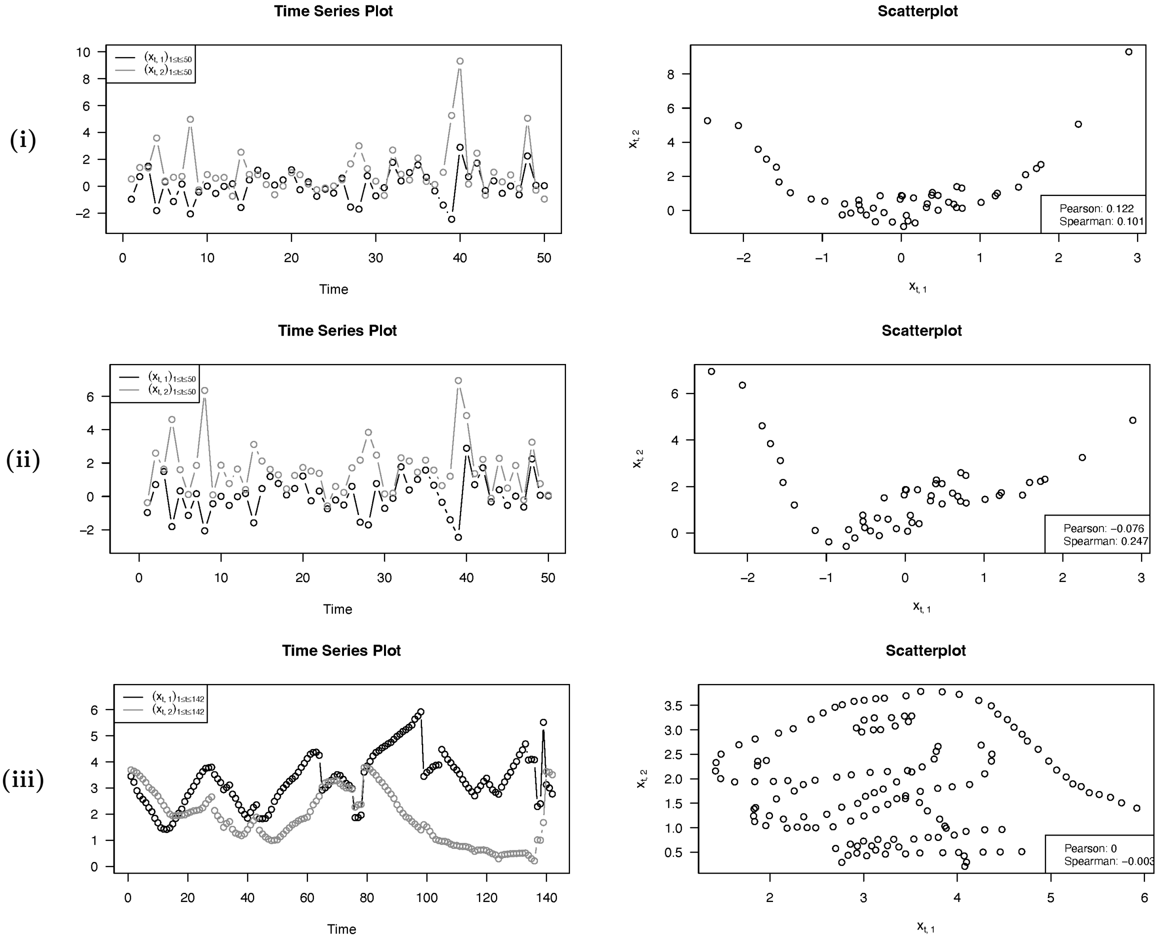


Figure 7.1: Data with regard to three different scenarios (rows), where the respective time series plots and scatterplots are given on the left and right, respectively. For each scenario, Pearson's and Spearman's correlation coefficients are indicated in the bottom right corner of the respective scatterplot, with none of them being significantly different from zero on the 5%-level.

The remainder of this chapter is organized as follows: First we derive a general central limit theorem for MOPs in Section 7.1.1, which allows us to establish the asymptotic distributions of various test statistics afterwards in Sections 7.1.2 and 7.1.3. As we will show in Section 7.2, also SOPs might form the starting point for constructing non-parametric tests of cross-dependence. The finite-sample performance of our novel ordinal pattern-based dependence tests is analyzed by simulations in Section 7.3, where it turns out that they constitute relevant complements of the popular Pearson and Spearman tests. Section 7.4 then presents the aforementioned real-world data examples to illustrate the applications of our tests in practice. Finally, we close in Section 7.5 with an interim conclusion and outline possible directions for future research in this regard.

This chapter as well as its introduction are based on the joint work [82] with C. H. Weiß and A. Schnurr.

7.1 Dependence Tests based on MOPs

Let $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ be a bivariate process, which we assume to be i.i.d. (but possibly cross-dependent) with some continuous bivariate cdf F, and recall that here we consider time series where the time component t runs column-wise from left to right. Furthermore, recall that the requirement for continuously distributed random variables implies that the probability of ties, both serially and between the components, equals zero. Under the null hypothesis, we suppose the additional independence between the components of $\mathbf{X}_t$, i.e., $F(x, y) = F_1(x) \cdot F_2(y)$ for all x, y.

7.1.1 A Central Limit Theorem for MOPs

For $i = 1, 2$, let the binary vectors $\mathbf{Z}_{t,i} = (Z_{t,i}^{(1)}, \ldots, Z_{t,i}^{(d!)})^\top \in \{0,1\}^{d!}$ be defined by $Z_{t,i}^{(k)} = 1$ iff the vector $(X_{t,i}, \ldots, X_{t+d-1,i})$ stemming from the i-th component of the time series at hand has the univariate ordinal pattern $\pi^{(k)}$, and $Z_{t,i}^{(k)} = 0$ otherwise ("one-hot encoding"). Then, the vector consisting of the respective marginal probabilities is given by $\mathbf{p}_i = (p_i^{(k)})_{1\le k\le d!} = \mathbb{E}(\mathbf{Z}_{t,i})$, $i = 1, 2$, while the vector consisting of the joint probabilities across components is given by $(p_{k,l})_{1\le k,l\le d!} = \mathbb{E}(\mathbf{Z}_{t,1}^\top \mathbf{Z}_{t,2})$. In favor of a more compact notation, we write $\tilde{\mathbf{Z}}_t = (\tilde{Z}_{t,r})_{1\le r\le 2d!+(d!)^2}$ with

$$\tilde{Z}_{t,r} = \begin{cases} Z_{t,1}^{(r)} & \text{if } 1 \le r \le d!, \\ Z_{t,2}^{(s)} & \text{if } r = d! + s \text{ with } 1 \le s \le d!, \\ Z_{t,1}^{(a)} \cdot Z_{t,2}^{(b)} & \text{if } r = 2d! + (a-1)d! + b = (a+1)d! + b \\ & \text{with } 1 \le a, b \le d!, \end{cases} \tag{7.1}$$

which encodes not only the respective univariate ordinal patterns of the marginals, but also the multivariate ordinal pattern at time t. Furthermore, we write $\tilde{\mathbf{p}} = \mathbb{E}(\tilde{\mathbf{Z}}_t) \in [0,1]^{2d!+(d!)^2}$ for the vector consisting of all aforementioned probabilities. Note that under the null of independent components, it holds that

$$\tilde{\mathbf{p}} = \Big(\underbrace{\frac{1}{d!}, \ldots, \frac{1}{d!}}_{(2\cdot d!)-\text{times}}, \underbrace{\frac{1}{(d!)^2}, \ldots, \frac{1}{(d!)^2}}_{(d!)^2-\text{times}}\Big)^\top.$$

If $(\mathbf{X}_t)_{t\in\mathbb{N}}$ is i.i.d., then the series of MOPs stemming from $(\mathbf{X}_t)_{t\in\mathbb{N}}$ is a stationary and $(d-1)$-dependent process itself, since MOPs are obtained from $(\mathbf{X}_t)_{t\in\mathbb{N}}$ by a sliding window approach. These properties transfer to $(\tilde{\mathbf{Z}}_t)_{t\in\mathbb{N}}$ accordingly. So estimating the above probabilities by the sample mean $\hat{\mathbf{p}} = \frac{1}{n}\sum_{t=1}^n \tilde{\mathbf{Z}}_t$, the application of a central limit theorem for $(d-1)$-dependent processes (see [37]) yields

$$\sqrt{n}(\hat{\mathbf{p}} - \tilde{\mathbf{p}}) \xrightarrow{D} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}) \tag{7.2}$$

with mean $\mathbf{0} = (0, \ldots, 0)^\top$ and cross-covariance matrix $\mathbf{\Sigma} = (\sigma_{k,l})_{1\le k,l\le 2d!+(d!)^2}$, where

$$\sigma_{r,s} = \mathrm{Cov}(\tilde{Z}_{0,r}, \tilde{Z}_{0,s}) + \sum_{h=1}^{d-1} \Big(\mathrm{Cov}(\tilde{Z}_{0,r}, \tilde{Z}_{h,s}) + \mathrm{Cov}(\tilde{Z}_{h,r}, \tilde{Z}_{0,s}) \Big).$$

According to (7.1), we obtain the following partition of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \left(\begin{array}{c|c|c} \sigma_{k,l} & \sigma_{k,\,d!+l} & \sigma_{k,\,(r+1)d!+s} \\ \hline \sigma_{d!+k,\,l} & \sigma_{d!+k,\,d!+l} & \sigma_{d!+k,\,(r+1)d!+s} \\ \hline \sigma_{(a+1)d!+b,\,l} & \sigma_{(a+1)d!+b,\,d!+l} & \sigma_{(a+1)d!+b,\,(r+1)d!+s} \end{array}\right)_{1\le k,l,a,b,r,s\le d!} . \tag{7.3}$$

The respective covariances can be computed exactly. We summarize the main results:

Lemma 7.1.1. *Let $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ be an i.i.d. series of random vectors. Under the null, we obtain the following results for Partition (7.3) of $\mathbf{\Sigma}$: For the blocks at positions $(1,1)$ and $(2,2)$ in (7.3), equality holds, i.e.,*

$$(\sigma_{d!+k,\,d!+l})_{1\le k,l\le d!} = (\sigma_{k,l})_{1\le k,l\le d!}. \tag{7.4}$$

The blocks at positions $(1,2)$ and $(2,1)$, in turn, vanish:

$$\sigma_{d!+k,\,l} \;=\; \sigma_{k,\,d!+l} = 0 \quad \text{for } 1\le k,l\le d!. \tag{7.5}$$

The blocks at positions $(1,3)$ and $(2,3)$ satisfy

$$\sigma_{k,\,(r+1)d!+s} = \frac{1}{d!}\cdot\sigma_{k,r} \quad \text{and} \quad \sigma_{d!+k,\,(r+1)d!+s} = \frac{1}{d!}\cdot\sigma_{k,s} \tag{7.6}$$

for $1\le k,r,s\le d!$, and those at positions $(3,1)$ and $(3,2)$ that

$$\sigma_{(a+1)d!+b,\,l} = \frac{1}{d!}\cdot\sigma_{a,l} \quad \text{and} \quad \sigma_{(a+1)d!+b,\,d!+l} = \frac{1}{d!}\cdot\sigma_{b,l} \tag{7.7}$$

for $1\le a,b,l\le d!$. Finally, the covariances of the block at position $(3,3)$ are given by

$$\begin{aligned} \sigma_{(a+1)d!+b,\,(r+1)d!+s} \;&=\; p^{(a,r)}(0)\cdot p^{(b,s)}(0) - \frac{1}{(d!)^4} \\ &+ \sum_{h=1}^{d-1}\Big(p^{(a,r)}(h)\cdot p^{(b,s)}(h) + p^{(r,a)}(h)\cdot p^{(s,b)}(h) - \frac{2}{(d!)^4}\Big), \end{aligned} \tag{7.8}$$

where $p^{(k,l)}(h) := \mathbb{E}(Z_{t,1}^{(k)}\cdot Z_{t+h,1}^{(l)})$, $1\le h\le d-1$, denotes the joint probability within one component with lag h.

Note the difference between $p^{(k,l)} = p^{(k,l)}(0) = \mathbb{E}(Z_{t,1}^{(k)}\cdot Z_{t,1}^{(l)}) = \mathbb{E}(Z_{t,2}^{(k)}\cdot Z_{t,2}^{(l)})$ and $p_{k,l} = \mathbb{E}(Z_{t,1}^{(k)}\cdot Z_{t,2}^{(l)})$. Furthermore, note that $p^{(k,l)}(h)$ for $d=2$ and $d=3$ has already been computed by Weiß [87], where also closed-form expressions for the respective cases $d=2$ and $d=3$ of (7.4) are provided. Hence, the covariances in (7.5)–(7.7) can also be directly derived from these. The covariances in (7.8), in turn, need to be computed separately. For $d=2$ and $d=3$, the exact cross-covariance matrices $\mathbf{\Sigma}$ are given in Appendix A.1.

Proof. Note that under the i.i.d.-assumption on $(\mathbf{X}_t)_{t\in\mathbb{N}}$, it follows that $(\mathbf{Z}_{t,1})_{t\in\mathbb{N}}$ and $(\mathbf{Z}_{t,2})_{t\in\mathbb{N}}$ are identically distributed such that $\sigma_{k,l} = \sigma_{d!+k,\,d!+l}$ holds for $1\le k,l\le d!$, which proves (7.4). In fact, for $1\le k,l\le d!$ and $0\le h\le d-1$, it holds that

$$\mathrm{Cov}(Z_{t,2}^{(k)}, Z_{t+h,2}^{(l)}) = \mathrm{Cov}(Z_{t,1}^{(k)}, Z_{t+h,1}^{(l)}) = \mathbb{E}(Z_{t,1}^{(k)}\cdot Z_{t+h,1}^{(l)}) - \frac{1}{(d!)^2}.$$

Due to independence of the components under the null, it holds

$$\mathrm{Cov}(Z_{t,1}^{(k)}, Z_{t+h,2}^{(l)}) = 0 = \mathrm{Cov}(Z_{t+h,1}^{(k)}, Z_{t,2}^{(l)})$$

for all $0 \leq h \leq d-1$, so (7.5) follows.

With the same reasoning, we obtain

$$\begin{aligned}\mathrm{Cov}(Z_{t,1}^{(j)}, Z_{t+h,1}^{(k)} \cdot Z_{t+h,2}^{(l)}) &= \mathbb{E}(Z_{t,1}^{(j)} \cdot Z_{t+h,1}^{(k)} \cdot Z_{t+h,2}^{(l)}) - \frac{1}{(d!)^3} \\ &= \frac{1}{d!}\Big(\mathbb{E}(Z_{t,1}^{(j)} \cdot Z_{t+h,1}^{(k)}) - \frac{1}{(d!)^2}\Big) \\ &= \frac{1}{d!}\,\mathrm{Cov}(Z_{t,1}^{(j)}, Z_{t+h,1}^{(k)}),\end{aligned}$$

and analogously $\mathrm{Cov}(Z_{t,2}^{(j)}, Z_{t+h,1}^{(k)} \cdot Z_{t+h,2}^{(l)}) = \frac{1}{d!}\,\mathrm{Cov}(Z_{t,2}^{(j)}, Z_{t+h,2}^{(l)})$. Therefore, we obtain (7.6). Furthermore, by the symmetry of the cross-covariance matrix, (7.7) follows.

Lastly, for $1 \leq a, b, r, s \leq d!$, consider

$$\begin{aligned}\mathrm{Cov}(Z_{t,1}^{(a)} \cdot Z_{t,2}^{(b)}, Z_{t+h,1}^{(r)} \cdot Z_{t+h,2}^{(s)}) &= \mathbb{E}(Z_{t,1}^{(a)} \cdot Z_{t+h,1}^{(r)} \cdot Z_{t,2}^{(b)} \cdot Z_{t+h,2}^{(s)}) - \frac{1}{(d!)^4} \\ &= \mathbb{E}(Z_{t,1}^{(a)} \cdot Z_{t+h,1}^{(r)}) \cdot \mathbb{E}(Z_{t,2}^{(b)} \cdot Z_{t+h,2}^{(s)}) - \frac{1}{(d!)^4} \\ &= p^{(a,r)}(h) \cdot p^{(b,s)}(h) - \frac{1}{(d!)^4}.\end{aligned}$$

□

7.1.2 The Permutation Entropy of MOPs

Regarding the PE of MOPs as suggested by Mohr et al. [58], it is sufficient to only consider the probabilities of the respective $(d!)^2$ MOPs, which correspond to the joint probabilities $(p_{k,l})_{1\leq k,l\leq d!}$ as well as their respective estimators. Then, the CLT (7.2) in Section 7.1.1 immediately yields

$$\sqrt{n}\big((\hat{p}_{k,l})_{1\leq k,l\leq d!} - \mathbf{p}\big) \xrightarrow{D} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}_{33}) \tag{7.9}$$

under the null of independence, where

$$\mathbf{p} = (p_{1,1}, \ldots, p_{1,d!},\ p_{2,1}, \ldots, p_{d!,d!})^\top = \left(\frac{1}{(d!)^2}, \ldots, \frac{1}{(d!)^2}\right)^\top,$$

and where $\mathbf{\Sigma}_{33} = (\sigma_{kl})_{2d!+1\leq k,l\leq 2d!+(d!)^2}$ equals the lower right submatrix of $\mathbf{\Sigma}$ in (7.3). Using the second-order Taylor expansion at $\mathbf{p}$, the standardized permutation entropy becomes

$$\begin{aligned}\mathrm{PE}\big((\hat{p}_{j,k})_{1\leq k,l\leq d!}\big) \approx \mathrm{PE}(\mathbf{p}) &- \frac{1}{2\log(d!)}\sum_{j,k=1}^{d!}(\log(p_{j,k})+1)\cdot(\hat{p}_{j,k} - p_{j,k}) \\ &- \frac{1}{4\log(d!)}\sum_{j,k=1}^{d!}\frac{(\hat{p}_{j,k} - p_{j,k})^2}{p_{j,k}}\end{aligned}$$

$$= 1 - \frac{(d!)^2}{4\log(d!)} \sum_{j,k=1}^{d!} \left(\hat{p}_{j,k} - \frac{1}{(d!)^2} \right)^2,$$

where $\mathrm{PE}(\mathbf{p}) = 1$ under the null of independence between components. Thus, it follows that

$$n \cdot \left(\mathrm{PE}\left((\hat{p}_{k,l})_{1 \le k,l \le d!} \right) - 1 \right) \approx -n \cdot \frac{(d!)^2}{4\log(d!)} \sum_{j,k=1}^{d!} \left(\hat{p}_{j,k} - \frac{1}{(d!)^2} \right)^2.$$

Now, application of (7.9) and Theorem 3.1 in [84] yields the following:

Theorem 7.1.2. *Let $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ be an i.i.d. series of random vectors. Under the null, it holds*

$$-n \cdot \frac{4\log(d!)}{(d!)^2} \left(\mathrm{PE}((\hat{p}_{k,l})_{1 \le k,l \le d!}) - 1 \right) \xrightarrow{D} \sum_{i=1}^{l} \lambda_i \cdot \chi^2_{r_i} =: Q_d,$$

where $\lambda_1, \dots, \lambda_l \neq 0$ denote the eigenvalues of $\mathbf{\Sigma}_{33}$ with algebraic multiplicities $r_1, \dots, r_l$, respectively, χ^2_r refers to a χ^2-distributed random variable with r degrees of freedom and $\chi^2_{r_1}, \dots, \chi^2_{r_l}$ are independent of each other.

Consequently, mean and variance of the limit distribution are equal to $\mathbb{E}Q_d = \sum_{i=1}^{l} r_i \cdot \lambda_i = \mathrm{trace}(\mathbf{\Sigma}_{33})$ and $\mathrm{Var}(Q_d) = 2\sum_{i=1}^{l} r_i \cdot \lambda_i^2$, respectively.

For $d = 2$, the non-zero eigenvalues of $\mathbf{\Sigma}_{33}$ are given by $\lambda_1 = 11/36$ and $\lambda_2 = 1/12$ with respective multiplicities $r_1 = 1$ and $r_2 = 2$. Also note that in case of the mean we obtain $\sum_{i=1}^{2} r_i \cdot \lambda_i = \frac{17}{36}$ and for the variance $2 \cdot \sum_{i=1}^{2} r_i \lambda_i^2 = \frac{139}{648}$.

Corollary 7.1.3. *Let $d = 2$. Under the assumptions of Theorem 7.1.2,*

$$-n \log(2) \left(\mathrm{PE}((\hat{p}_{k,l})_{1 \le k,l \le d!}) - 1 \right)$$

is asymptotically distributed like the quadratic form

$$Q_2 = \frac{11}{36} \cdot \chi^2_1 + \frac{1}{12} \cdot \chi^2_2.$$

In particular, its asymptotic mean and variance are given by $\frac{17}{36}$ and $\frac{139}{648}$, respectively.

For $d = 3$, some of the eigenvalues are roots of third- and fourth-order polynomials such that they cannot be written down exactly in compact form. Thus, numerical approximations to the non-zero eigenvalues of $1\,000 \cdot \mathbf{\Sigma}_{33}$ together with their respective multiplicities are enlisted in Table 7.1.

Corollary 7.1.4. *Let $d = 3$. Under the assumptions of Theorem 7.1.2,*

$$-n \cdot \frac{\log(6)}{9} \left(\mathrm{PE}((\hat{p}_{k,l})_{1 \le k,l \le d!}) - 1 \right)$$

is asymptotically distributed like the quadratic form

$$Q_3 = \sum_{i=1}^{22} \lambda_i \cdot \chi^2_{r_i},$$

where the eigenvalues λ_i and multiplicities r_i are provided by Table 7.1. In particular, its asymptotic mean and variance are given by $\frac{203}{225}$ and $\frac{95\,431}{1\,620\,000}$, respectively.

Table 7.1: Non-zero eigenvalues (rounded) of $1\,000 \cdot \mathbf{\Sigma}_{33}$ from Corollary 7.1.4 and their respective multiplicities.

i	1	2	3	4	5	6	7	8
λ_i	0.631118	0.474196	0.406304	0.376367	0.364922	0.350167	0.315313	0.3125
r_i	1	2	1	2	2	1	1	1

i	9	10	11	12	13	14	15	16
λ_i	0.293148	0.277778	0.268432	0.263119	0.261076	0.258333	0.223257	0.222222
r_i	1	1	2	2	1	1	2	2

i	17	18	19	20	21	22
λ_i	0.218622	0.212237	0.208333	0.183974	0.166667	0.0813592
r_i	1	2	1	1	2	2

Based on Theorem 7.1.2 (and its Corollaries 7.1.3 and 7.1.4), we can now define the following (one-sided) hypothesis tests on level α. Let $q_{d,1-\alpha}$ denote the $(1-\alpha)$-quantile of the quadratic form distribution Q_d. Then, we reject the null of independent components if

$$\mathrm{PE}\left((\hat{p}_{k,l})_{k,l}\right) < 1 - \frac{(d!)^2}{n \cdot 4\log(d!)} q_{d,1-\alpha}. \tag{7.10}$$

The critical values from (7.10) which are most frequently required in practice are provided by Table 7.2.

Table 7.2: Critical values (rounded) from (7.10) for $\alpha \in \{0.10, 0.05, 0.01\}$ and $d \in \{2, 3\}$ depending on the sample size n.

α	0.10	0.05	0.01
$d=2$	$1-1.5/n$	$1-1.992/n$	$1-3.214/n$
$d=3$	$1-6.145/n$	$1-6.719/n$	$1-7.91/n$

7.1.3 Dependence Tests via Ordinal Pattern Dependence

As OPD evaluates the agreement between the ordinal patterns in the respective components $\mathbf{X} = (X_{1,1}, \dots, X_{d,1})$ and $\mathbf{X}^* = (X_{1,2}, \dots, X_{d,2})$, it can also be interpreted as a special MOP-based statistic (different to PE), where only MOPs with equal components are considered. Hence, noting that the OPD between $\mathbf{X}$ and $\mathbf{X}^*$ can be written as

$$\mathrm{OPD}_d(\mathbf{X}, \mathbf{X}^*) = \frac{\sum_{k=1}^{d!} p_{k,k} - \sum_{k=1}^{d!} p_1^{(k)} \cdot p_2^{(k)}}{1 - \sum_{k=1}^{d!} p_1^{(k)} \cdot p_2^{(k)}},$$

we derive the limit distribution of OPD under the null. The following result encompasses Theorem 2 of Betken et al. [17], since here we explicitly derive the covariance structure. The proofs are still very similar though.

Theorem 7.1.5. *Let $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ be an i.i.d. series of random vectors. Let $\mathbf{\Sigma}^* \in [-1,1]^{3d!\times 3d!}$ denote that submatrix of $\mathbf{\Sigma}$ from* (7.3)*, which refers to*

$$(Z_{t,1}^{(1)}, \ldots, Z_{t,1}^{(d!)}, Z_{t,2}^{(1)}, \ldots, Z_{t,2}^{(d!)}, Z_{t,1}^{(1)} \cdot Z_{t,2}^{(1)}, Z_{t,1}^{(2)} \cdot Z_{t,2}^{(2)}, \ldots, Z_{t,1}^{(d!)} \cdot Z_{t,2}^{(d!)})^\top.$$

Furthermore, denote

$$\mathbf{D} := \Bigg(\underbrace{\frac{-1}{d!-1}, \ldots, \frac{-1}{d!-1}}_{(2d!)\text{-}times}, \underbrace{\frac{d!}{d!-1}, \ldots, \frac{d!}{d!-1}}_{(d!)\text{-}times}\Bigg).$$

Then, under the null of independence, it holds that

$$\sqrt{n}\ \mathrm{OPD}_d(\mathbf{X}, \mathbf{X}^*) \xrightarrow{D} N(0, \sigma^2), \quad \text{where } \sigma^2 = \mathbf{D}\,\mathbf{\Sigma}^*\,\mathbf{D}^\top.$$

Proof. The proof is done by applying the "Delta method" (see [77, Theorem 3.3.A]) to (7.2). For this purpose, define the function $f : \mathbb{R}^{d!} \times \mathbb{R}^{d!} \times \mathbb{R}^{d!} \to \mathbb{R}$ by

$$f(\mathbf{u}, \mathbf{v}, \mathbf{w}) := \frac{\mathbf{w}^\top \mathbf{1} - \mathbf{u}^\top \mathbf{v}}{1 - \mathbf{u}^\top \mathbf{v}} = 1 + \frac{\mathbf{w}^\top \mathbf{1} - 1}{1 - \mathbf{u}^\top \mathbf{v}},$$

such that $\mathrm{OPD}_d(\mathbf{X}, \mathbf{X}^*) = f\left((p_1^{(k)})_{1\le k\le d!}, (p_2^{(k)})_{1\le k\le d!}, (p_{k,k})_{1\le k\le d!}\right)$. Then, the gradient ∇f is given by

$$\nabla f(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \left(\mathbf{v} \cdot \frac{\mathbf{w}^\top \mathbf{1} - 1}{(1 - \mathbf{u}^\top \mathbf{v})^2}, \mathbf{u} \cdot \frac{\mathbf{w}^\top \mathbf{1} - 1}{(1 - \mathbf{u}^\top \mathbf{v})^2}, \mathbf{1} \cdot \frac{1}{1 - \mathbf{u}^\top \mathbf{v}}\right)^\top.$$

Under the null, it holds that

$$\left((p_1^{(k)})_{1\le k\le d!}, (p_2^{(k)})_{1\le k\le d!}, (p_{k,k})_{1\le k\le d!}\right) = \Bigg(\underbrace{\frac{1}{d!}, \ldots, \frac{1}{d!}}_{(2d!)\text{-times}}, \underbrace{\frac{1}{(d!)^2}, \ldots, \frac{1}{(d!)^2}}_{(d!)\text{-times}}\Bigg),$$

so that

$$\nabla f\left((p_1^{(k)})_{1\le k\le d!}, (p_2^{(k)})_{1\le k\le d!}, (p_{k,k})_{1\le k\le d!}\right) = \Bigg(\underbrace{\frac{-1}{d!-1}, \ldots, \frac{-1}{d!-1}}_{(2d!)\text{-times}}, \underbrace{\frac{d!}{d!-1}, \ldots, \frac{d!}{d!-1}}_{(d!)\text{-times}}\Bigg).$$

This expression for $\mathbf{D}$ together with the Delta method completes the proof. □

We were able to compute the cross-covariance matrix $\mathbf{\Sigma}$ for $d = 2$ and $d = 3$ exactly (under the null), recall Appendix A.1. Consequently, we are able to give exact limiting variances σ^2 in Theorem 7.1.5 for these cases.

Corollary 7.1.6. *Under the assumptions of Theorem 7.1.5, it follows*

$$\sqrt{n}\ \mathrm{OPD}_2(\mathbf{X}, \mathbf{X}^*) \xrightarrow{D} N\Big(0, \frac{11}{9}\Big),$$

$$\sqrt{n}\ \mathrm{OPD}_3(\mathbf{X}, \mathbf{X}^*) \xrightarrow{D} N\Big(0, \frac{401}{1250}\Big).$$

Based on the results of Corollary 7.1.6, two-sided critical values can be derived. The null of independence is rejected if

$$
\begin{aligned}
|\operatorname{OPD}_2(\mathbf{X},\mathbf{X}^*)| &> z_{1-\alpha/2}\sqrt{11/(9n)},\\
|\operatorname{OPD}_3(\mathbf{X},\mathbf{X}^*)| &> z_{1-\alpha/2}\sqrt{401/(1250n)},
\end{aligned}
$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of a standard normal distribution.

7.2 A Spatial Approach towards Cross-Dependence

Bandt and Wittfeld [12] and Weiß and Kim [88] have considered rectangular spatial data occurring in a regular grid, being generated by a stationary random field with continuous distribution. For analyzing the corresponding spatial dependence structure, they have used a spatial extension of ordinal patterns, namely SOPs, which are a kind of rectangular ordinal pattern. In the present research, we are not concerned with spatial data, but with data generated by a bivariate process. Hence, the MOPs discussed in Section 7.1.1 are the natural approach for analyzing the cross-dependence structure. But as a bivariate time series of length n can also be interpreted as a $2\times n$-rectangle, there appears a possibility to utilize SOPs as well. The motivation is that MOPs analyze the bivariate data in a componentwise manner, whereas SOPs consider the order within the extracted segment holistically and thus, they might be better suited to uncover certain forms of cross-dependence. For this reason, as a further competitor to MOP-based tests, we shall also investigate whether SOP-based tests might be useful in some situations for uncovering cross-dependence in bivariate time series data.

Recall that $\mathcal{S}$ is defined as the set consisting of the $(2d)!$ SOPs. For a specified partition $\mathcal{S}=\mathcal{S}_1\cup\cdots\cup\mathcal{S}_K$, $1\le K\le(2d)!$, we define again binary vectors $\mathbf{Z}_t\in\{0,1\}^K$, where $Z_t^{(j)}=1$ iff the SOP $\mathbf{\Pi}_t$ of $\mathbf{X}_t$ belongs to $\mathcal{S}_j$ and $Z_t^{(j)}=0$ otherwise. We denote the mean of $\mathbf{Z}_t$ by $\mathbf{p}=\mathbb{E}\mathbf{Z}_t\in[0,1]^K$, whereas we denote its sample mean by $\hat{\mathbf{p}}=\overline{\mathbf{Z}}=\frac{1}{n}\sum_{t=1}^{n}\mathbf{Z}_t$. Note that $p^{(j)}=\mathbb{P}(\mathbf{\Pi}_t\in\mathcal{S}_j)$.

Remark 7.2.1. There is a fundamental difference in applying MOPs and SOPs. For MOPs, the distribution of the components $X_{t,i}$ does not matter, it is just important that the pairs $(\mathbf{X}_t)_{t\in\mathbb{N}}=(X_{t,1},X_{t,2})_{t\in\mathbb{N}}$ are i.i.d. with independent components under the null hypothesis. SOPs, in turn, require an identical (continuous) distribution for the whole rectangular data set (plus independence under the null, i.e., altogether an i.i.d. random field as the data generating process). Hence, in order to analyze an i.i.d. bivariate process $(\mathbf{X}_t)_{t\in\mathbb{N}}=(X_{t,1},X_{t,2})_{t\in\mathbb{N}}$ for cross-dependence, we have to assume that $X_{t,1}$ and $X_{t,2}$ are also identically distributed. If then the null of mutually independent components $X_{t,1}$ and $X_{t,2}$ holds, it even follows that $X_{t,1}$ and $X_{t,2}$ are i.i.d. (and not just independent).

Altogether, we have a more specific field of application for SOPs than for MOPs. In some applications, the assumption of identically distributed components is fulfilled naturally, e.g., if the considered process refers to parallel production lines in manufacturing industry. But often, it will not be satisfied. In such cases, see the details in Section 7.3, we recommend to apply the subsequent SOP-based statistics only after a prior standardization of the components $(X_{t,1})_{t\in\mathbb{N}}$ and $(X_{t,2})_{t\in\mathbb{N}}$, respectively. This does not guarantee exactly identically distributed components, but at least approximately ones. In Section 7.3, we shall analyze by simulations if this approximate approach is sufficient for practice.

For the rest of this section, let us assume that $(\mathbf{X}_t)_{t\in\mathbb{N}} = (X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ is an i.i.d. process with identically and continuously distributed components. Furthermore, assume that $X_{t,1}$ and $X_{t,2}$ are independent (null hypothesis). Then, $\mathbf{p} = \big(|\mathcal{S}_1|/(2d)!, \ldots, |\mathcal{S}_q|/(2d)!\big)^\top$ holds. If extracting SOPs from a bivariate process, horizontal overlaps arise naturally. These overlaps again result in some kind of $(d-1)$-dependence between successive SOPs. This dependence has to be considered regarding the asymptotic distribution of $\hat{\mathbf{p}}$.

Focusing on $d = 2$, Weiß and Kim [88] have deduced the limiting behavior of the SOP frequencies $\hat{\mathbf{p}}$ (in a more general setting than considered here, see [88, Theorem 2.1]). Adapting their results to the bivariate time series case, we obtain

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{D} \mathrm{N}(\mathbf{0}, \tilde{\mathbf{\Sigma}}) \tag{7.11}$$

with

$$\tilde{\mathbf{\Sigma}} = \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top + (1 - \tfrac{1}{n})(\mathbf{H} + \mathbf{H}^\top - 2\mathbf{p}\mathbf{p}^\top), \tag{7.12}$$

where the matrix $\mathbf{H}$ refers to the (horizontal) overlaps of successive SOPs. In [88], the authors have computed $\mathbf{H}$ for $K = (2d)!$, see Table A.3 in Appendix A.1.3, from which the respective matrices for cases $K < (2d)!$ can be immediately derived. In Appendix A.1.3, we also provide the result for the large-sample approximation of $\tilde{\mathbf{\Sigma}}$, where we set the factor $(1 - \frac{1}{n}) \approx 1$.

As there are 24 different SOPs of length $d = 2$, the vector $\mathbf{p}$ is 24-dimensional and, thus, difficult to estimate from small data sets. Therefore, we follow the recommendation in [12, 88], where the asymptotic behavior regarding so-called "types" instead of all individual SOPs has been derived (see Section 6.2).

For partition (6.3), $\mathbf{p} = (1/3, 1/3, 1/3)^\top$ under the null, and the matrix $\mathbf{H}$ becomes the symmetric matrix

$$\mathbf{H} \;=\; \frac{1}{180}\begin{pmatrix} 21 & 20 & 19 \\ 20 & 21 & 19 \\ 19 & 19 & 22 \end{pmatrix}.$$

Therefore, $\tilde{\mathbf{\Sigma}}$ from (7.12) simplifies to

$$\begin{aligned}
\tilde{\mathbf{\Sigma}} \;&=\; \mathrm{diag}(\mathbf{p}) - \mathbf{p}\,\mathbf{p}^\top \;+\; 2(1 - \tfrac{1}{n})\Big(\mathbf{H} - \mathbf{p}\,\mathbf{p}^\top\Big) \\
&=\; \frac{1}{9}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \;+\; 2\cdot\big(1 - \tfrac{1}{n}\big)\cdot\frac{1}{180}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix} \\
&\to\; \frac{1}{90}\begin{pmatrix} 21 & -10 & -11 \\ -10 & 21 & -11 \\ -11 & -11 & 22 \end{pmatrix} \quad \text{for } n \to \infty,
\end{aligned} \tag{7.13}$$

see [88] for detailed derivations.

Remark 7.2.2. The larger d, the more kinds of horizontal overlaps have to be considered, which we refer to as overlaps of order 1 to $d-1$. This is illustrated in Figure 7.2 for SOPs of length $d = 3$.

Bearing this in mind, we may generalize (7.11) and (7.12) for $d \geq 2$. In fact, (7.11) holds for any $d \geq 2$ by the central limit theorem for $(d-1)$-dependent processes (see [37]), but the limit covariance matrix $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{\Sigma}}(d)$ is dependent on d. In general it holds

$$n \cdot \tilde{\mathbf{\Sigma}}(d) = n^2 \cdot \mathrm{Cov}(\overline{\mathbf{Z}}, \overline{\mathbf{Z}}^\top) = \sum_{t_1,t_2=1}^{n} \mathrm{Cov}(\mathbf{Z}_{t_1}, \mathbf{Z}_{t_2}^\top)$$

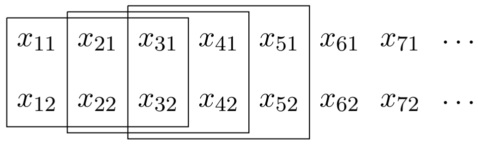


Figure 7.2: Overlapping SOPs of length $d = 3$ in a bivariate time series: consecutive blocks have overlaps of order $d - 1 = 2$, while skipping a block (delay by 1) leads to overlaps of order $d - 2 = 1$. Delays > 1 lead to non-overlapping blocks.

with

$$\mathrm{Cov}(\mathbf{Z}_{t_1}, \mathbf{Z}_{t_2}^\top) = \mathbf{0}_{(2d)! \times (2d)!} \quad \text{for } |t_1 - t_2| \geq d,$$

where $\mathbf{0}_{(2d)! \times (2d)!}$ denotes the $(2d)! \times (2d)!$ zero matrix. These constitute the non-overlaps. Now, considering the total overlaps as well as the overlaps of order $o \in \{1, \ldots, d-1\}$, it follows

$$n \cdot \tilde{\mathbf{\Sigma}}(d) = n \cdot \mathrm{Cov}(\mathbf{Z}_0, \mathbf{Z}_0^\top) + \sum_{o=1}^{d-1} (n - (d - o)) \Big(\mathrm{Cov}(\mathbf{Z}_0, \mathbf{Z}_{d-o}^\top) + \mathrm{Cov}(\mathbf{Z}_{d-o}, \mathbf{Z}_0^\top) \Big).$$

Then, defining $\mathbf{H}_o := \mathbb{E}(\mathbf{Z}_t \mathbf{Z}_{t+d-o}^\top) = \mathbb{E}(\mathbf{Z}_t \mathbf{Z}_{t+k}^\top)$ for $k \in \{1, \ldots, d-1\}$, this yields

$$\tilde{\mathbf{\Sigma}}(d) = \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top + \sum_{o=1}^{d-1} \left(1 - \frac{d-o}{n}\right) \left(\mathbf{H}_o \mathbf{H}_o^\top - 2\mathbf{p}\mathbf{p}^\top\right).$$

Depending on d, we then need to estimate the probabilities for (jointly) observing SOPs, that is, we need to consider the frequencies of SOPs of length up to $2d - 1$. Hence, in order to obtain reasonable frequencies either a very large data set or forming $K \ll (2d)!$ groups of SOPs is necessary (see [88]). With regard to the latter, it is part of future research to find such an appropriate partition for $d > 2$ or at least $d = 3$ ('extended types', recall Section 6.2). However, in the remaining we focus on $d = 2$.

Similar to the case of MOPs in Theorem 7.1.2, using the second-order Taylor expansion as well as (7.11) and Theorem 3.1 of [84], we obtain the following asymptotic result.

Theorem 7.2.3. *Let $(\mathbf{X}_t)_{t \in \mathbb{N}} = (X_{t,1}, X_{t,2})_{t \in \mathbb{N}}$ be an i.i.d. series of random vectors. Under the null, it holds*

$$-n \cdot \frac{2\log(K)}{K} \big(\mathrm{PE}(\hat{\mathbf{p}}) - 1\big) \xrightarrow{D} \sum_{i=1}^{l} \lambda_i \cdot \chi_{r_i}^2,$$

where $\lambda_1, \ldots, \lambda_l \neq 0$ denote the eigenvalues of $\tilde{\mathbf{\Sigma}}$ with algebraic multiplicities $r_1, \ldots, r_l$, respectively. Here, χ_r^2 refers to a χ^2-distributed random variable with r degrees of freedom, and $\chi_{r_1}^2, \ldots, \chi_{r_l}^2$ are independent of each other.

Therefore, asymptotically, we obtain again a quadratic-form distribution. Note that if considering types ($K = 3$), the required non-zero eigenvalues of the large-sample approximation to $\tilde{\mathbf{\Sigma}}$ in (7.13) are given by $\frac{11}{30}$ and $\frac{31}{90}$ (multiplicity 1 each), so

$$-n \cdot \tfrac{2\log(3)}{3} \big(\mathrm{PE}(\hat{\mathbf{p}}) - 1\big) \xrightarrow{D} \tfrac{11}{30} \cdot \chi_1^2 + \tfrac{31}{90} \cdot \chi_1^2 \quad \text{with mean } \tfrac{32}{45} \text{ and variance } \tfrac{41}{81}. \tag{7.14}$$

Table 7.3: Non-zero eigenvalues and their respective multiplicities for $\tilde{\mathbf{\Sigma}}$ according to Table A.4.

i	1	2	3	4	5	6	7	8	9	10
λ_i	$\frac{1}{12}$	$\frac{7}{120}$	$\frac{17}{360}$	$\frac{2}{45}$	$\frac{1}{24}$	$\frac{7}{180}$	$\frac{13}{360}$	$\frac{1}{40}$	$\frac{7}{360}$	$\frac{1}{120}$
r_i	1	3	1	3	6	3	1	2	1	1

For the large-sample approximation to $\tilde{\mathbf{\Sigma}}$ in case of individual SOPs ($K = 24$), see Table A.4, the non-zero eigenvalues are given in Table 7.3. Hence, in this case, the corresponding quadratic-form distribution has mean $\frac{331}{360}$ and variance $\frac{623}{7200}$.

The asymptotics according to Theorem 7.2.3 can now be used to define a (one-sided) hypothesis test on level α. Let Q denote the asymptotic quadratic-form distribution obtained in Theorem 7.2.3, and let $q_{1-\alpha}$ be its $(1-\alpha)$-quantile. Then, we reject the null if

$$\mathrm{PE}(\hat{\mathbf{p}}) < 1 - \frac{K}{n \cdot 2\log(K)} q_{1-\alpha}.$$

In case of considering types, further statistics than just the PE are possible. Bandt and Wittfeld [12] have proposed

$$\tau = p^{(1)} - 1/3 \quad \text{and} \quad \kappa = p^{(2)} - p^{(3)},$$

which have range $[-1/3, 2/3]$ and $[-1, 1]$, respectively, and become zero under the null. As an alternative, they have proposed

$$\tau' = p^{(3)} - 1/3 \quad \text{and} \quad \kappa' = p^{(1)} - p^{(2)}.$$

We denote the respective test statistics by $\hat{\tau} = \overline{Z}_1 - 1/3$, $\hat{\kappa} = \overline{Z}_2 - \overline{Z}_3$, $\hat{\tau}' = \overline{Z}_3 - 1/3$ and $\hat{\kappa}' = \overline{Z}_1 - \overline{Z}_2$, where $\overline{\mathbf{Z}} = (\overline{Z}_1, \overline{Z}_2, \overline{Z}_3)^\top$ (recall that $K = 3$). Then, adapting the derivation of Corollary 3.1 in [88] to (7.13), we obtain that under the null, the vectors $(\hat{\tau}, \hat{\kappa})$ and $(\hat{\tau}', \hat{\kappa}')$ are asymptotically normally distributed, namely $\sqrt{n}(\hat{\tau}, \hat{\kappa}) \xrightarrow{D} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}')$ and $\sqrt{n}(\hat{\tau}', \hat{\kappa}') \xrightarrow{D} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}'')$, where

$$\begin{aligned} \mathbf{\Sigma}' &= \frac{2}{9}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + (1 - \tfrac{1}{n}) \cdot \frac{1}{90}\begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} \approx \frac{1}{90}\begin{pmatrix} 21 & 1 \\ 1 & 65 \end{pmatrix} \\ \mathbf{\Sigma}'' &= \frac{2}{9}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + (1 - \tfrac{1}{n}) \cdot \frac{1}{45}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \approx \frac{1}{45}\begin{pmatrix} 11 & 0 \\ 0 & 31 \end{pmatrix} \end{aligned} \tag{7.15}$$

for n large. On this basis, Weiß and Kim [88] defined four kinds of hypothesis test on level α. As an illustrative example, the null is rejected by the τ'-test if

$$\sqrt{n}\,|\hat{\tau}'| > z_{1-\alpha/2}\sqrt{\frac{2}{9} + \frac{1}{45}\left(1 - \frac{1}{n}\right)}, \tag{7.16}$$

where $z_{1-\alpha/2}$ again denotes the $(1-\alpha/2)$-quantile of the standard normal distribution. For large n, one may use the respective approximate expression for the variance in (7.15) instead. In our simulation study in Section 7.3, we have used the critical values based on the exact variances.

7.3 Performance of Dependence Tests

In this section, the finite-sample performance of the novel dependence tests is investigated by simulations. While we considered all statistics discussed before in our simulation study, we restrict our presentation to the PE of 2nd-order MOPs (MOP_2) from Corollary 7.1.3, the PE of 3rd-order MOPs (MOP_3) from Corollary 7.1.4, the dependence measures OPD_2 and OPD_3 from Corollary 7.1.6, and to two-SOP-based statistics, namely the PE of types (PE_{typ}) from (7.14) as well as τ' from (7.16), because these six statistics generally turned out to perform best. Compare in this regard the recommendation in Bandt and Wittfeld [12], who preferred τ and κ in the context of image analysis instead of τ', and Weiß and Kim [88]. As the competitors, we use the classical Pearson's and Spearman's correlation ("Pears" and "Spear", respectively) as well as the rank-based Chatterjee's correlation coefficient (CCC) from Chatterjee [22], which has recently become quite popular. Note that Chatterjee constitutes a directed measure of dependence, so we always consider both versions, CCC_{12} and CCC_{21}.

Throughout our simulations, we considered time series of i.i.d. pairs $(X_{t,1}, X_{t,2})_{t=1,\dots,n}$ of length $n \in \{50, 100, 250, 500\}$ and 10,000 repetitions. The nominal level of the tests was chosen as 5%. Moreover, as identically distributed components (which hardly ever holds in practice) are required for the SOPs to be uniformly distributed (under the null of independence), and as the SOP distribution is very sensitive to different scalings in the respective components in particular, we applied the SOP-based statistics only after first performing a standardization of the components: Given $(X_{1,1}, X_{1,2}), \dots, (X_{n,1}, X_{n,2})$, we computed the component-wise sample means $\overline{x}_1, \overline{x}_2$ and sample standard deviations s_1, s_2. Then, we substituted the original data $X_{t,i}$ by $Y_{t,i} = (X_{t,i} - \overline{x}_i)/s_i$ for $i = 1, 2$, and applied the SOP statistics to these standardized data afterwards. All other tests were performed on the original (non-standardized) data.

All computations of this section (as well as the subsequent data analyses) have been carried out in R, where we used the packages "`ordinalpattern`" and "`xicor`" for the computation of OPD_d and Chatterjee's coefficients (as well as p-values of the latter), respectively. The results are tabulated in Appendix A.2.

7.3.1 Analysis of Sizes

The starting point of our discussion is size analysis, i.e., if $(X_{t,1}, X_{t,2})_{t\in\mathbb{N}}$ has independent components. There, we have investigated the following scenarios:

1. Both components are identically distributed according to

    N-N: $X_{t,1}, X_{t,2} \sim N(0, 1)$;

    E1-E1: $X_{t,1}, X_{t,2} \sim Exp(1)$;

    P1-P1: $X_{t,1}, X_{t,2} \sim Poi(1)$.

    Here, $N(0, 1)$ denotes the standard normal distribution, $Exp(1)$ the exponential distribution with parameter value 1, and $Poi(1)$ the Poisson distribution with mean 1.

2. The components have distributions from different continuous distribution families:

    N-E1: $X_{t,1} \sim N(0, 1)$ but $X_{t,2} \sim Exp(1)$.

3. Components have distributions from different distribution families, where at least one component is discrete-valued:

   N-P1: $X_{t,1} \sim \mathrm{N}(0,1)$, but $X_{t,2} \sim \mathrm{Poi}(1)$;

   N-P5: $X_{t,1} \sim \mathrm{N}(0,1)$, but $X_{t,2} \sim \mathrm{Poi}(5)$.

Note that if a component $X_{t,i}$ is discrete-valued with the range being coded by integers, as it is the case here for the Poisson-distribution, then ordinal pattern-based statistics require to first apply "jittering" to that component, see [52, 89], the randomization by standard-uniform noise. The same applies to Chatterjee's correlation coefficient, see [22], while we have computed the classical cross-correlations from the original data (without noise). For example, in the last scenario, we have considered a second component of the form $X_{t,2} = Z_t + \varepsilon_t$ with $Z_t \sim \mathrm{Poi}(5)$ and $\varepsilon_t \sim \mathrm{U}[0,1]$, a uniform distribution on $[0,1]$, except for the classical correlations.

We report the tests' empirical sizes in Table A.5 in Appendix A.2.1. The rejection rates of most statistics are reasonably close to the nominal level of 5 % except for MOP_3. There, oversizing for small sample sizes $n \in \{50, 100\}$ can be recognized. However, this is not surprising. There are $(3!)^2 = 36$ MOPs of length $d = 3$, to which only $n - 2$ MOPs stemming from the sample of size n can be assigned. Therefore, for small sample sizes, the MOP distribution can deviate from the uniform distribution quite easily. On the contrary, the empirical size converges to the theoretical significance level quite fast for larger sample sizes, and it is already quite close to the significance level for $n = 250$. At this point, also note that the scenario at hand does not seem to have a strong influence on the degree of oversizing. In this context, it is also worth noting that the SOP-based statistics hold the level reasonably well also in the mixed cases **N-E1**, **N-P1**, and **N-P5**. While we performed a component-wise standardization prior to SOP computation, this does still not ensure an identical distribution across the components (only equal means and variances). Nevertheless, the SOP-based statistics appear to be robust against the remaining deviations from identical distributions.

7.3.2 Analysis of Sizes under Outliers

In this section, we still restrict to mutually independent components such that the null hypothesis is satisfied. But now, we complement our analysis of sizes by consideration of the following scenarios, where outliers can be observed.

1. Both components are identically t_1-distributed, that is,

   t1-t1: $X_{t,1}, X_{t,2} \sim t_1$.

2. Standard normal variates are randomly contaminated by additive outliers (AOs), which occur either simultaneously (sim) in both components or not (nsim), and either a few strong or many weak AOs:

   $\mathbf{AO}^{10}_{\mathbf{nsim}}$: 10 percent outliers with mean 1, which occur randomly but non-simultaneously in both components;

   $\mathbf{AO}^{2}_{\mathbf{nsim}}$: 2 percent outliers with mean 10, which occur randomly but non-simultaneously in both components;

AO^{10}_{sim}: 10 percent outliers with mean 1, which occur randomly and simultaneously in both components;

AO^{2}_{sim}: 2 percent outliers with mean 10, which occur randomly and simultaneously in both components.

From Table A.6, we again observe an oversizing-effect for MOP_3. In addition, with regard to the t_1-distributed marginals, there is strong oversizing for SOP-based tests. This can be explained by the fact that the SOP-based tests require prior standardization, which, in turn, is affected by the outliers. While both mean and variance are generally sensitive to outliers, in case of the t_1-distribution, these do not even exist anymore. For the same reason, also the sizes of the Pearson correlation are misleading, being too large for $n = 50$ (oversizing) and too small (undersizing) otherwise. Regarding the non-simultaneous outliers, we do not observe any systematic effects on the rejection rates.

However, this changes for simultaneous outliers. There, oversizing can be noted for all tests, though to varying extent. In fact, this behavior is reasonable (although not desirable), as the simultaneous occurrence of outliers causes a kind of positive dependence. In case of many weak outliers, tests based on Chatterjee's coefficient are most robust, followed by SOP-based tests. Apart from the difficulties being typical for MOPs of length $d = 3$, the rejection rates for the MOP-based tests (including OPD) are relatively similar, while Pearson's and Spearman's correlation perform worst in comparison. In case of few strong simultaneous outliers, SOP-based tests are superior followed by MOP_2 and OPD_2. For larger sample sizes, even OPD_3 and MOP_3 outperform Spearman's and Chatterjee's coefficients. Pearson's correlation is clearly worst with a rejection rate of 100%, as simultaneous outliers cause a kind of bogus linearity.

Finally, if we compare the tests regarding few strong or many weak outliers, then MOP_2, OPD_2 and Spearman's correlation perform better with regard to few weak, while it is the other way round for tests based on Pearson's and Chatterjee's coefficients. For tests based on MOP_3 and OPD_3 as well as SOP-based tests, no notable difference can be reported. Overall, our proposed tests seem to be more robust with respect to simultaneous outliers.

7.3.3 Power Analysis

For the power analysis without outliers, we have considered the same six scenarios as in the size analysis, but with additonal cross-dependence. For the sake of uniqueness, we always used a Gaussian copula for causing cross-dependencies, where the dependence parameter was set to ± 0.3, respectively. The respective results are reported in Tables A.7 and A.8 in Appendix A.2.2. Except for OPD_3 and SOP-based tests, the rejection rates for both tables are quite similar. Therefore, we first consider the performance for cross-dependence $+0.3$, and subsequently, we point out the differences in terms of the rejection rates for cross-dependence -0.3.

With regard to ordinal pattern-based statistics under cross-dependence $+0.3$, OPD_3 performs best for smaller sample sizes $n \in \{50, 100\}$, whereas OPD_2 has the largest empirical powers for larger sample sizes $n \in \{250, 500\}$. A modest performance can be reported for tests based on SOPs. Interestingly, the rejection rates of Chatterjee's correlation coefficient are inferior to every other statistic considered, whereas the best empirical powers are generally obtained for Pearson's and Spearman's correlations.

For cross-dependence -0.3, the rejection rates for OPD_3 clearly deteriorate, which is due to the fact that we have only considered positive OPD as discussed in Remark 5.1.2, that is, the co-occurence of the same ordinal patterns while disregarding inversed patterns. Also the SOP-based statistics perform worse for negative cross-dependence (especially τ'), although not to such an extent as OPD_3. The ordinal pattern-tests based on MOP_2, MOP_3, and OPD_2, by contrast, are at most slightly affected by the change in the sign of dependence such that OPD_2 and MOP_2 now show the best empirical powers within the ordinal pattern framework. Also Chatterjee's correlation coefficient is not affected by the sign of cross-dependence, but still it shows the worst power performance among all dependence tests.

7.3.4 Power Analysis under Outliers

In order to check the robustness of the power results reported in Section 7.3.3, we have again complemented our analysis by considering additional outliers, that is, we have considered cross-dependence (both ± 0.3) together with t_1-distributed marginals, and or with random contamination of normal variates by additive non-simultaneous outliers. Note that we do not consider simultaneous outliers in the context of existing cross-dependence as this would even intensify the dependence structure. The results are summarized in Table A.9.

Recalling that the SOP-based rejection rates for t_1-variates are not interpretable due to size problems, the relations for the respective empirical powers with regard to the ordinal pattern framework remain as in the case without outliers (compare Tables A.7 and A.8). In particular, we recognize that OPD_3 and the SOP-based tests show better performance for positive rather than for negative dependence, whereas MOP_2, MOP_3, and OPD_2 are (nearly) not affected by the sign of dependence. The latter statement also holds for Chatterjee's coefficient, but it again performs worst in most scenarios anyway.

Interestingly, it can be observed that all tests except Pearson show larger rejection rates for few strong outliers rather than for many weak ones. This is plausible as we are concerned with non-simultaneous outliers, where more outliers can break the dependence structure in more places of the time series. For the Pearson test, it is the other way round, i.e., already few strong outliers or t_1-marginals severely deteriorate the power, whereas the power was robust against only weak outliers.

While the performance of Pearson's test is severely affected by outliers, the ordinal pattern-based tests (as well as the test based on Spearman's correlation, which generally performs best in the scenarios of Table A.9) are rather robust against outliers. In particular, the rejection rates for MOP_2 and MOP_3 for the scenario **t1-t1** coincide with the results for **N-N** in Tables A.7 and A.8, which is plausible as the t_1-distribution is just a continuous distribution, although tending to produce extreme observations.

7.3.5 Power Analysis for Non-monotone Scenarios

While the dependence scenarios considered so far are rather simple (Gaussian copula with or without outliers), let us finally investigate some more demanding kinds of cross-dependence. More precisely, we consider non-monotone dependence structures, which are chosen such that the components have zero correlation:

$$X_{t,1} \sim \mathrm{N}(0,1), \qquad X_{t,2} = f(X_{t,1}) + \varepsilon_t^a,$$

where functions f are chosen such that $E(X\,f(X)) = 0$, while $\varepsilon_t^a \sim \mathrm{U}[-a,a]$, $a > 0$, denotes some noise. Let φ and Φ denote the density and distribution function of a standard normal distribution, respectively. Then, we consider data generating processes based on the following choices of f and a:

abs: $f(x) = |x|$ and $a = 0.25$;

square: $f(x) = x^2$ and $a = 1$;

cos: $f(x) = \cos(x)$ and $a = 0.25$;

asym1: $f(x) = \begin{cases} \frac{-\Phi(1)}{1-\Phi(1)} \cdot (x+1) & \text{if } x \leq -1, \\ x+1 & \text{if } x > -1, \end{cases}$ and $a = 1$;

asym2: $f(x) = \begin{cases} \frac{3\varphi(1)-\Phi(1)}{1-\Phi(1)} \cdot (x+1) + 3 & \text{if } x \leq -1 \\ x+1, & \text{if } x > -1, \end{cases}$ and $a = 0.5$.

The last two scenarios describe some form of asymmetric monotone dependence, where **asym1** is illustrated by Figure 7.1 (i). There, we can clearly observe similar or opposite/inverse behavior in terms of ups and downs. Note that for **asym1** and **asym2**, the property $E(X\,f(X)) = 0$ was shown by using [91].

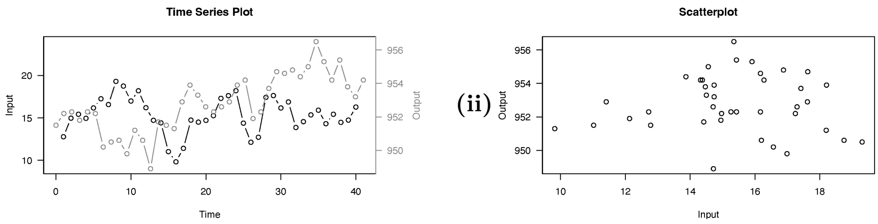
The obtained power values are summarized in Table A.10. Since we are concerned with non-monotone dependencies, it is not surprising that Pearson's and Spearman's correlation coefficients show a poor power performance. In fact, Pearson's correlation is not powerful throughout, while we note a certain power for Spearman's correlation in both asymmetric scenarios. Nevertheless, its power is worse than for the best ordinal pattern-based tests.

So let us next turn our attention to the ordinal pattern-based tests in Table A.10. Among the MOP-based tests, MOP_3 and OPD_3 exhibit better power than MOP_2 and OPD_2. This is plausible as MOP_3 and OPD_3 are based on three consecutive data points such that they can account for the apparent non-monotonicity (recall Figure 7.1 (i) and (ii)). For the SOP-based tests, the PE of the types performs better than the τ', and in the case of **asym2**, it even outperforms the MOP-based tests. Recall that τ' solely focuses on type 3 whereas the PE of types considers all three types jointly. So in case of the non-monotone dependencies, it seems that all three types provide useful information on the actual dependence structure, whereas in the previous scenarios, it was often sufficient to concentrate on type 3 only.

Let us conclude the discussion with a look at Chatterjee's coefficient. While it showed a very poor power performance in the "standard scenarios" of Sections 7.3.3 and 7.3.4, it excels in all non-monotone cases with a rejection rate of (almost) 100% for $n \geq 50$ (provided that the considered direction of dependence is correct). So Chatterjee's coefficient seems to have a highly specialized field of application, whereas the OPD_3-test, for example, is universally applicable, with reasonable power for both monotone and non-monotone dependencies.

All in all, dependence tests based on multivariate extensions of ordinal patterns constitute a valuable complement to the classical dependence tests. They perform reasonably well across various kinds of dependence, and they are robust with respect to outliers. The existing tests appear to have quite specialized fields of application (Pearson correlation only linear and without outliers, Spearman correlation only monotone, Chatterjee's coefficient only non-monotone), whereas the proposed ordinal pattern-based tests may be used to fill these gaps and constitute more of an all-rounder.

(i) (ii)



(iii)

	MOP$_2$	OPD$_2$	OPD$_3$	PE$_{\text{typ}}$	τ'	Pears	Spear	CCC$_{12}$	CCC$_{21}$
value	0.854	0.584	0.234	0.688	-0.282	0.001	-0.036	-0.058	0.021
decision	1	1	1	1	1	0	0	0	0

Figure 7.3: Cascade process data of Section 7.4.1: (i) time series plot and (ii) scatterplot of the first input and output variable. Test statistics and decisions in (iii), where "1" ("0") indicates (non-)rejection at 5% level.

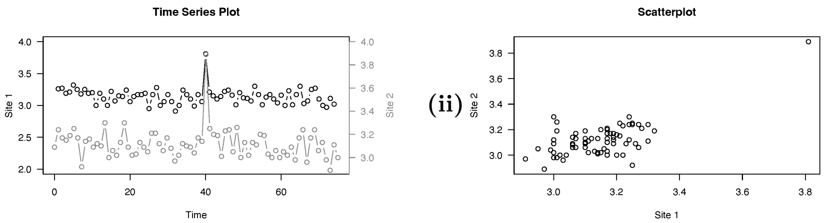
7.4 Data Applications

In the following, we apply our novel ordinal pattern-based test statistics as well as the previously considered competitors to two real-world data examples. The first data set in Section 7.4.1 consists of sequentially observed input and output variables, which makes it a candidate for a regression model after first having analyzed the cross-dependence structure. The second data set in Section 7.4.2 refers to measurements at three sites on a product wafer for consecutive lots.

7.4.1 Cascade Process Data

The first example are the cascade process data taken from [59, Table 11.5]. The eleven process variables have an inherent hierarchy, namely nine input variables and two output variables (with sample size $n = 40$). For illustrative purposes, we focus on the dependence between the first input and first output variable. As we shall see, these two variables exhibit a demanding dependence structure. In addition, this combination of variables contains only few ties, namely zero ties for ordinal patterns of length 2, and two ties for ordinal patterns of length 3. For SOPs and MOPs, these are captured by definition, whereas we have used the randomization predefined in the respective R-packages "`ordinalpattern`" and "`xicor`". The cascade data are plotted in Figure 7.3. While the scatterplot does not exhibit a pronounced dependence structure, one can note a simultaneous up and down behavior in the time series plot.

The computed statistics as well as the respective test decisions at level 5% are given in Figure 7.3 (iii), where '1' denotes rejection of the null. Note that we omitted the MOP$_3$-test due to the size problems for small sample sizes (here: $n = 40$), recall Section 7.3.1. It is immediately apparent that all the ordinal pattern-based statistics reject the null. By contrast, none of the competitors leads to a rejection of null of independence. These opposite test decisions can be explained by the aforementioned discrepancy between scatterplot and time series plot, where only the latter shows the coupled behavior of input and output variable. Hence, our novel ordinal pattern-based tests constitute a valuable complement of the existing tests for cross-dependence.

(i) (ii)


(iii)

		MOP_2	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
$n = 74$	value	0.955	0.285	0.300	0.981	-0.087	0.668	0.436	0.111	0.060
	decision	1	1	1	0	0	1	1	0	0
$n = 73$	value	0.958	0.276	0.291	0.983	-0.083	0.404	0.413	0.057	0.116
	decision	1	1	1	0	0	1	1	0	0

Figure 7.4: Product wafer data of Section 7.4.2: (i) time series plot and (ii) scatterplot of the first input and output variable. Test statistics and decisions in (iii), where "1" ("0") indicates (non-)rejection at 5% level.

7.4.2 Product Wafer Data

The second data set consists of trivariate observations, measured at three sites on a product wafer from a photolithographic process, for $n = 74$ consecutive lots, which is taken from [40, Appendix A]. Here, we have considered the observations for Sites 1 and 2, see the plots in Figure 7.4. However, the data exhibit a single simultaneous outlier in lot 40. It is well known that Pearson's correlation is sensitive with respect to outliers, also recall our simulation study in Sections 7.3.2 and 7.3.4, whereas our ordinal pattern-based tests turned out to be robust. Therefore, to further investigate the possible effect of an outlier on the different dependence measures, let us not only analyze the full data set, but also the subset where the outlier in Figure 7.4 is omitted (thus $n = 73$).

The obtained results are summarized in Figure 7.4 (iii), where we again omitted the MOP_3-test due to the small sample sizes. Interestingly, MOP_2, OPD_2, and OPD_3 as well as Pearson's and Spearman's correlations indicate significant dependence between these sites of a wafer, whereas both SOP-based tests, i.e., the PE of types and τ', as well as Chatterjee's coefficient do not. However, the decision to reject the null of independence appears reasonable as observations at different sites on a wafer are often correlated in practice, see [40, p. 346]. The most interesting feature of the product wafer data is the outlier at $t = 40$. If comparing the statistics for the reduced data set (row "$n = 73$" in (iii)) to those of the full data set (row "$n = 74$"), we observe that our statistics and Spearman's correlation are robust against the single outlier. This is not the case for Chatterjee's coefficient and especially for the Pearson correlation, although the test decision itself is not affected in this case. Altogether, the robustness against outliers for our novel ordinal pattern-based tests is an appealing property for applications in practice.

7.5 Interim Conclusion

In this chapter, we developed ordinal pattern-based methods for uncovering various forms of cross-dependence in sequentially observed random vectors. The main approaches are statistics related to MOPs, where we considered permutation entropy and OPD. Furthermore, we also showed that SOPs can be adapted to this framework, although they were originally developed for spatial data. For all considered statistics, we derived closed-form asymptotics, which allow for a computationally efficient implementation of the corresponding non-parametric tests for cross-dependence. In a comprehensive simulation study, we analyzed the size and power properties of our novel tests compared to popular competitors. It turned out that the ordinal pattern-based tests constitute a valuable complement to the classical dependence tests. They show an appealing performance across various kinds of dependence as well as robustness with respect to outliers. They are less specialized than the existing tests and constitute more of an all-rounder. Finally, we illustrated the practical application of our tests by two real-world data examples.

There are various directions for future research. In Section 7.2, we restricted to SOPs of length $d = 2$ in view of keeping the complexity at a feasible level. However, SOPs of length $d > 2$ might be better able to uncover non-monotone forms of dependence (in analogy to MOP_3 and OPD_3). In order to obtain manageable asymptotics, one could try to find an appropriate partition for such SOPs with $d > 2$ (see Remark 7.2.2). As a second topic for future research, one could try to combine the approach of MOPs from Section 7.1.1 with the generalized ordinal patterns considered by Weiß and Schnurr [89]. Doing this, it might be possible to get an improved power performance for discrete-valued bivariate data. Finally, a sequential implementation of our proposed MOP-tests appears desirable for practice, e.g., by developing corresponding control charts in analogy to Weiß and Testik [90]. This would allow for an online monitoring of the bivariate time series in order to detect changes in the cross-dependence structure.

However, our approaches are still limited to i.i.d pairs. Even though this has many practical application, e.g., in the field of manufacturing, this means that we cannot directly apply our theory to serially dependent time series. Further development of the theory in this respect could also be interesting for future research, although it is not yet clear to what extent it could be used for hypothesis testing in terms of size and power.

8 Conclusion and Outlook

Motivated by the advantages of ordinal patterns if compared to classical methods, we have made various contributions to the ordinal pattern analysis throughout this thesis. First and foremost is a comprehensive comparative analysis of different representations of classical ordinal patterns (Chapter 3), which is extended to multivariate generalizations of ordinal patterns in Chapter 6.

In Chapter 4 we have complemented the work of Caballero-Pintado et al. [21] by deriving limit theorems for the symbolic correlation integral, and hence also the Rényi-2 permutation entropy, for the short-range dependent case. Furthermore, we have provided a consistent estimator for the limit variance. All in all, this allows to test whether two data series stem from the same data generating processes.

However, challenges arise for the cases where the limit variance equals zero as it is, e.g., in the i.i.d. case (see Proposition 4.4.1). There, the limit distribution follows a one-point distribution. In this particular example, the results of Caballero-Pintado et al. [21] can be used, who considered the i.i.d. case with regard to another convergence rate, but caution is advised in the other cases. This highlights how little is known about ordinal pattern distributions stemming from classical models of time series analysis. To our knowledge, so far only Gaussian and ARMA-models haven been addressed (see Bandt and Shiha [11]). Therefore, this gap should become subject of future research. Another direction for future research is to use the above limit theorems to develop a test for time-reversibility or even Gaussianity.

In Chapter 5 we have corrected a result by Betken et al. [17] and have proved that OPD does not satisfy concordance ordering. Grothe et al. [32] have imposed this condition in their proposed axiomatic framework for multivariate measures of dependence between random vectors of same dimension. Arguably one could have formulated this axiom differently. As an example, we have considered supermodular ordering as a special case of concordance ordering. In this regard we have proved that our counterexamples already fulfill supermodular ordering. Hence, OPD does not respect supermodular ordering either. However, making an alteration to the assumptions of concordance ordering, we have shown that OPD satisfies ordering defined in terms of (conditional) cdfs and (conditional) survival functions, though these sets of assumptions are less intuitive and very difficult to check in practice.

All the orderings considered so far operate on the level of the actual values of the random vectors. Since OPD works on the level of ordinal patterns, that is, the relation of consecutive elements in terms of position and rank, here it seems to be more appropriate to consider an ordering which also operates on this level. Establishing such an order relation remains future research.

In Chapter 7, we have introduced a general framework for dependence tests between time series under the assumption of serial independence with regard to multivariate extensions of ordinal patterns. This also includes OPD as it can be embedded into the context of multivariate ordinal patterns. To this end, we have proved general limit theorems of multivariate pattern distributions. These encompass some existing results. Based on the derived asymptotics, we have proposed non-parametric tests for cross-dependence. We have done a comprehensive simulation study of our novel tests where we have analyzed their performance compared to popular competitors. There, we have shown that our ordinal pattern-based tests constitute a valuable complement to the classical dependence tests. In particular, we would like to emphasize that our tests are less specialized than the existing ones and hence, constitute more of an all-rounder.

Our contributions open up various directions for future research. First of all, consideration of SOPs of length $d > 2$ might be better suited to uncover forms of non-monotone dependence. In this regard, an appropriate partition of such SOPs would be reasonable in order to keep the asymptotics as well as the computational costs manageable. Furthermore, it would be interesting to explore the effects of generalized MOPs (as well as SOPs) in terms of generalized ordinal patterns. Finally, we suggest the development of corresponding control charts to allow for an online monitoring for changes in the cross-dependence structure.

A Appendix

A.1 Computation of Cross-Covariance Matrices

A.1.1 Ordinal Patterns of Length $d = 2$

To compute the cross-covariance matrix $\mathbf{\Sigma}$ for $d = 2$, we need to compute the 3×3 blocks according to (7.3). All blocks except the one at position $(3, 3)$ can be computed from the covariances $\sigma_{k,l}$ given in formula (13) of [87], namely

$$(\sigma_{k,l})_{k,l=1,2} \;=\; \frac{1}{12}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

by just applying formulae (7.4) to blocks $(1, 1)$ and $(2, 2)$, (7.5) to blocks $(1, 2)$ and $(2, 1)$, (7.6) to blocks $(1, 3)$ and $(2, 3)$, and (7.7) to blocks $(3, 1)$ and $(3, 2)$. The block at position $(3, 3)$, in turn, is computed via (7.8), where generally $p^{(k,l)}(0) = 1/d! \cdot (\delta_{kl} - 1/d!)$ with $\delta_{kl} := \mathbb{1}_{\{k=l\}}$ denoting the Kronecker delta, and where

$$\left(p^{(k,l)}(1)\right)_{k,l=1,2} \;=\; \begin{pmatrix} 1/6 & 1/3 \\ 1/3 & 1/6 \end{pmatrix}$$

for $d = 2$ according to formula (12) in [87]. The resulting expression for $\mathbf{\Sigma}$ is shown in Table A.1.

Table A.1: Entries of the matrix $144 \cdot \mathbf{\Sigma}$ for $d = 2$.

12	-12	0	0	6	6	-6	-6
-12	12	0	0	-6	-6	6	6
0	0	12	-12	6	-6	6	-6
0	0	-12	12	-6	6	-6	6
6	-6	6	-6	17	-11	-11	5
6	-6	-6	6	-11	17	5	-11
-6	6	6	-6	-11	5	17	-11
-6	6	-6	6	5	-11	-11	17

A.1.2 Ordinal Patterns of Length $d = 3$

The case $d = 3$ is computed in the same way as in Appendix A.1.1, but the involved expressions for $\sigma_{k,l}$ and $p^{(k,l)}(h)$ differ. According to Section III.B in [87], these are given by

$$\left(p^{(k,l)}(1)\right)_{k,l=1,\ldots,6} \;=\; \frac{1}{24}\begin{pmatrix} 1 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 1 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 1 \end{pmatrix},$$

by

$$\left(p^{(k,l)}(2)\right)_{k,l=1,\dots,6} = \frac{1}{120}\begin{pmatrix} 1 & 1 & 3 & 3 & 6 & 6 \\ 3 & 3 & 4 & 4 & 3 & 3 \\ 1 & 1 & 3 & 3 & 6 & 6 \\ 6 & 6 & 3 & 3 & 1 & 1 \\ 3 & 3 & 4 & 4 & 3 & 3 \\ 6 & 6 & 3 & 3 & 1 & 1 \end{pmatrix},$$

and by

$$(\sigma_{k,l})_{k,l=1,\dots,6} = \frac{1}{360}\begin{pmatrix} 46 & -23 & -23 & 7 & 7 & -14 \\ -23 & 28 & 10 & -20 & -2 & 7 \\ -23 & 10 & 28 & -2 & -20 & 7 \\ 7 & -20 & -2 & 28 & 10 & -23 \\ 7 & -2 & -20 & 10 & 28 & -23 \\ -14 & 7 & 7 & -23 & -23 & 46 \end{pmatrix}.$$

The resulting expression for $\mathbf{\Sigma}$ is shown in Table A.2.

Table A.2: Entries of the matrix $129600 \cdot \mathbf{\Sigma}$ for $d = 3$.

16560	-8280	-8280	2520	2520	-5040	0	0	0	0	0	0	2760	2760	2760	2760	2760	2760	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	420	420	420	420	420	420	420	420	420	420	420	420	-840	-840	-840	-840	-840	-840
-8280	10080	3600	-7200	-720	2520	0	0	0	0	0	0	-1380	-1380	-1380	-1380	-1380	-1380	1680	1680	1680	1680	1680	1680	600	600	600	600	600	600	-1200	-1200	-1200	-1200	-1200	-1200	-120	-120	-120	-120	-120	-120	420	420	420	420	420	420
-8280	3600	10080	-720	-7200	2520	0	0	0	0	0	0	-1380	-1380	-1380	-1380	-1380	-1380	600	600	600	600	600	600	1680	1680	1680	1680	1680	1680	-120	-120	-120	-120	-120	-120	-1200	-1200	-1200	-1200	-1200	-1200	420	420	420	420	420	420
2520	-7200	-720	10080	3600	-8280	0	0	0	0	0	0	420	420	420	420	420	420	-1200	-1200	-1200	-1200	-1200	-1200	-120	-120	-120	-120	-120	-120	1680	1680	1680	1680	1680	1680	600	600	600	600	600	600	-1380	-1380	-1380	-1380	-1380	-1380
2520	-720	-7200	3600	10080	-8280	0	0	0	0	0	0	420	420	420	420	420	420	-120	-120	-120	-120	-120	-120	-1200	-1200	-1200	-1200	-1200	-1200	600	600	600	600	600	600	1680	1680	1680	1680	1680	1680	-1380	-1380	-1380	-1380	-1380	-1380
-5040	2520	2520	-8280	-8280	16560	0	0	0	0	0	0	-840	-840	-840	-840	-840	-840	420	420	420	420	420	420	420	420	420	420	420	420	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	-1380	2760	2760	2760	2760	2760	2760
0	0	0	0	0	0	16560	-8280	-8280	2520	2520	-5040	2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840
0	0	0	0	0	0	-8280	10080	3600	-7200	-720	2520	-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420
0	0	0	0	0	0	-8280	3600	10080	-720	-7200	2520	-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420
0	0	0	0	0	0	2520	-7200	-720	10080	3600	-8280	420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380
0	0	0	0	0	0	2520	-720	-7200	3600	10080	-8280	420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380
0	0	0	0	0	0	-5040	2520	2520	-8280	-8280	16560	-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760
2760	-1380	-1380	420	420	-840	2760	-1380	-1380	420	420	-840	3567	-238	-238	31	31	-391	-238	-184	-445	139	-364	-283	-238	-445	-184	-364	139	-283	31	139	-364	805	-175	-13	31	-364	139	-175	805	-13	-391	-283	-283	-13	-13	148
2760	-1380	-1380	420	420	-840	-1380	1680	600	-1200	-120	420	-238	3153	220	-409	4	31	-445	-391	-211	-301	-166	139	-184	-391	67	-337	-166	-364	-364	-256	112	-67	193	805	139	-256	643	-121	193	-175	-283	-175	-229	40	-175	-13
2760	-1380	-1380	420	420	-840	-1380	600	1680	-120	-1200	420	-238	220	3153	4	-409	31	-184	67	-391	-166	-337	-364	-445	-211	-391	-166	-301	139	139	643	-256	193	-121	-175	-364	112	-256	193	-67	805	-283	-229	-175	-175	40	-13
2760	-1380	-1380	420	420	-840	420	-1200	-120	1680	600	-1380	31	-409	4	3153	220	-238	-364	-337	-166	-391	67	-184	139	-301	-166	-391	-211	-445	-175	-121	193	-256	643	139	805	-67	193	-256	112	-364	-13	40	-175	-175	-229	-283
2760	-1380	-1380	420	420	-840	420	-120	-1200	600	1680	-1380	31	4	-409	220	3153	-238	139	-166	-301	-211	-391	-445	-364	-166	-337	67	-391	-184	805	193	-67	112	-256	-364	-175	193	-121	643	-256	139	-13	-175	40	-229	-175	-283
2760	-1380	-1380	420	420	-840	-840	420	420	-1380	-1380	2760	-391	31	31	-238	-238	3567	-283	-364	139	-445	-184	-238	-283	139	-364	-184	-445	-238	-13	-175	805	-364	139	31	-13	805	-175	139	-364	31	148	-13	-13	-283	-283	-391
-1380	1680	600	-1200	-120	420	2760	-1380	-1380	420	420	-840	-238	-445	-184	-364	139	-283	3153	-391	-391	-256	-256	-175	220	-211	67	112	643	-229	-409	-301	-337	-67	-121	40	4	-166	-166	193	193	-175	31	139	-364	805	-175	-13
-1380	1680	600	-1200	-120	420	-1380	1680	600	-1200	-120	420	-184	-391	67	-337	-166	-364	-391	3262	-365	-229	-337	-256	67	-365	778	-301	309	112	-337	-229	-301	-31	-229	-67	-166	-337	309	-229	112	193	-364	-256	112	-67	193	805
-1380	1680	600	-1200	-120	420	-1380	600	1680	-120	-1200	420	-445	-211	-391	-166	-301	139	-391	-365	3262	-337	-229	-256	-211	472	-365	309	-247	643	-301	-247	-229	-229	-67	-121	-166	309	-337	112	-229	193	139	643	-256	193	-121	-175
-1380	1680	600	-1200	-120	420	420	-1200	-120	1680	600	-1380	139	-301	-166	-391	-211	-445	-256	-229	-337	3262	-365	-391	643	-247	309	-365	472	-211	-121	-67	-229	-229	-247	-301	193	-229	112	-337	309	-166	-175	-121	193	-256	643	139
-1380	1680	600	-1200	-120	420	420	-120	-1200	600	1680	-1380	-364	-166	-337	67	-391	-184	-256	-337	-229	-365	3262	-391	112	309	-301	778	-365	67	-67	-229	-31	-301	-229	-337	193	112	-229	309	-337	-166	805	193	-67	112	-256	-364
-1380	1680	600	-1200	-120	420	-840	420	420	-1380	-1380	2760	-283	139	-364	-184	-445	-238	-175	-256	-256	-391	-391	3153	-229	643	112	67	-211	220	40	-121	-67	-337	-301	-409	-175	193	193	-166	-166	4	-13	-175	805	-364	139	31
-1380	600	1680	-120	-1200	420	2760	-1380	-1380	420	420	-840	-238	-184	-445	139	-364	-283	220	67	-211	643	112	-229	3153	-391	-391	-256	-256	-175	4	-166	-166	193	193	-175	-409	-337	-301	-121	-67	40	31	-364	139	-175	805	-13
-1380	600	1680	-120	-1200	420	-1380	1680	600	-1200	-120	420	-445	-391	-211	-301	-166	139	-211	-365	472	-247	309	643	-391	3262	-365	-229	-337	-256	-166	-337	309	-229	112	193	-301	-229	-247	-67	-229	-121	139	-256	643	-121	193	-175
-1380	600	1680	-120	-1200	420	-1380	600	1680	-120	-1200	420	-184	67	-391	-166	-337	-364	67	778	-365	309	-301	112	-391	-365	3262	-337	-229	-256	-166	309	-337	112	-229	193	-337	-301	-229	-229	-31	-67	-364	112	-256	193	-67	805
-1380	600	1680	-120	-1200	420	420	-1200	-120	1680	600	-1380	-364	-337	-166	-391	67	-184	112	-301	309	-365	778	67	-256	-229	-337	3262	-365	-391	193	-229	112	-337	309	-166	-67	-31	-229	-229	-301	-337	805	-67	193	-256	112	-364
-1380	600	1680	-120	-1200	420	420	-120	-1200	600	1680	-1380	139	-166	-301	-211	-391	-445	643	309	-247	472	-365	-211	-256	-337	-229	-365	3262	-391	193	112	-229	309	-337	-166	-121	-229	-67	-247	-229	-301	-175	193	-121	643	-256	139
-1380	600	1680	-120	-1200	420	-840	420	420	-1380	-1380	2760	-283	-364	139	-445	-184	-238	-229	112	643	-211	67	220	-175	-256	-256	-391	-391	3153	-175	193	193	-166	-166	4	40	-67	-121	-301	-337	-409	-13	805	-175	139	-364	31
420	-1200	-120	1680	600	-1380	2760	-1380	-1380	420	420	-840	31	-364	139	-175	805	-13	-409	-337	-301	-121	-67	40	4	-166	-166	193	193	-175	3153	-391	-391	-256	-256	-175	220	67	-211	643	112	-229	-238	-184	-445	139	-364	-283
420	-1200	-120	1680	600	-1380	-1380	1680	600	-1200	-120	420	139	-256	643	-121	193	-175	-301	-229	-247	-67	-229	-121	-166	-337	309	-229	112	193	-391	3262	-365	-229	-337	-256	-211	-365	472	-247	309	643	-445	-391	-211	-301	-166	139
420	-1200	-120	1680	600	-1380	-1380	600	1680	-120	-1200	420	-364	112	-256	193	-67	805	-337	-301	-229	-229	-31	-67	-166	309	-337	112	-229	193	-391	-365	3262	-337	-229	-256	67	778	-365	309	-301	112	-184	67	-391	-166	-337	-364
420	-1200	-120	1680	600	-1380	420	-1200	-120	1680	600	-1380	805	-67	193	-256	112	-364	-67	-31	-229	-229	-301	-337	193	-229	112	-337	309	-166	-256	-229	-337	3262	-365	-391	112	-301	309	-365	778	67	-364	-337	-166	-391	67	-184
420	-1200	-120	1680	600	-1380	420	-120	-1200	600	1680	-1380	-175	193	-121	643	-256	139	-121	-229	-67	-247	-229	-301	193	112	-229	309	-337	-166	-256	-337	-229	-365	3262	-391	643	309	-247	472	-365	-211	139	-166	-301	-211	-391	-445
420	-1200	-120	1680	600	-1380	-840	420	420	-1380	-1380	2760	-13	805	-175	139	-364	31	40	-67	-121	-301	-337	-409	-175	193	193	-166	-166	4	-175	-256	-256	-391	-391	3153	-229	112	643	-211	67	220	-283	-364	139	-445	-184	-238
420	-120	-1200	600	1680	-1380	2760	-1380	-1380	420	420	-840	31	139	-364	805	-175	-13	4	-166	-166	193	193	-175	-409	-301	-337	-67	-121	40	220	-211	67	112	643	-229	3153	-391	-391	-256	-256	-175	-238	-445	-184	-364	139	-283
420	-120	-1200	600	1680	-1380	-1380	1680	600	-1200	-120	420	-364	-256	112	-67	193	805	-166	-337	309	-229	112	193	-337	-229	-301	-31	-229	-67	67	-365	778	-301	309	112	-391	3262	-365	-229	-337	-256	-184	-391	67	-337	-166	-364
420	-120	-1200	600	1680	-1380	-1380	600	1680	-120	-1200	420	139	643	-256	193	-121	-175	-166	309	-337	112	-229	193	-301	-247	-229	-229	-67	-121	-211	472	-365	309	-247	643	-391	-365	3262	-337	-229	-256	-445	-211	-391	-166	-301	139
420	-120	-1200	600	1680	-1380	420	-1200	-120	1680	600	-1380	-175	-121	193	-256	643	139	193	-229	112	-337	309	-166	-121	-67	-229	-229	-247	-301	643	-247	309	-365	472	-211	-256	-229	-337	3262	-365	-391	139	-301	-166	-391	-211	-445
420	-120	-1200	600	1680	-1380	420	-120	-1200	600	1680	-1380	805	193	-67	112	-256	-364	193	112	-229	309	-337	-166	-67	-229	-31	-301	-229	-337	112	309	-301	778	-365	67	-256	-337	-229	-365	3262	-391	-364	-166	-337	67	-391	-184
420	-120	-1200	600	1680	-1380	-840	420	420	-1380	-1380	2760	-13	-175	805	-364	139	31	-175	193	193	-166	-166	4	40	-121	-67	-337	-301	-409	-229	643	112	67	-211	220	-175	-256	-256	-391	-391	3153	-283	139	-364	-184	-445	-238
-840	420	420	-1380	-1380	2760	2760	-1380	-1380	420	420	-840	-391	-283	-283	-13	-13	148	31	-364	139	-175	805	-13	31	139	-364	805	-175	-13	-238	-445	-184	-364	139	-283	-238	-184	-445	139	-364	-283	3567	-238	-238	31	31	-391
-840	420	420	-1380	-1380	2760	-1380	1680	600	-1200	-120	420	-283	-175	-229	40	-175	-13	139	-256	643	-121	193	-175	-364	-256	112	-67	193	805	-184	-391	67	-337	-166	-364	-445	-391	-211	-301	-166	139	-238	3153	220	-409	4	31
-840	420	420	-1380	-1380	2760	-1380	600	1680	-120	-1200	420	-283	-229	-175	-175	40	-13	-364	112	-256	193	-67	805	139	643	-256	193	-121	-175	-445	-211	-391	-166	-301	139	-184	67	-391	-166	-337	-364	-238	220	3153	4	-409	31
-840	420	420	-1380	-1380	2760	420	-1200	-120	1680	600	-1380	-13	40	-175	-175	-229	-283	805	-67	193	-256	112	-364	-175	-121	193	-256	643	139	139	-301	-166	-391	-211	-445	-364	-337	-166	-391	67	-184	31	-409	4	3153	220	-238
-840	420	420	-1380	-1380	2760	420	-120	-1200	600	1680	-1380	-13	-175	40	-229	-175	-283	-175	193	-121	643	-256	139	805	193	-67	112	-256	-364	-364	-166	-337	67	-391	-184	139	-166	-301	-211	-391	-445	31	4	-409	220	3153	-238
-840	420	420	-1380	-1380	2760	-840	420	420	-1380	-1380	2760	148	-13	-13	-283	-283	-391	-13	805	-175	139	-364	31	-13	-175	805	-364	139	31	-283	139	-364	-184	-445	-238	-283	-364	139	-445	-184	-238	-391	31	31	-238	-238	3567

A.1.3 Spatial Ordinal Patterns of Dimension $d = 2$

Table A.3 shows the entries of the matrix $\mathbf{H}$, whereas Table A.4 provides the large sample approximation of $\tilde{\mathbf{\Sigma}}$ according to (7.12):

$$\tilde{\mathbf{\Sigma}} \;\approx\; \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top \;+\; \mathbf{H} + \mathbf{H}^\top - 2\mathbf{p}\mathbf{p}^\top .$$

Table A.3: Entries of matrix $720 \cdot \mathbf{H}$, where row horizontally before column, see [88].

	$\pi^{(1)}$	$\pi^{(2)}$	$\pi^{(3)}$	$\pi^{(4)}$	$\pi^{(5)}$	$\pi^{(6)}$	$\pi^{(7)}$	$\pi^{(8)}$	$\pi^{(9)}$	$\pi^{(10)}$	$\pi^{(11)}$	$\pi^{(12)}$	$\pi^{(13)}$	$\pi^{(14)}$	$\pi^{(15)}$	$\pi^{(16)}$	$\pi^{(17)}$	$\pi^{(18)}$	$\pi^{(19)}$	$\pi^{(20)}$	$\pi^{(21)}$	$\pi^{(22)}$	$\pi^{(23)}$	$\pi^{(24)}$
$\pi^{(1)}$	2	3	1	3	1	2	2	4	0	4	0	2	0	3	0	3	0	0	0	0	0	0	0	0
$\pi^{(2)}$	2	1	3	1	3	2	4	2	0	2	0	4	0	3	0	3	0	0	0	0	0	0	0	0
$\pi^{(3)}$	1	1	1	1	1	1	3	3	0	3	0	3	0	6	0	6	0	0	0	0	0	0	0	0
$\pi^{(4)}$	0	0	0	0	0	0	0	0	3	0	3	0	4	0	2	0	2	4	2	3	1	3	1	2
$\pi^{(5)}$	0	0	0	0	0	0	0	0	1	0	1	0	3	0	1	0	1	3	3	6	1	6	1	3
$\pi^{(6)}$	0	0	0	0	0	0	0	0	1	0	1	0	2	0	2	0	2	2	4	3	3	3	3	4
$\pi^{(7)}$	3	6	1	6	1	3	1	3	0	3	0	1	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(8)}$	4	3	3	3	3	4	2	2	0	2	0	2	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(9)}$	1	1	1	1	1	1	3	3	0	3	0	3	0	6	0	6	0	0	0	0	0	0	0	0
$\pi^{(10)}$	0	0	0	0	0	0	0	0	3	0	3	0	2	0	4	0	4	2	2	1	3	1	3	2
$\pi^{(11)}$	0	0	0	0	0	0	0	0	1	0	1	0	3	0	1	0	1	3	3	6	1	6	1	3
$\pi^{(12)}$	0	0	0	0	0	0	0	0	1	0	1	0	1	0	3	0	3	1	3	1	6	1	6	3
$\pi^{(13)}$	3	6	1	6	1	3	1	3	0	3	0	1	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(14)}$	3	1	6	1	6	3	3	1	0	1	0	3	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(15)}$	2	3	1	3	1	2	2	4	0	4	0	2	0	3	0	3	0	0	0	0	0	0	0	0
$\pi^{(16)}$	0	0	0	0	0	0	0	0	6	0	6	0	3	0	3	0	3	3	1	1	1	1	1	1
$\pi^{(17)}$	0	0	0	0	0	0	0	0	1	0	1	0	2	0	2	0	2	2	4	3	3	3	3	4
$\pi^{(18)}$	0	0	0	0	0	0	0	0	1	0	1	0	1	0	3	0	3	1	3	1	6	1	6	3
$\pi^{(19)}$	4	3	3	3	3	4	2	2	0	2	0	2	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(20)}$	3	1	6	1	6	3	3	1	0	1	0	3	0	1	0	1	0	0	0	0	0	0	0	0
$\pi^{(21)}$	2	1	3	1	3	2	4	2	0	2	0	4	0	3	0	3	0	0	0	0	0	0	0	0
$\pi^{(22)}$	0	0	0	0	0	0	0	0	6	0	6	0	3	0	3	0	3	3	1	1	1	1	1	1
$\pi^{(23)}$	0	0	0	0	0	0	0	0	3	0	3	0	4	0	2	0	2	4	2	3	1	3	1	2
$\pi^{(24)}$	0	0	0	0	0	0	0	0	3	0	3	0	2	0	4	0	4	2	2	1	3	1	3	2

Table A.4: Entries of matrix $2880 \cdot \tilde{\mathbf{\Sigma}}$, where $\tilde{\mathbf{\Sigma}}$ denotes the large-sample approximation to (7.12).

	$\pi^{(1)}$	$\pi^{(2)}$	$\pi^{(3)}$	$\pi^{(4)}$	$\pi^{(5)}$	$\pi^{(6)}$	$\pi^{(7)}$	$\pi^{(8)}$	$\pi^{(9)}$	$\pi^{(10)}$	$\pi^{(11)}$	$\pi^{(12)}$	$\pi^{(13)}$	$\pi^{(14)}$	$\pi^{(15)}$	$\pi^{(16)}$	$\pi^{(17)}$	$\pi^{(18)}$	$\pi^{(19)}$	$\pi^{(20)}$	$\pi^{(21)}$	$\pi^{(22)}$	$\pi^{(23)}$	$\pi^{(24)}$
$\pi^{(1)}$	121	5	−7	−3	−11	−7	5	17	−11	1	−15	−7	−3	9	−7	−3	−15	−15	1	−3	−7	−15	−15	−15
$\pi^{(2)}$	5	113	1	−11	−3	−7	25	5	−11	−7	−15	1	9	1	−3	−3	−15	−15	−3	−11	−11	−15	−15	−15
$\pi^{(3)}$	−7	1	113	−11	−11	−11	1	9	−11	−3	−15	−3	−11	33	−11	9	−15	−15	−3	9	−3	−15	−15	−15
$\pi^{(4)}$	−3	−11	−11	105	−15	−15	9	−3	1	−15	−3	−15	25	−11	5	−15	−7	1	5	1	−7	−3	−11	−7
$\pi^{(5)}$	−11	−3	−11	−15	105	−15	−11	−3	−7	−15	−11	−15	1	9	−7	−15	−11	−3	9	33	1	9	−11	−3
$\pi^{(6)}$	−7	−7	−11	−15	−15	105	−3	1	−7	−15	−11	−15	5	−3	1	−15	−7	−7	17	9	5	−3	−3	1
$\pi^{(7)}$	5	25	1	9	−11	−3	113	5	−3	−3	−15	−11	−11	1	−7	−11	−15	−15	−7	−3	1	−15	−15	−15
$\pi^{(8)}$	17	5	9	−3	−3	1	5	121	−3	−7	−15	−7	−3	−7	1	−11	−15	−15	−7	−11	−7	−15	−15	−15
$\pi^{(9)}$	−11	−11	−11	1	−7	−7	−3	−3	105	9	−11	1	−15	9	−15	33	−11	−11	−15	−15	−15	9	−3	−3
$\pi^{(10)}$	1	−7	−3	−15	−15	−15	−3	−7	9	105	−3	−15	5	−11	17	−15	1	−7	1	−7	5	−11	−3	−7
$\pi^{(11)}$	−15	−15	−15	−3	−11	−11	−15	−15	−11	−3	113	−11	−3	−15	−11	9	−7	1	−3	9	−11	33	1	9
$\pi^{(12)}$	−7	1	−3	−15	−15	−15	−11	−7	1	−15	−11	105	−7	−3	5	−15	−3	−11	5	1	25	−11	9	−3
$\pi^{(13)}$	−3	9	−11	25	1	5	−11	−3	−15	5	−3	−7	105	−11	−15	1	−7	−11	−15	−15	−15	−3	1	−7
$\pi^{(14)}$	9	1	33	−11	9	−3	1	−7	9	−11	−15	−3	−11	113	−3	−11	−15	−15	−11	−11	−3	−15	−15	−15
$\pi^{(15)}$	−7	−3	−11	5	−7	1	−7	1	−15	17	−11	5	−15	−3	105	9	−7	−3	−15	−15	−15	−3	−7	1
$\pi^{(16)}$	−3	−3	9	−15	−15	−15	−11	−11	33	−15	9	−15	1	−11	9	105	−3	−3	−7	−7	1	−11	−11	−11
$\pi^{(17)}$	−15	−15	−15	−7	−11	−7	−15	−15	−11	1	−7	−3	−7	−15	−7	−3	121	5	1	−3	−3	9	5	17
$\pi^{(18)}$	−15	−15	−15	1	−3	−7	−15	−15	−11	−7	1	−11	−11	−15	−3	−3	5	113	−3	−11	9	1	25	5
$\pi^{(19)}$	1	−3	−3	5	9	17	−7	−7	−15	1	−3	5	−15	−11	−15	−7	1	−3	105	−15	−15	−11	−7	−7
$\pi^{(20)}$	−3	−11	9	1	33	9	−3	−11	−15	−7	9	1	−15	−11	−15	−7	−3	−11	−15	105	−15	−11	−3	−11
$\pi^{(21)}$	−7	−11	−3	−7	1	5	1	−7	−15	5	−11	25	−15	−3	−15	1	−3	9	−15	−15	105	−11	−11	−3
$\pi^{(22)}$	−15	−15	−15	−3	9	−3	−15	−15	9	−11	33	−11	−3	−15	−3	−11	9	1	−11	−11	−11	113	1	−7
$\pi^{(23)}$	−15	−15	−15	−11	−11	−3	−15	−15	−3	−3	1	9	1	−15	−7	−11	5	25	−7	−3	−11	1	113	5
$\pi^{(24)}$	−15	−15	−15	−7	−3	1	−15	−15	−3	−7	9	−3	−7	−15	1	−11	17	5	−7	−11	−3	−7	5	121

A.2 Simulation Study

A.2.1 Analysis of Sizes

Tables A.5 and A.6 show the rejection rates for the scenarios described in Sections 7.3.1 and 7.3.2, respectively. The strongly oversized values, i.e., where the simulated size is ≥ 0.10, are indicated in italics.

Table A.5: Empirical sizes for the scenarios described in Section 7.3.1 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
N-N	50	0.056	*0.179*	0.051	0.047	0.054	0.042	0.053	0.049	0.047	0.045
	100	0.054	*0.128*	0.046	0.052	0.048	0.052	0.048	0.045	0.046	0.049
	250	0.051	0.066	0.050	0.050	0.047	0.048	0.051	0.052	0.050	0.054
	500	0.048	0.060	0.051	0.052	0.051	0.051	0.052	0.052	0.048	0.049
E1-E1	50	0.050	*0.178*	0.049	0.043	0.053	0.041	0.051	0.052	0.051	0.047
	100	0.053	*0.121*	0.045	0.048	0.049	0.052	0.050	0.050	0.049	0.053
	250	0.051	0.068	0.051	0.049	0.050	0.045	0.050	0.050	0.050	0.049
	500	0.049	0.059	0.052	0.053	0.050	0.054	0.050	0.051	0.049	0.053
P1-P1	50	0.051	*0.180*	0.050	0.047	0.055	0.040	0.049	0.052	0.044	0.048
	100	0.057	*0.127*	0.048	0.049	0.050	0.053	0.048	0.049	0.048	0.050
	250	0.050	0.067	0.051	0.050	0.050	0.047	0.051	0.051	0.050	0.047
	500	0.050	0.061	0.052	0.049	0.051	0.053	0.046	0.048	0.051	0.051
N-E1	50	0.052	*0.177*	0.053	0.045	0.058	0.041	0.053	0.051	0.047	0.050
	100	0.055	*0.129*	0.048	0.054	0.055	0.058	0.052	0.054	0.050	0.052
	250	0.054	0.067	0.050	0.048	0.065	0.053	0.052	0.049	0.049	0.049
	500	0.053	0.057	0.054	0.052	0.079	0.062	0.049	0.049	0.051	0.050
N-P1	50	0.051	*0.182*	0.048	0.045	0.060	0.044	0.048	0.045	0.046	0.046
	100	0.052	*0.130*	0.044	0.049	0.050	0.052	0.052	0.051	0.050	0.050
	250	0.053	0.068	0.052	0.049	0.048	0.047	0.047	0.051	0.048	0.050
	500	0.049	0.055	0.053	0.049	0.053	0.051	0.047	0.047	0.050	0.051
N-P5	50	0.051	*0.181*	0.048	0.042	0.057	0.045	0.050	0.047	0.045	0.046
	100	0.054	*0.130*	0.047	0.050	0.050	0.050	0.052	0.055	0.051	0.045
	250	0.051	0.071	0.052	0.048	0.050	0.044	0.047	0.049	0.044	0.049
	500	0.051	0.061	0.052	0.052	0.049	0.049	0.047	0.046	0.053	0.047

Table A.6: Empirical sizes for the scenarios described in Section 7.3.2 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
t1-t1	50	0.056	*0.179*	0.051	0.047	*0.229*	*0.193*	0.061	0.049	0.047	0.045
	100	0.054	*0.128*	0.046	0.052	*0.346*	*0.328*	0.044	0.045	0.046	0.049
	250	0.051	0.066	0.050	0.050	*0.510*	*0.468*	0.037	0.052	0.050	0.054
	500	0.048	0.060	0.051	0.052	*0.623*	*0.585*	0.029	0.052	0.048	0.049
$\text{AO}^{10}_{\text{nsim}}$	50	0.054	*0.186*	0.051	0.045	0.058	0.044	0.047	0.049	0.046	0.047
	100	0.049	*0.128*	0.045	0.051	0.045	0.052	0.049	0.051	0.047	0.051
	250	0.054	0.073	0.052	0.051	0.050	0.045	0.051	0.052	0.051	0.048
	500	0.051	0.058	0.052	0.053	0.047	0.048	0.050	0.051	0.050	0.049
$\text{AO}^{2}_{\text{nsim}}$	50	0.050	*0.175*	0.047	0.043	0.053	0.039	0.028	0.049	0.047	0.046
	100	0.054	*0.131*	0.047	0.050	0.051	0.058	0.051	0.054	0.047	0.047
	250	0.053	0.075	0.052	0.054	0.055	0.050	0.056	0.049	0.052	0.051
	500	0.053	0.060	0.051	0.053	0.052	0.051	0.050	0.053	0.051	0.052
$\text{AO}^{10}_{\text{sim}}$	50	0.065	*0.186*	0.062	0.072	0.058	0.043	0.085	0.075	0.058	0.053
	100	0.072	*0.150*	0.063	0.087	0.048	0.055	*0.128*	*0.109*	0.057	0.059
	250	*0.111*	*0.105*	*0.112*	*0.118*	0.065	0.065	*0.259*	*0.209*	0.068	0.063
	500	*0.157*	*0.133*	*0.165*	*0.166*	0.087	0.089	*0.455*	*0.366*	0.072	0.065
$\text{AO}^{2}_{\text{sim}}$	50	0.056	*0.181*	0.052	0.064	0.057	0.043	*1.000*	0.058	0.054	0.056
	100	0.061	*0.141*	0.052	0.077	0.051	0.055	*1.000*	0.076	0.099	*0.102*
	250	0.086	*0.103*	0.085	*0.118*	0.060	0.059	*1.000*	*0.147*	*0.198*	*0.199*
	500	*0.121*	*0.123*	*0.124*	*0.164*	0.074	0.084	*1.000*	*0.253*	*0.354*	*0.342*

A.2.2 Power Analysis

Tables A.7–A.10 show the rejection rates for the scenarios described in Sections 7.3.3–7.3.5, respectively. Those empirical powers, which are not interpretable due to the size problems indicated in Appendix A.2.1, are written in italic font. The best empirical powers among the remaining OP-statistics are printed in bold font.

Table A.7: Empirical powers for the scenarios described in Section 7.3.3 with cross-dependence +0.3 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
N-N	50	0.233	*0.341*	0.229	**0.266**	0.117	0.110	0.570	0.520	0.138	0.139
	100	0.428	*0.422*	0.411	**0.432**	0.172	0.210	0.868	0.833	0.193	0.202
	250	0.792	0.694	**0.800**	0.766	0.392	0.454	0.998	0.996	0.356	0.344
	500	0.975	0.953	**0.978**	0.960	0.699	0.758	1.000	1.000	0.543	0.535
E1-E1	50	0.233	*0.341*	0.229	**0.266**	0.121	0.116	0.449	0.520	0.138	0.139
	100	0.428	*0.422*	0.411	**0.432**	0.189	0.229	0.719	0.833	0.193	0.202
	250	0.792	0.694	**0.800**	0.766	0.402	0.470	0.978	0.996	0.356	0.344
	500	0.975	0.953	**0.978**	0.960	0.703	0.768	1.000	1.000	0.543	0.535
P1-P1	50	0.161	*0.273*	0.159	**0.186**	0.090	0.075	0.438	0.428	0.102	0.108
	100	0.277	*0.303*	0.261	**0.281**	0.118	0.142	0.725	0.717	0.140	0.137
	250	0.579	0.471	**0.586**	0.554	0.252	0.292	0.982	0.982	0.225	0.228
	500	0.856	0.772	**0.867**	0.813	0.478	0.555	1.000	1.000	0.349	0.341
N-E1	50	0.233	*0.341*	0.229	**0.266**	0.134	0.133	0.497	0.520	0.138	0.139
	100	0.428	*0.422*	0.411	**0.432**	0.222	0.280	0.793	0.833	0.193	0.202
	250	0.792	0.694	**0.800**	0.766	0.507	0.582	0.994	0.996	0.356	0.344
	500	0.975	0.953	**0.978**	0.960	0.822	0.875	1.000	1.000	0.543	0.535
N-P1	50	0.186	*0.297*	0.185	**0.216**	0.100	0.093	0.497	0.476	0.116	0.117
	100	0.349	*0.366*	0.331	**0.352**	0.150	0.188	0.807	0.784	0.156	0.167
	250	0.683	0.572	**0.693**	0.650	0.316	0.374	0.993	0.990	0.263	0.284
	500	0.933	0.885	**0.940**	0.901	0.604	0.676	1.000	1.000	0.418	0.440
N-P5	50	0.222	*0.334*	0.218	**0.260**	0.106	0.102	0.563	0.521	0.135	0.139
	100	0.413	*0.418*	0.398	**0.427**	0.169	0.210	0.868	0.834	0.191	0.195
	250	0.776	0.680	**0.785**	0.746	0.369	0.432	0.998	0.996	0.331	0.335
	500	0.972	0.947	**0.975**	0.953	0.686	0.745	1.000	1.000	0.517	0.520

Table A.8: Empirical powers for the scenarios described in Section 7.3.3 with cross-dependence −0.3 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
N-N	50	0.233	*0.341*	**0.241**	0.083	0.124	0.115	0.570	0.520	0.139	0.138
	100	**0.428**	*0.422*	0.415	0.243	0.177	0.219	0.868	0.833	0.202	0.193
	250	0.792	0.694	**0.799**	0.588	0.357	0.393	0.998	0.996	0.344	0.356
	500	0.975	0.953	**0.978**	0.896	0.642	0.686	1.000	1.000	0.535	0.543
E1-E1	50	0.233	*0.341*	**0.241**	0.083	0.123	0.110	0.372	0.520	0.139	0.138
	100	**0.428**	*0.422*	0.415	0.243	0.178	0.213	0.702	0.833	0.202	0.193
	250	0.792	0.694	**0.799**	0.588	0.356	0.389	0.982	0.996	0.344	0.356
	500	0.975	0.953	**0.978**	0.896	0.639	0.684	1.000	1.000	0.535	0.543
P1-P1	50	**0.161**	*0.269*	0.160	0.061	0.100	0.094	0.400	0.418	0.104	0.107
	100	**0.275**	*0.301*	0.261	0.154	0.124	0.165	0.712	0.715	0.138	0.136
	250	0.569	0.459	**0.577**	0.386	0.239	0.285	0.981	0.982	0.222	0.224
	500	0.859	0.777	**0.867**	0.719	0.433	0.530	1.000	1.000	0.334	0.339
N-E1	50	0.233	*0.341*	**0.241**	0.083	0.116	0.097	0.492	0.520	0.139	0.138
	100	**0.428**	*0.422*	0.415	0.243	0.156	0.182	0.794	0.833	0.202	0.193
	250	0.792	0.694	**0.799**	0.588	0.330	0.312	0.993	0.996	0.344	0.356
	500	0.975	0.953	**0.978**	0.896	0.601	0.573	1.000	1.000	0.535	0.543
N-P1	50	0.195	*0.307*	**0.201**	0.072	0.106	0.103	0.497	0.476	0.115	0.129
	100	**0.350**	*0.367*	0.336	0.199	0.147	0.184	0.807	0.784	0.161	0.179
	250	0.682	0.567	**0.691**	0.480	0.280	0.319	0.994	0.992	0.257	0.288
	500	0.934	0.880	**0.941**	0.822	0.515	0.588	1.000	1.000	0.410	0.441
N-P5	50	0.228	*0.334*	**0.232**	0.085	0.116	0.110	0.562	0.521	0.137	0.139
	100	**0.424**	*0.420*	0.409	0.236	0.165	0.210	0.865	0.832	0.198	0.199
	250	0.771	0.674	**0.779**	0.565	0.343	0.376	0.997	0.995	0.323	0.332
	500	0.974	0.942	**0.977**	0.889	0.618	0.663	1.000	1.000	0.512	0.526

Table A.9: Empirical powers for the scenarios described in Section 7.3.4 with cross-dependencies ± 0.3 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
t1-t1	50	0.233	*0.341*	0.229	**0.266**	*0.398*	*0.400*	0.172	0.520	0.138	0.139
+0.3	100	0.428	*0.422*	0.411	**0.432**	*0.581*	*0.617*	0.183	0.833	0.193	0.202
	250	0.792	0.694	**0.800**	0.766	*0.827*	*0.854*	0.189	0.996	0.356	0.344
	500	0.975	0.953	**0.978**	0.960	*0.951*	*0.964*	0.199	1.000	0.543	0.535
$\text{AO}^{10}_{\text{nsim}}$	50	0.161	*0.266*	0.157	**0.185**	0.088	0.078	0.349	0.336	0.100	0.101
+0.3	100	0.277	*0.302*	0.259	**0.286**	0.118	0.145	0.617	0.608	0.124	0.122
	250	0.565	0.450	**0.572**	0.553	0.247	0.290	0.944	0.945	0.195	0.192
	500	0.850	0.763	**0.862**	0.830	0.471	0.539	0.999	0.998	0.293	0.295
$\text{AO}^{2}_{\text{nsim}}$	50	0.202	*0.306*	0.199	**0.237**	0.103	0.097	0.053	0.453	0.121	0.133
+0.3	100	0.369	*0.379*	0.351	**0.390**	0.155	0.190	0.112	0.767	0.183	0.176
	250	0.727	0.618	**0.736**	0.707	0.350	0.403	0.274	0.991	0.318	0.321
	500	0.952	0.914	**0.956**	0.933	0.629	0.702	0.549	1.000	0.492	0.484
t1-t1	50	0.233	*0.341*	**0.241**	0.083	*0.206*	*0.135*	0.172	0.520	0.139	0.138
−0.3	100	**0.428**	*0.422*	0.415	0.243	*0.323*	*0.230*	0.183	0.833	0.202	0.193
	250	0.792	0.694	**0.799**	0.588	*0.542*	*0.354*	0.189	0.996	0.344	0.356
	500	0.975	0.953	**0.978**	0.896	*0.737*	*0.475*	0.199	1.000	0.535	0.543
$\text{AO}^{10}_{\text{nsim}}$	50	0.157	*0.273*	**0.164**	0.061	0.092	0.087	0.347	0.335	0.099	0.096
−0.3	100	**0.276**	*0.298*	0.263	0.151	0.115	0.149	0.612	0.606	0.127	0.124
	250	0.568	0.453	**0.579**	0.372	0.236	0.264	0.947	0.946	0.195	0.197
	500	0.852	0.762	**0.862**	0.687	0.417	0.484	0.999	0.999	0.295	0.294
$\text{AO}^{2}_{\text{nsim}}$	50	0.203	*0.305*	**0.210**	0.078	0.108	0.104	0.073	0.450	0.131	0.120
−0.3	100	**0.369**	*0.377*	0.356	0.198	0.150	0.192	0.150	0.765	0.178	0.182
	250	0.727	0.620	**0.736**	0.514	0.321	0.358	0.352	0.991	0.318	0.308
	500	0.950	0.914	**0.957**	0.844	0.568	0.633	0.631	1.000	0.483	0.492

Table A.10: Empirical powers for the scenarios described in Section 7.3.5 for varying sample sizes n and nominal level 5%.

	n	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	OPD_2	OPD_3	PE_{typ}	τ'	Pears	Spear	CCC_{12}	CCC_{21}
abs	50	0.081	*0.861*	0.081	**0.192**	0.110	0.054	0.253	0.140	1.000	0.608
	100	0.087	*0.996*	0.076	**0.280**	0.154	0.060	0.245	0.138	1.000	0.834
	250	0.080	**1.000**	0.082	0.509	0.346	0.054	0.255	0.141	1.000	0.991
	500	0.081	**1.000**	0.085	0.750	0.625	0.062	0.249	0.142	1.000	1.000
square	50	0.075	*0.587*	0.071	**0.132**	0.068	0.053	0.346	0.124	0.994	0.361
	100	0.075	*0.844*	0.067	**0.174**	0.078	0.074	0.343	0.120	1.000	0.554
	250	0.070	**0.999**	0.071	0.316	0.121	0.088	0.352	0.124	1.000	0.835
	500	0.073	**1.000**	0.075	0.497	0.195	0.132	0.351	0.126	1.000	0.976
cos	50	0.084	*0.750*	0.083	**0.172**	0.089	0.054	0.277	0.136	1.000	0.490
	100	0.081	*0.969*	0.072	**0.233**	0.107	0.070	0.263	0.131	1.000	0.705
	250	0.081	**1.000**	0.079	0.414	0.203	0.074	0.274	0.134	1.000	0.950
	500	0.074	**1.000**	0.076	0.659	0.362	0.091	0.271	0.138	1.000	0.998
asym1	50	0.464	*0.746*	0.456	**0.719**	0.364	0.283	0.334	0.652	0.999	0.876
	100	0.754	*0.958*	0.742	**0.935**	0.654	0.586	0.325	0.888	1.000	0.992
	250	0.983	**1.000**	0.985	**1.000**	0.981	0.945	0.327	0.998	1.000	1.000
	500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.999	0.320	1.000	1.000	1.000
asym2	50	0.417	*0.947*	0.405	0.858	**0.969**	0.173	0.183	0.329	1.000	1.000
	100	0.689	*0.999*	0.678	0.984	**1.000**	0.327	0.179	0.468	1.000	1.000
	250	0.963	**1.000**	0.967	**1.000**	**1.000**	0.611	0.185	0.732	1.000	1.000
	500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.869	0.181	0.927	1.000	1.000

Bibliography

[1] José M. Amigó. The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized. *Physica D: Nonlinear Phenomena*, 241(7):789–793, 2012.

[2] José M. Amigó and Osvaldo A. Rosso. Ordinal methods: Concepts, applications, new developments, and challenges – in memory of Karsten Keller (1961–2022). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(8), 2023.

[3] José M. Amigó, Karsten Keller, and Valentina A. Unakafova. Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2034):20140091, 2015.

[4] José M. Amigó, Sámuel G. Balogh, and Sergio Hernández. A brief review of generalized entropies. *Entropy*, 20(11):813, 2018.

[5] José M. Amigó, Matthew B. Kennel, and Ljupco Kocarev. The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. *Physica D: Nonlinear Phenomena*, 210(1):77–95, 2005.

[6] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001.

[7] Christoph Bandt. Small order patterns in big time series: A practical guide. *Entropy*, 21(6):613, 2019.

[8] Christoph Bandt. Order patterns, their variation and change points in financial time series and Brownian motion. *Statistical Papers*, 61(4):1565–1588, 2020.

[9] Christoph Bandt. Statistics and contrasts of order patterns in univariate time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(3), 2023.

[10] Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88(17):174102, 2002.

[11] Christoph Bandt and Faten Shiha. Order patterns in time series. *Journal of Time Series Analysis*, 28(5):646–665, 2007.

[12] Christoph Bandt and Katharina Wittfeld. Two new parameters for the ordinal analysis of images. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(4), 2023.

[13] Christoph Bandt, Gerhard Keller, and Bernd Pompe. Entropy of interval maps via permutations. *Nonlinearity*, 15(5):1595, 2002.

[14] Sebastian Berger, Andrii Kravtsiv, Gerhard Schneider, and Denis Jordan. Teaching ordinal patterns to a computer: Efficient encoding algorithms based on the Lehmer code. *Entropy*, 21(10), 2019.

[15] Annika Betken and Alexander Schnurr. Depth patterns. Preprint. arXiv:2401.13532, 2024.

[16] Annika Betken, Jannis Buchsteiner, Herold Dehling, Ines Münker, Alexander Schnurr, and Jeannette H.C. Woerner. Ordinal patterns in long-range dependent time series. *Scandinavian Journal of Statistics*, 48(3):969–1000, 2021.

[17] Annika Betken, Herold Dehling, Ines Nüßgen, and Alexander Schnurr. Ordinal pattern dependence as a multivariate dependence measure. *Journal of Multivariate Analysis*, 186:104798, 2021.

[18] Svetlana Borovkova, Robert Burton, and Herold Dehling. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society*, 353(11):4261–4318, 2001.

[19] Richard C. Bradley. *Introduction to strong mixing conditions*, volume 1. Kendrick, Heber City, Utah, 2007.

[20] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1987.

[21] M. Victoria Caballero-Pintado, Mariano Matilla-García, and Manuel Ruiz Marín. Symbolic correlation integral. *Econometric Reviews*, 38(5):533–556, 2019.

[22] Sourav Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, 2021.

[23] Rudolf Clausius. *The mechanical theory of heat*. McMillan and Co., 1865.

[24] James Davidson. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

[25] Robert M. De Jong and James Davidson. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2):407–423, 2000.

[26] Herold Dehling, Daniel Vogel, Martin Wendler, and Dominik Wied. Testing for changes in Kendall's tau. *Econometric Theory*, 33(6):1352–1386, 2017.

[27] Richard M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2nd edition, 2002.

[28] Svenja Fischer, Andreas Schumann, and Alexander Schnurr. Ordinal pattern dependence between hydrological time series. *Journal of Hydrology*, 548:536–551, 2017.

[29] Svenja Fischer, Marco Oesting, and Alexander Schnurr. Multivariate motion patterns and applications to rainfall radar data. *Stochastic Environmental Research and Risk Assessment*, 2023.

[30] Gery Geenens. Towards a universal representation of statistical dependence. Preprint. arXiv:2302.08151, 2023.

[31] Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9(1-2):189–208, 1983.

[32] Oliver Grothe, Julius Schnieders, and Johan Segers. Measuring association and dependence between random vectors. *Journal of Multivariate Analysis*, 123:96–110, 2014.

[33] Tim Gutjahr. *Determining the Kolmogorov–Sinai entropy by using ordinal patterns.* PhD thesis, University of Lübeck, 2021.

[34] Paul R. Halmos. The Theory of Unbiased Estimation. *The Annals of Mathematical Statistics*, 17(1):34 – 43, 1946.

[35] Wassily Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325, 1948.

[36] Wassily Hoeffding. The strong law of large numbers for U-statistics. Technical report, North Carolina State University. Dept. of Statistics, 1961. Mimeograph Series No. 302.

[37] I. A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382, 1962.

[38] Harry Joe. Multivariate concordance. *Journal of multivariate analysis*, 35(1):12–30, 1990.

[39] Harry Joe. *Multivariate Models and Dependence Concepts.* Chapman and Hall, London, 1997.

[40] Madhukar Joshi and Kimberley Sprague. *24. Obtaining and Using Statistical Process Control Limits in the Semiconductor Industry*, pages 337–356. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[41] Olav Kallenberg. *Foundations of modern probability*, volume 99 of *Probability Theory and Stochastic Modelling.* Springer Cham, 3rd edition, 2021.

[42] Karsten Keller and Mathieu Sinn. Ordinal analysis of time series. *Physica A: Statistical Mechanics and its Applications*, 356(1):114–120, 2005.

[43] Karsten Keller and Mathieu Sinn. Kolmogorov–Sinai entropy from the ordinal viewpoint. *Physica D: Nonlinear Phenomena*, 239(12):997–1000, 2010.

[44] Karsten Keller, Mathieu Sinn, and Jan Emonds. Time series from the ordinal viewpoint. *Stochastics and Dynamics*, 07(02):247–272, 2007.

[45] Karsten Keller, Anton M. Unakafov, and Valentina A. Unakafova. On the relation of KS entropy and permutation entropy. *Physica D: Nonlinear Phenomena*, 241(18):1477–1481, 2012.

[46] Karsten Keller, Teresa Mangold, Inga Stolz, and Jenna Werner. Permutation entropy: New ideas and challenges. *Entropy*, 19(3):134, 2017.

[47] Aleksandr Ja. Khinchin. *Mathematical foundations of information theory.* Dover, New York, 1957.

[48] Kalimuthu Krishnamoorthy. *Handbook of statistical distributions with applications.* CRC Press, 2016.

[49] Zhenhu Liang, Yinghua Wang, Xue Sun, Duan Li, Logan J. Voss, Jamie W. Sleigh, Satoshi Hagihira, and Xiaoli Li. EEG entropy measures in anesthesia. *Frontiers in computational neuroscience*, 9:16, 2015.

[50] M. Loève. *Probability Theory I.* Graduate Texts in Mathematics 45. Springer New York, New York, NY, 4th edition, 1977.

[51] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis.* Springer, Berlin, Heidelberg, 1st edition, 2005.

[52] José A. F Machado and J. M. C. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.

[53] Nathaniel F. G. Martin and James W. England. *Entropy and Information*, page 51–116. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1984.

[54] Mariano Matilla-García. A non–parametric test for independence based on symbolic dynamics. *Journal of Economic Dynamics and Control*, 31(12):3889–3903, 2007.

[55] Mariano Matilla-García and Manuel Ruiz Marín. A non-parametric independence test using permutation entropy. *Journal of Econometrics*, 144(1):139–155, 2008.

[56] Mariano Matilla-García, Manuel Ruiz Marín, and Mohammed I. Dore. A permutation entropy based test for causality: The volume–stock price relation. *Physica A: Statistical Mechanics and its Applications*, 398:280–288, 2014.

[57] Mariano Matilla-García, José Miguel Rodríguez, and Manuel Ruiz Marín. A symbolic test for testing independence between time series. *Journal of Time Series Analysis*, 31 (2):76–85, 2010.

[58] Marisa Mohr, Florian Wilhelm, Mattis Hartwig, Ralf Möller, and Karsten Keller. New approaches in ordinal pattern representations for multivariate time series. In *The Thirty-Third International Flairs Conference*, 2020.

[59] C. Montgomery, Douglas. *Introduction to Statistical Quality Control.* Wiley, Hoboken, NJ, 6th edition, 2009.

[60] Alfred Müller and Marco Scarsini. Some remarks on the supermodular order. *Journal of multivariate analysis*, 73(1):107–119, 2000.

[61] Alfred Müller and Dietrich Stoyan. *Comparison methods for stochastic models and risks.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, 2002.

[62] Ines Nüßgen and Alexander Schnurr. Ordinal pattern dependence in the context of long-range dependence. *Entropy*, 23(6), 2021.

[63] Marco Oesting and Alexander Schnurr. Ordinal patterns in clusters of subsequent extremes of regularly varying time series. *Extremes*, 23(4):521–545, 2020.

[64] Jagdish K. Patel and Campbell B. Read. *Handbook of the normal distribution.* Statistics 150. Dekker, New York, 2nd edition, 1996.

[65] Karl Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[66] Vladas Pipiras and Murad S. Taqqu. *Long-range dependence and self-similarity.* Cambridge series in statistical and probabilistic mathematics 45. Cambridge University Press, Cambridge, 2017.

[67] José C. Principe. *Information theoretic learning: Renyi's entropy and kernel perspectives.* Information Science and Statistics. Springer New York, NY, 2010.

[68] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press, 1961.

[69] Friedrich Schmid, Rafael Schmidt, Thomas Blumentritt, Sandra Gaißer, and Martin Ruppert. Copula-based measures of multivariate association. In *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, pages 209–236. Springer, 2010.

[70] Alexander Schnurr. An ordinal pattern approach to detect and to model leverage effects and dependence structures between financial time series. *Statistical Papers*, 55:919–931, 2014.

[71] Alexander Schnurr and Herold Dehling. Testing for structural breaks via ordinal pattern dependence. *Journal of the American Statistical Association*, 112(518):706–720, 2017.

[72] Alexander Schnurr and Svenja Fischer. Generalized ordinal patterns allowing for ties and their applications in hydrology. *Computational Statistics & Data Analysis*, 171:107472, 2022.

[73] Alexander Schnurr and Svenja Fischer. An ordinal procedure to detect change points in the dependence structure between non–stationary time series. *Engineering Proceedings*, 18(1), 2022.

[74] Alexander Schnurr and Angelika Silbernagel. A comparison of different representations of ordinal patterns and their usability in data analysis. In *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–7. IEEE, 2023.

[75] Alexander Schnurr, Angelika Silbernagel, and Manuel Ruiz Marín. Limit theorems for the symbolic correlation integral and the Rényi-2 entropy under short-range dependence. Preprint, 2024+.

[76] P. K. Sen. *Introduction to Hoeffding (1948) A Class of Statistics with Asymptotically Normal Distribution*, pages 299–307. Springer New York, 1992.

[77] Robert J Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 1980.

[78] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[79] Angelika Silbernagel. A comparison between multivariate measures of dependence. Unpublished master's thesis, Ruhr-University Bochum, Bochum, Germany, 2022.

[80] Angelika Silbernagel and Alexander Schnurr. Ordinal pattern dependence and multivariate measures of dependence. *Journal of Multivariate Analysis*, 203:105337, 2024.

[81] Angelika Silbernagel and Alexander Schnurr. An overview of various applications of ordinal patterns in data analysis and mathematical statistics. In *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–5. IEEE, 2024.

[82] Angelika Silbernagel, Christian H. Weiß, and Alexander Schnurr. Non–parametric tests for cross–dependence based on multivariate extensions of ordinal patterns. Preprint, 2024+.

[83] Mathieu Sinn and Karsten Keller. Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments. *Computational Statistics & Data Analysis*, 55(4): 1781–1790, 2011.

[84] W. Y. Tan. On the distribution of quadratic forms in normal random variables. *Canadian Journal of Statistics*, 5(2):241–250, 1977.

[85] Valentina A. Unakafova and Karsten Keller. Efficiently measuring complexity on the basis of real-world data. *Entropy*, 15(10):4392–4415, 2013.

[86] Boto von Querenburg and Gunter Bengel. *Mengentheoretische Topologie*, volume 3. Springer, 1973.

[87] Christian H. Weiß. Non–parametric tests for serial dependence in time series based on asymptotic implementations of ordinal–pattern statistics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(9):093107, 2022.

[88] Christian H. Weiß and Hee-Young Kim. Using spatial ordinal patterns for non-parametric testing of spatial dependence. *Spatial Statistics*, 59:100800, 2024.

[89] Christian H. Weiß and Alexander Schnurr. Generalized ordinal patterns in discrete–valued time series: Non–parametric testing for serial dependence. *Journal of Nonparametric Statistics*, pages 1–27, 2023.

[90] Christian H. Weiß and Murat Caner Testik. Nonparametric control charts for monitoring serial dependence based on ordinal patterns. *Technometrics*, pages 1–11, 2023.

[91] Robert L. Winkler, Gary M. Roodman, and Robert R. Britney. The determination of partial moments. *Management Science*, 19(3):290–296, 1972.

[92] L. Zunino, D. G. Pérez, A. Kowalski, M. T. Martín, M. Garavaglia, A. Plastino, and O. A. Rosso. Fractional Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy. *Physica A: Statistical Mechanics and its Applications*, 387(24):6057–6068, 2008.

Notation

Basics and Conventions

$\mathbb{N}$, $\mathbb{N}_0$	1, 2, 3, ..., $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$]a,b[\,,]a,b]\,,[a,b[\,,[a,b]$	open, left-open, right-open, closed interval
$\mathbb{R}^+$, $\mathbb{R}$, $\overline{\mathbb{R}}$	$]0,\infty[$, $]-\infty,\infty[$, $[-\infty,\infty]$
$n!$	n factorial
log	natural logarithm
$a_k \in \mathcal{O}(b_k)$	$\limsup_{k\to\infty} \frac{a_k}{b_k} < \infty$, Landau Big O notation
$\lfloor a \rfloor$	$\max\{n \in \mathbb{Z} : n \leq a\}$
a^+	$\max\{0, a\}$
$\lvert a \rvert$	absolute value of $a \in \mathbb{R}$
$\lvert A \rvert$	cardinality of a set A
$\subset$	subset (incl. "=")
$\mathbb{1}_A$	indicator function of set A
$\delta_{kl} = \mathbb{1}_{\{k=l\}}$	Kronecker delta
x	value x
$\mathbf{x} = (x_1, \ldots, x_d)$	vector $\mathbf{x}$
$\mathbf{x} \wedge \mathbf{y}$	$(\min\{x_1,y_1\}, \ldots, \min\{x_d,y_d\})$
$\mathbf{x} \vee \mathbf{y}$	$(\max\{x_1,y_1\}, \ldots, \max\{x_d,y_d\})$
$\mathbf{0}_{p\times r}$	$(p \times r)$-zero matrix
$\lVert\cdot\rVert$	norm

Probabilistic Notation

$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
$(S, \mathcal{S})$	state space
$\mathcal{A} \subset \mathcal{F}$	sub-σ-algebra
$\sigma(\mathcal{A}), \sigma(Z)$	generated σ-algebra
$\mathcal{A}_k^l$	$\sigma(Z_k, \ldots, Z_l) \subset \mathcal{F}$
X, Y, Z	random variable
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	random vector
$L_p := \{\mathbf{X} \text{ rv} : \mathbb{E}\,\lVert\mathbf{X}\rVert^p < \infty\}$	set of p-integrable random variables
$\lVert\mathbf{X}\rVert_p := (\mathbb{E}\,\lVert\mathbf{X}\rVert^p)^{1/p}$	L_p-norm
F, G	cumulative distribution function (cdf)
$\overline{F}, \overline{G}$	survival function
$f = f_X$	density function of X
$N(\mu, \sigma^2)$	normal distribution
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate normal distribution
Q_d	quadratic form distributed random variable
χ_r^2	χ^2-distributed random variable with r degrees of freedom
$\perp\!\!\!\perp$	independence
$p \in [0,1]$	probability
$\mathbf{p} = (p^{(1)}, \ldots, p^{(d)})^\top$	probability mass function
$\overset{D}{=}$	equality in distribution
$\xrightarrow{D}$, $\xrightarrow{\mathbb{P}}$	convergence in distribution, probability

Time Series Analysis

$(X_t)_{t\in T}, (Z_t)_{t\in T}$	stochastic process/time series
$(x_t)_{t\in T}, (X_t)_{t\in T}(\omega)$	data, realizations of $(X_t)_{t\in T}$
$\gamma_X(\cdot)$	autocovariance function of $(X_t)_{t\in\mathbb{Z}}$ weakly stationary

Ordinal Pattern Analysis

π	ordinal pattern
d	length/order/embedding dimension of ordinal pattern
S_d	set consisting of ordinal patterns of length d (symmetric group)
$\Pi : \mathbb{R}^d \to S_d$	function which assigns to each vector its ordinal pattern
Δ	delay parameter
$p^{(j)}$	probability of j-th ordinal pattern
$p_{j,k}$	joint ordinal pattern probability
$\mathrm{OPD}_d(\mathbf{X}, \mathbf{Y})$	ordinal pattern dependence of $\mathbf{X}, \mathbf{Y}$
$\boldsymbol{\pi}$	multivariate ordinal pattern
$\boldsymbol{\Pi}$	spatial ordinal pattern
$\mathcal{S}$	set consisting of spatial ordinal patterns of fixed order d

Miscellaneous

$I : [0,1] \to \mathbb{R}$	information function $I(p)$ w.r.t. p
$H(\mathbf{p})$	(Shannon) entropy w.r.t. $\mathbf{p}$
$\mathrm{PE}(\mathbf{p})$	permutation entropy w.r.t. $\mathbf{p}$
$R_q(\mathbf{p})$	Rényi-q entropy w.r.t. $\mathbf{p}$

$U_n(h)$	U-statistic with kernel h
h_1	$h_1(x_1) = \mathbb{E}h(x_1, X_2)$
$(\beta_k)_{k\in\mathbb{N}_0}$	absolute regularity coefficients
$(a_k)_{k\in\mathbb{N}_0}$	approximating constants
$\phi :]0,\infty[\to]0,\infty[$	p-continuity function
S^d, S_n^d	symbolic correlation integral and its estimator
$\preccurlyeq_C$	concordance ordering
$\preccurlyeq_{sm}$	supermodular ordering

Abbrevations

a.s.	almost surely
cdf	cumulative distribution function
CLT	central limit theorem
i.i.d.	independent and identically distributed
LLN	law of large numbers
L_r-NED	L_r-near-epoch dependence
MOP	multivariate ordinal pattern
$\mathbb{P}$-NED	near-epoch dependence in probability
OPD	ordinal pattern dependence
pmf	probability mass function
SCI	symbolic correlation integral
SOP	spatial ordinal pattern

List of Figures

List of Tables