technische universität
dortmund

fakultät
statistik

Dissertation

in partial fulfillment of the requirements for the degree of
Doktor der Naturwissenschaften

K-sign depth: Asymptotic distribution, efficient computation and applications

by M.Sc. Dennis Malcherczyk

Referees:	Prof. Dr. Christine Müller
	Prof. Dr. Carsten Jentsch
	Prof. Dr. Anne Leucht
Comission chairperson:	Prof. Dr. Roland Fried
Submitted:	6. October 2021
Day of Oral Exam:	9. February 2022

Department of Statistics
Statistics with Applications in the Field of Engineering Sciences
TU Dortmund University

Acknowledgments

Meine Dissertation wurde von Prof. Dr. Christine Müller betreut und baut auf der Forschung ihres Lehrstuhls der letzten Jahre auf. Sie hat mich während meines Promotionsstudiums immer unterstützt, viele meiner Beweise gelesen und war stets offen für Diskussionen und Austausch neuer Ideen. Für ihre Unterstützungen möchte ich mich hiermit bedanken. Weiterhin hat sie mir Freiraum für die Entfaltung meiner Ideen gegeben. Am Ende meines Promotionsstudiums sind einige Ideen übrig geblieben, die ich der Nachwelt nicht verschweigen wollte, aber in aller Ausführlichkeit den Rahmen dieser Dissertation gesprengt hätten. Deswegen habe ich mich dafür entschieden, in den letzten beiden Kapiteln viele kleine Ausblicke anzubieten, statt eine einzige Ideen ausführlich zu diskutieren.
Außerdem möchte ich die Zusammenarbeit mit Dr. Kevin Leckey hervor heben. Die Diskussionen mit ihm inspirierten mich, neue Ideen zu verfolgen und halfen mir, bestehende Argumente auszuarbeiten, zu verfeinern und Denkfehler zu erkennen. Einige Resultate aus den Diskussionen mit ihm sind auch in diese Dissertation geflossen und in andere wissenschaftliche Publikationen gemündet. Darüber hinaus hat Kevin die Vorversion meiner Dissertation gelesen und mir nützliche Hinweise vor der abschließenden Einreichung auf den Weg gegeben.
An dieser Stelle möchte ich mich auch bei Lars Schroeder, Su Jin Park und Sven Popovici bedanken, die ebenfalls Teile meiner Dissertation zur Korrektur gelesen haben. Alle haben mich individuell mit hilfreichen Ratschlägen unterstützt.
Abschließend möchte ich meiner Familie Marzanna, Andreas und Aleksandra Malcherczyk und nochmals Sven Popovici für ihre stetige Unterstützung in meinem Leben meinen herzlichen Dank aussprechen.

Table of Contents

1 Introduction

The regression analysis is a fundamental tool in statistics to find relationships between several variables. Estimators and tests for parameters in the classical linear model with independent identically normally distributed errors have already been widely studied. The ordinary least-squares estimator and tests based on this estimator yield statistical procedures with optimality properties under the given assumptions, see Clark (2008); Lehmann and Romano (2005). However, procedures based on the ordinary least-squares estimator can have a high loss of efficiency if the assumptions are not satisfied. Heteroscedastic, skewed, heavy tailed or correlated errors are examples for violated assumptions. The research of Marcinko (2014), Scariano and Davenport (1987) and Rousseeuw and Leroy (1987) are examples for the study of the loss of efficiency under such violations.

The literature offers a number of other statistical methods which can be used under possibly violated assumptions, e.g., the MM-estimators and the associated asymptotic Wald-tests (Yohai, 1987; Maronna et al., 2006, p. 124). Although these robust alternatives have consistency properties (Huber and Ronchetti, 2009, p. 126), the estimation is based on an optimization procedure and thus not trivial to understand and implement. Furthermore, optimization in high dimensions is very challenging (Törn and Žilinskas, 1989, p. 11), especially when a high dimensional covariance matrix has to be estimated (Gomez and Gallón, 2011). All in all, the number of robust tests is still limited in the regression context. Therefore, research for methods with weak assumptions, which are also easy to comprehend, is worthwhile.

In this thesis, the K-sign depth will be introduced and discussed based on theoretical and practical aspects. This statistic evaluates the fit of a suggested parameter from a class of models for observed data. In contrast to the likelihood, this depth is not computed by density functions of a given class of distributions, but rather by the sign-structure from the residuals. Under the independence assumption of errors with a continuous distribution, the statistical behavior of the sign-structure can be derived. If the number of sign changes is too low, the suggested parameter can be declared as a bad fit. The choice of the hyper-parameter K denotes how many consecutive sign changes are counted. For $K = 2$, we obtain the statistic of the classical sign test. Hence, tests based on the K-sign depth can be understood as a class of extended versions of the classical sign test for $K > 2$.

Apart from the indepedence and continuous distribution, no more essential assumptions are needed for applying this depth. The independence can further be weakened if the errors have sufficiently quickly decaying autocorrelations. The restrictions for

applying this statistic are very low, so the high flexibility of this depth allows applications in many contexts potentially. Based on the idea to consider sign-structures, this depth is not difficult to comprehend as well. Therefore, research for this depth seems valuable. Historically, the K-sign depth is developed from the regressions depth introduced by Rousseeuw and Hubert (1999) which can be another motivation for its research.

This thesis is structured as follows: In Chapter 2, the assumptions on the model and the K-sign depth will be introduced. Further, a historical background of the concept of several types of regression depths will be given briefly. In Chapter 3, the asymptotic distribution of the K-sign depth will be derived for arbitrary K by continuing research work from the past: For $K = 2$ and for the more difficult case $K = 3$, Müller (2005) and Kustosz et al. (2016a) derived the asymptotic distribution, respectively. In the Master thesis Malcherczyk (2018a), the proof for $K = 3$ has been simplified which helped to find a proof for $K = 4$ there as well. In Malcherczyk et al. (2021), a proof of the derivation for general K has finally been shown. This proof is a shortened version of the derivation in Chapter 3.

The computation of the depth by its definition leads to an algorithm having an undesirable polynomial time complexity of degree K. Therefore, improved algorithms in linear time will be discussed in Chapter 4 and 5. In Chapter 4, we use a representation of the K-sign depth from Chapter 3 describing a decomposition of an asymptotically relevant part and a negligible part. The asymptotic part can be computed in linear time as an approximation of the K-depth. In Chapter 5, another idea for an efficient computation based on summarizing blocks with the same signs from the residuals will be presented. This algorithm has already been introduced in Leckey et al. (2020) to compute the K-sign depth without using the results from the asymptotic derivation. Furthermore, the algorithm from Chapter 5 leads to some new theoretical results as well.

Chapter 6 will present varieties of test procedures based on the K-sign depth for applications. Testing the fit of models or independence are possible for example. The performance of both applications will be discussed. Some conjectures for the choice of the hyper-parameter K and procedures how to use the K-sign depth in practice will be presented as well. Generalizations of the K-sign depth are introduced in Chapter 7 as an outlook for further research.

If the reader is only interested on a theoretical or practical focus, we recommend to consider Figure 1. The upper path is more focused on theoretical aspects such as the asymptotic derivation in Chapter 3, the discussion of an interesting conjecture from Chapter 5.4 or the generalization approaches in Chapter 7. In order to understand

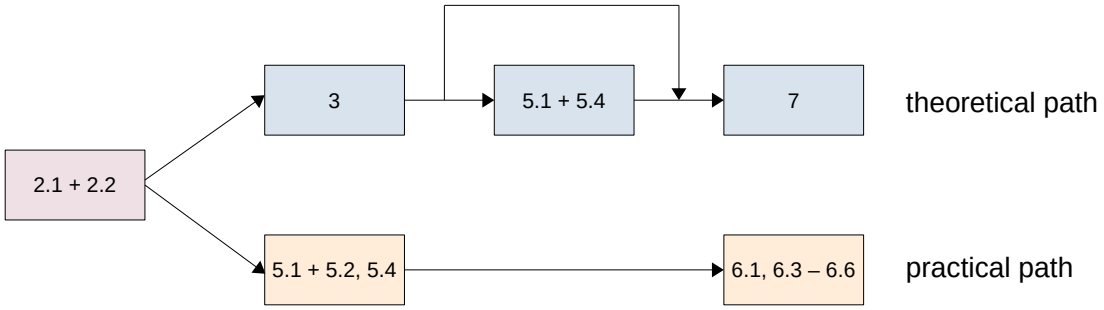

Figure 1: Two different suggested paths for reading this thesis based on a theoretical or practical focus

Chapter 7, it is highly recommended to read Chapter 3 first. Chapter 5 can be skipped if desired.

The lower path is more focused on the implementation and the statistical methods based on the K-sign depth. In Chapter 5, the more efficient implementation is presented. Chapter 5.4 is theoretical but useful for the interpretations in Chapter 6.

For computations and implementations, `R` is mostly used in this thesis (R Core Team, 2021). For some computer algebraic applications, `Mathematica` is used as well (Wolfram Research, 2019). All presentend algorithms in this thesis are implemented by Dennis Malcherczyk and included in the `R`-package `GSignTest` Horn (2021b) which is available on GitHub. These algorithms are implemented in `C++` using the `R`-package `Rcpp` for running `C++` code in `R` (Eddelbuettel and Francois, 2011).

2 The K-sign depth

Chapter 2 is structured as follows. In Chapter 2.1, the class of models we consider mostly in this thesis will be introduced. Connections between sign changes and the model fit are explained as well. In Chapter 2.2, the K-sign depth will be defined. In Chapter 2.3, some extra information concerning the historical background of the K-sign depth will be presented.

2.1 Model and motivation

In the whole thesis, we will consider the following type of models:

$$Y_n = g(\mathbf{X}_n, \boldsymbol{\theta}^*) + E_n \text{ for } n = 1, \dots, N. \tag{1}$$

The unknown and true parameter $\boldsymbol{\theta}^*$ is contained in a p-dimensional parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$. The explanatory variables $\mathbf{X}_1, \dots, \mathbf{X}_N$ have values in $\mathbb{R}^d$. The class of model functions is denoted by $g : \mathbb{R}^d \times \boldsymbol{\Theta} \to \mathbb{R}$ for given explanatory variables and a parameter in $\boldsymbol{\Theta}$. The additive, unobservable errors $E_1, \dots, E_N$ are random variables on a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$. E.g., polynomial regression models of degree $q \in \mathbb{N}$ can be expressed by the model function $g(x, \boldsymbol{\theta}) = \theta_0 + \sum_{i=1}^{q} \theta_i x^i$ with $q+1$ parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_q)^\top \in \mathbb{R}^{q+1}$ and $x \in \mathbb{R}$.
We allow the explanatory variables $\mathbf{X}_1, \dots, \mathbf{X}_N$ to be random variables with values in $\mathbb{R}^d$ (for the sake of simplicity, also on the same probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ as the errors) in order to have also stochastic explanatory variables instead of only deterministic planned explanatory variables. Stochastic explanatory variables are common in various applications. In an example of our research group from technometrics, it is of interest to model the crack growth of a reinforced-concrete bridge by using the air temperature, the temperature of the bridge and the traffic on the bridge as explanatory variables (Abbas et al., 2019). All of these explanatory variables (temperatures, traffic) cannot be set arbitrarily as in laboratory experiments. Additionally, we can model time series by allowing stochastic explanatory variables.
E.g., AR(p)-processes (Peña et al., 2001, p. 53) can be modeled by Formula (1) such that $Y_n = g((Y_{n-1}, \dots, Y_{n-p})^\top, \boldsymbol{\theta}) + E_n$ for $n = p, \dots, N$ and given initial values $Y_0, \dots, Y_{p-1}$ where the last p observations $Y_{n-1}, \dots, Y_{n-p}$ represent the stochastic explanatory variables for modeling Y_n.
Other residual-based models can be considered as well, e.g., nonparametric models (Hastie et al., 2009, p. 191). Since we will be interested in testing model classes, a semi-parametric settings as in Formula (1) is considered in this thesis.

For the errors $E_1, \ldots, E_N$, the following assumptions are given for Chapter 2 to 6:

Assumption 2.1. *Let $E_1, \ldots, E_N$ be random variables in $\mathbb{R}$ which satisfy the following conditions:*

$$E_1, \ldots, E_N \text{ are independent,} \tag{A1}$$

$$\mathbb{P}(E_n > 0) = \mathbb{P}(E_n < 0) = \frac{1}{2}, n = 1, \ldots, N. \tag{A2}$$

The independence of the errors can be assumed in many applications, so assumption (A1) is not very restrictive. Moreover, we will see later in Chapter 7.2 that assumption (A1) can be weakened if necessary. If the errors have a continuous distribution with a connected support, the second assumption (A2) is equivalent to the condition that the median of the errors is zero. If the model has an intercept, this assumption is no real restriction on the errors. Note that we neither assume identically distributed random variables nor make assumptions on the moments of the errors. In particular, skewed, heavy-tailed or heteroscedastic errors are allowed. The residuals of a model based on $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ are defined by

$$R_n(\boldsymbol{\theta}) := Y_n - g(\mathbf{X}_n, \boldsymbol{\theta}) \text{ for } n = 1, \ldots, N. \tag{2}$$

If we insert the true parameter $\boldsymbol{\theta}^*$, the residuals coincide with the errors:

$$R_n(\boldsymbol{\theta}^*) = E_n \text{ for } n = 1, \ldots, N. \tag{3}$$

We often omit the parameter and denote $R_1, \ldots, R_N$ for the residuals instead. Random variables will be denoted by capital letters and their realizations by small letters, e.g., $r_1, \ldots, r_N$ for realizations of the residuals.
Our aim is to construct statistical tests for hypotheses of the following form

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0, \; H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1 \tag{4}$$

where $\boldsymbol{\Theta_0}$ and $\boldsymbol{\Theta}_1$ are a disjoint decomposition of the parameter space denoted by $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \uplus \boldsymbol{\Theta}_1$. These statistical tests should only be based on Assumption 2.1 and the functional structure of the model class. According to Formula (3) and Assumption 2.1, statistical properties of the signs in the residual vector are known under the true parameter $\boldsymbol{\theta}^*$ and can lead to statistical tests.
Figure 2 visualizes an example for a typical statistical behavior of the residuals under the true model (left side) and a situation with a poorly fitting model (right side). The blue points are realizations of the real model, the dashed line represents a fitted

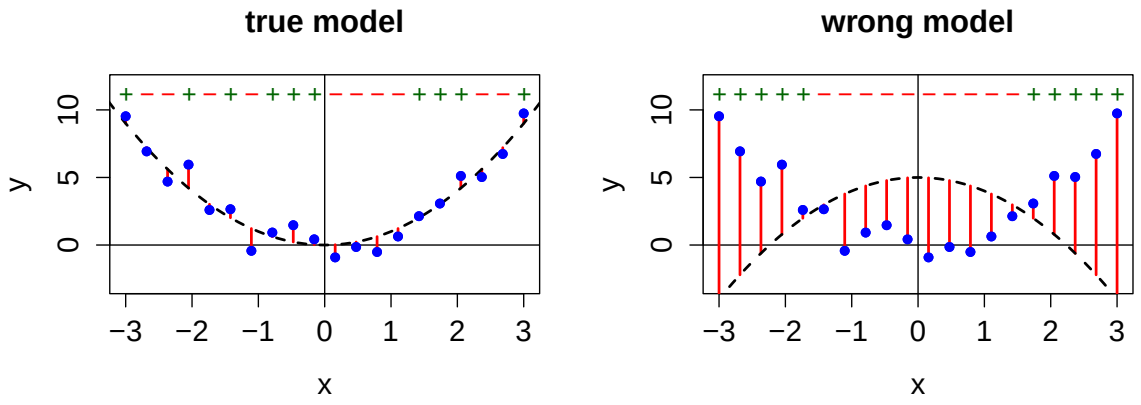

Figure 2: Sign changes as an indication for a good fit under Assumption 2.1.

model and the red lines are the residuals. On the top of each figure, the signs of the residuals are given. Since the errors $E_1, \ldots, E_N$ satisfy the condition in Formula (A2), we expect $\frac{N-1}{2}$ sign changes (cf. Lemma 5.13, p. 89, for details) under the true parameter. Moreover, due to the independence of the errors, we expect no systematic pattern (left side). For the wrong model, the probability of positive or negative signs are not necessarily equal and depends on the explanatory variables. Due to unexplained or badly explained parts of the real model, we obtain block structures of the same signs which are atypical under Assumption 2.1 for the residual vector (right side). Thus, signs with such block patterns are an indication for a bad fit of the given parameter. This leads to the idea to evaluate a model by the resulting number of *sign changes*.

2.2 Definition of the K-sign depth

Based on the notation of the model and the residuals given in Formula (1) and (2), we will define the *K-sign depth*, or shortly denoted as *K-depth*. Before we give the definition, we want to introduce the following notation. For arbitrary coefficients $a(n_1, \ldots, n_K) \in \mathbb{R}$, we define the *ordered sum*:

$$\sum_{1 \le n_1 < \ldots < n_K \le N} a(n_1, \ldots, n_K) := \sum_{n_1=1}^{N-K+1} \sum_{n_1+1 \le n_2 < \ldots < n_K \le N} a(n_1, \ldots, n_K).$$

By recursion, we obtain the following representation of K nested sums:

$$\sum_{1 \le n_1 < \ldots < n_K \le N} a(n_1, \ldots, n_K) = \sum_{n_1=1}^{N-K+1} \sum_{n_2=n_1+1}^{N-K+2} \cdots \sum_{n_K=n_{K-1}+1}^{N} a(n_1, \ldots, n_K).$$

Definition 2.2. *Let* $\mathbf{R}(\boldsymbol{\theta}) = (R_1(\boldsymbol{\theta}), \dots, R_N(\boldsymbol{\theta}))^\top$ *be a residual vector of the model under* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*. (We shortly denote* $\mathbf{R} = (R_1, \dots, R_N)^\top$ *if the parameter* $\boldsymbol{\theta}$ *is known from the context.) For* $K \in \mathbb{N} \setminus \{1\}$*, we define the* ***K-sign depth*** *as*

$$d_K(\mathbf{R}) := \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \dots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(-1)^k < 0\} \right).$$

The K-depth is the relative frequency of ordered K-tuples with $K-1$ sign changes. If a K-tuple has $K-1$ sign changes, we say that the signs are *alternating*. Under Assumption 2.1, we can understand the statistical behavior of the number of alternating signs and by that the statistical behavior of the K-depth.
At first, we can consider boundaries of the K-depth in order to obtain first impressions for interpretations whether a parameter $\boldsymbol{\theta}$ fits well or poorly. The lower boundary of the K-depth is zero and occurs if there exists no K-tuple with $K-1$ sign changes. This is an indication for a bad fit. Conversely, the higher the value of the K-depth, the more the given parameters fit well. Under residuals $R_1, \dots, R_N$ with $N-1$ sign changes, the K-depth is supposed to be maximal. The suspected asymptotic upper boundary $\frac{1}{2^{K-1}}$ and the suspected exact upper boundary are computed in Leckey et al. (2020). However, a suspiciously high number of sign changes is also an indication for negative correlation. Note further that the depth of a parameter is a very similar concept to the likelihood.
The next remark explains the influence of the order of the residuals.

Remark 2.3. *A crucial aspect for the application of the* K*-depth is the choice of an ordering criterion for the residual vector* $\mathbf{R} = (R_1, \dots, R_N)^\top$*. The value of the* K*-depth can change drastically if we change the ordering afterwards. For* $K = 3$*, we can construct an example of two residual vectors*

$$\mathbf{r}^{(1)} := (1, -1, 1)^\top, \mathbf{r}^{(2)} := (1, 1, -1)^\top,$$

which contain the same values but in a different order. Due to different orders, we obtain different values of their 3-depth: $d_3(\mathbf{r}^{(1)}) = 1$ *and* $d_3(\mathbf{r}^{(2)}) = 0$*.*
For regression with univariate explanatory variables $X_1, \dots, X_N$*, we can use their natural order on* $\mathbb{R}$ *or the passage of time in time series. However, for multiple regression, an obvious choice how to order multivariate data is usually not given.*
Under the true parameter $\boldsymbol{\theta}^*$ *(or for residuals satisfying Assumption 2.1), the distribution of the* K*-depth does not depend on the order if the rule of order fulfills the*

following two sufficient conditions:

1. *The rule of order is chosen in advance.*
2. *If $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independent from $E_1, \ldots, E_N$, then the realizations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ can be used for ordering the data.*

However, the power of the tests based on the K-depth (introduced in Chapter 6) varies when we use different orders. In Horn (2021b), several orders of multivariate explanatory variables, e.g., based on the shortest Hamiltonian path or the k-nearest neighbor algorithm, are considered and the resulting power of tests based on the K-depth with those ordering approaches are investigated. The results in this thesis are usually independent from the order criterion if not otherwise specified.

2.3 Excursion to the history of the K-sign depth

Originally, the K-sign depth has a long historical background in the research field of depth functions which we want to introduce here shortly.
Tukey (1975) introduced the *halfspace depth* as a statistical tool to measure the depth of a location parameter $\boldsymbol{\mu} \in \mathbb{R}^q$ with respect to a fixed data set $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in $\mathbb{R}^q$. The halfspace depth of $\boldsymbol{\mu}$ with respect to the data is defined by the minimal number of data points $\mathbf{x}_1, \ldots, \mathbf{x}_N$ which are elements of a halfspace containing $\boldsymbol{\mu}$:

$$d^H(\boldsymbol{\mu}, (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top) := \frac{1}{N} \min_{\mathbf{u} \in \mathbb{R}^q} |\{n; \mathbf{u}^\top \mathbf{x}_n \geq \mathbf{u}^\top \boldsymbol{\mu}\}|$$

where $|A|$ denotes the cardinality of a given set A. This depth concept was originally introduced for the case $q = 2$ for visualizing bivariate data. Moreover, the set of elements in $\mathbb{R}^q$ with the maximal halfspace depth can be understood as a multivariate version of the median of $\mathbf{x}_1, \ldots, \mathbf{x}_N$.
Liu (1988) introduced another depth concept called *simplicial depth* for the location of multivariate data. This depth is defined by the relative number of $(q+1)$-dimensional simplices with edges in $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ which cover $\boldsymbol{\mu}$:

$$d^S(\boldsymbol{\mu}, (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top) := \frac{1}{\binom{N}{q+1}} \sum_{1 \leq n_1 < \ldots < n_{q+1} \leq N} \mathbb{1}\{\boldsymbol{\mu} \in \triangle(\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_{q+1}})\}$$

where $\triangle(\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_{q+1}}) \subseteq \mathbb{R}^q$ denotes the set of the simplicial having the edges $\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_{q+1}}$. The simplicial depth can be represented by counting all ordered

$(q+1)$-tuples with positive halfspace depths:

$$d^S(\boldsymbol{\mu}, (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top) = \frac{1}{\binom{N}{q+1}} \sum_{1 \leq n_1 < \ldots < n_{q+1} \leq N} \mathbb{1}\{d^H(\boldsymbol{\mu}, (\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_{q+1}})^\top) > 0\}.$$

This connection between the halfspace depth and the simplicial depth is often used for defining new depth concepts from previous depth notions by the sum of all ordered $(q+1)$-tuples (Müller, 2005; Wellmann and Müller, 2010).

Rousseeuw and Hubert (1999) gave further ideas of depth approaches in the regression case for the depth of parameters $\boldsymbol{\theta}$ in a given parameter class $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$ for observed data points $\mathbf{z}_1 = (\mathbf{x}_1, y_1)^\top, \ldots, \mathbf{z}_N = (\mathbf{x}_N, y_N)^\top \in \mathbb{R}^{d+1}$. For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we can define the associated residuals $r_1(\boldsymbol{\theta}), \ldots, r_N(\boldsymbol{\theta})$.

The *regression depth* $d^R(\boldsymbol{\theta}, (\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top)$ of a parameter $\boldsymbol{\theta}$ and data $(\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top$ is defined by the minimal number of data points that must be omitted to obtain a regression *non-fit*. A regression parameter $\boldsymbol{\theta}$ is called a regression non-fit for the data if and only if there exists an affine hyperplane $\mathbf{V} \subseteq \mathbb{R}^d$ such that $\mathbf{x}_1, \ldots, \mathbf{x}_N \notin \mathbf{V}$ and such that $r_n(\boldsymbol{\theta}) > 0$ for all n with $\mathbf{x}_n$ in one of its open halfspace and $r_n(\boldsymbol{\theta}) < 0$ for all n with $\mathbf{x}_n$ in the other halfspace. However, the computational costs for the regression depth are expensive for increasing p and only approximations are available for $p > 4$ (Rousseeuw and Struyf, 1998).

In the case of the linear regression, Rousseeuw and Hubert mentioned that the regression depth of three data points is positive if and only if the signs of the residuals are alternating. This leads to the idea to consider only tuples of the data. The *simplicial regression depth* is defined by counting the relative number of $(p+1)$-tuples which have a positive regression depth (Müller, 2005):

$$d^{SR}(\boldsymbol{\theta}, (\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top) := \frac{1}{\binom{N}{p+1}} \sum_{1 \leq n_1 < \ldots < n_{p+1} \leq N} \mathbb{1}\{d^R(\boldsymbol{\theta}, (\mathbf{z}_{n_1}, \ldots, \mathbf{z}_{n_{p+1}})^\top) > 0\}.$$

This depth concept is called 'simplicial', because it mimics the theoretical connection between Tukey's halfspace depth and Liu's simplicial depth for the regression depth. Mizera (2002) considered the regression depth again and found an equivalent representation of a regression non-fit under convexity conditions on the model class by directional derivatives and introduced the *tangent depth*:

$$d^T(\boldsymbol{\theta}, (\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top) := \frac{1}{N} \min_{\mathbf{u} \in \mathbb{R}^p} \left| \left\{ n; \mathbf{u}^\top \frac{\partial}{\partial \boldsymbol{\theta}} r_n(\boldsymbol{\theta})^2 \geq 0 \right\} \right|.$$

Originally, also other quality functions than the squares of the residuals are introduced by Mizera (2002).

Then, the *simplicial tangent depth* is defined by

$$d^{ST}(\boldsymbol{\theta}, (\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top) := \frac{1}{\binom{N}{p+1}} \sum_{1 \leq n_1 < \ldots < n_{p+1} \leq N} \mathbb{1}\{d^T(\boldsymbol{\theta}, (\mathbf{z}_{n_1}, \ldots, \mathbf{z}_{n_{p+1}})^\top) > 0\}.$$

Kustosz et al. (2016b) found conditions on the model class which yields for the simplicial tangent depth:

$$\begin{aligned} d^{ST}(\boldsymbol{\theta}, (\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top) = & \frac{1}{\binom{N}{p+1}} \sum_{1 \leq n_1 < \ldots < n_{p+1} \leq N} \left(\prod_{k=1}^{p+1} \mathbb{1}\{r_{n_k}(\boldsymbol{\theta})(-1)^k > 0\} \right. \\ & \left. + \prod_{k=1}^{p+1} \mathbb{1}\{r_{n_k}(\boldsymbol{\theta})(-1)^k < 0\} + \left(1 - \prod_{k=1}^{p+1} \mathbb{1}\{r_{n_k}(\boldsymbol{\theta}) \neq 0\}\right) \right). \end{aligned} \tag{5}$$

Thus, the tangent depth of a $(p+1)$-tuple of residuals is positive if and only if their signs are alternating or if one residual is exactly zero. In the case that the residuals cannot be zero, we obtain the equality of the simplicial tangent depth and the K-sign depth for $K = p + 1$ from Definition 2.2. The scenario that residuals are equal to zero can be usually omitted if the errors of the model are realizations of random variables with a continuous distribution.
The K-sign depth has the advantage that it is easier to comprehend than the original depth functions. Therefore, efficient computational tools for computation in linear time can also be constructed as it will be shown in this thesis.

Dealing with zeros in the residual vector

The difference between Formula (5) and Definition 2.2 is the third part in the sum of Formula (5). This part increases the value of the simplicial tangent depth if at least one residual of a $(p+1)$-tuple is zero. This is a plausible procedure to deal with zeros since residuals equal to zero denotes a good fit. The K-sign depth in Definition 2.2 ignores this case by assuming that zeros never or only sometimes occur and such cases are neglected therefore. We will assume this in the following parts of this thesis since an analysis under the third term in Formula (5) would lead to more complicated derivations which we would like to avoid. For the algorithms in Chapter 4 and 5, we remove all zeros before computing the K-sign depth.
If the number of zeros is not negligible, the original K-sign depth can lead to unfitting results. For the extreme case that each residual is equal to zero, we expect a strong fit. However, the K-sign depth in Definition 2.2 is zero in this case which expresses mistakenly a bad fit.

There are several ways how to deal with zeros when we apply the K-sign depth:

(i) We can *remove zeros* as in all implementations in (Horn, 2021b). This procedure should be used under the assumption that zeros rarely occur.

(ii) The original K-sign depth in Definition 2.2 can be replaced by

$$\tilde{d}_K(\mathbf{R}) := \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{1}\{\varphi(R_{n_k})(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{\varphi(R_{n_k})(-1)^k < 0\} \right)$$

with $\varphi(x) = \mathbb{1}\{x \geq 0\} - \mathbb{1}\{x < 0\}$, i.e., *zeros are interpreted as positive signs*. However, the decision to set zeros as positive is arbitrary. Moreover, this procedure handles residuals with multiple consecutive zeros badly either.

(iii) We can *replace zeros by the sign* of sampled random variables satisfying (A1) and (A2). This can be understood as an extension of (ii) by decomposing the value zero into two possible random values. However, randomized decision rules only make sense from a theoretical perspective but can be absurd in a practical context. In the worst case, the decision of the sampled random variables can be relevant whether a hypothesis is rejected or not.

(iv) In order to eliminate the stochastic components in (iii), we can replace the zeros by deterministic rules that increase the K-sign depth under the occurrence of zeros. E.g., we can *replace a zero by the reversed sign from the entry before*. This leads to a sign change which increases the K-sign depth. If multiple consecutive zeros occur, they should have alternating signs. If the first entry of the residual vector is already zero than the next non-zero entry can be used instead and the signs can be constructed alternating in the reversed direction. If only zeros occur than we translate this residual vector to a vector with alternating signs. Note that a high number of sign changes can be misunderstood as negative correlation although higher number of sign changes are only artificially added.

(v) The third part in Formula (5) takes a similar idea as in (iv) in consideration. Its value is determined by *the number of zeros* in the data as the third term in Formula (5) and can be added to Definition 2.2 as an extra part. But as mentioned before, this part is more complicated to handle and omitted in this thesis therefore.

3 Asymptotic distribution of the K-sign depth

The aim of this chapter is to analyze the asymptotic distribution of the K-depth under the true parameter $\boldsymbol{\theta}^*$ with errors $\mathbf{E} = (E_1, \ldots, E_N)^\top$ fulfilling Assumption 2.1. For the true parameter, the errors $\mathbf{E} = (E_1, \ldots, E_N)^\top$ corresponds to the residual vector according to Formula (3), p. 5, and $d_K(\mathbf{E})$ denotes their K-depth. The expected value of the K-depth is then given by

$$\mathbb{E}(d_K(\mathbf{E})) = \frac{1}{2^{K-1}}$$

as the following calculation shows:

$$\begin{aligned}
\mathbb{E}(d_K(\mathbf{E})) &= \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \mathbb{E}\left(\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} \right) \\
&= \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\mathbb{P}\left(\bigcap_{k=1}^{K} \left\{ E_{n_k}(-1)^k > 0 \right\} \right) + \mathbb{P}\left(\bigcap_{k=1}^{K} \left\{ E_{n_k}(-1)^k < 0 \right\} \right) \right) \\
&\overset{\text{(A1)}}{=} \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{P}\left(E_{n_k}(-1)^k > 0 \right) + \prod_{k=1}^{K} \mathbb{P}\left(E_{n_k}(-1)^k < 0 \right) \right) \\
&\overset{\text{(A2)}}{=} \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\frac{1}{2^K} + \frac{1}{2^K} \right) = \frac{1}{2^{K-1}}.
\end{aligned}$$

Note that the independence of the errors $E_1, \ldots, E_N$ and the probability of observing positive or negative values are needed to compute this expected value. In order to obtain a non-degenerated asymptotic distribution of the K-depth, we consider its rescaled version:

$$T_K(\mathbf{E}) = N \left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}} \right). \tag{6}$$

Centering by the expected value $\frac{1}{2^{K-1}}$ and the scaling factor N are necessary to apply the (Functional) Central Limit Theorem later in the derivation of the asymptotic distribution.

Chapter 3 is split into three parts. In Chapter 3.1, the derivation of an equivalent representation of the 3-depth will be derived. Kustosz et al. (2016a) have already derived an asymptotic distribution, but the derivation has been strongly simplified in the Master thesis Malcherczyk (2018a). This simplification of the derivation is useful for an easier understanding of the more general case. It is recommened to study Chapter 3.1 first before reading the general case. Nevertheless, Chapter 3.1 can be

skipped, because the more general case also includes the case $K = 3$. The 2-depth is equivalent to counting the number of positive or negative signs (Leckey et al., 2020) so it is straight forward to derive its asymptotic distribution. In Chapter 3.2, the general case for the previous chapter is presented and references to the special case $K = 3$ are given to understand the extension better. In Chapter 3.3, the asymptotic distribution of the K-depth is derived for general K by using Donsker's invariance principle. In Malcherczyk et al. (2021), a similar derivation is given, but the resulting asymptotic representation is given in another form.

Landau-Bachmann notation for asymptotic behavior

In the following, we will use the Landau-Bachmann big-$\mathcal{O}$ and small-o notation for the asymptotic analysis (Sedgewick and Flajolet, 2012).

Definition 3.1. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions.*

(a) g is an asymptotic upper bound for f denoted by $f = \mathcal{O}(g)$ if and only if

$$\exists C > 0 \, \exists x_0 > 0 \, \forall x > x_0 : |f(x)| \leq C \cdot |g(x)|.$$

(b) f is asymptotically less than g denoted by $f = o(g)$ if and only if

$$\forall C > 0 \, \exists x_0 > 0 \, \forall x > x_0 : |f(x)| < C \cdot |g(x)|.$$

(c) f and g are asymptotically equivalent denoted by $f = \Theta(g)$ if and only if

$$f = \mathcal{O}(g) \text{ and } g = \mathcal{O}(f).$$

The original Landau-Bachmann notation does not consider random variables and their asymptotic behavior. Therefore, we introduce the following notation.

Definition 3.2. *Let $(X_N)_{N\in\mathbb{N}}$ and $(Y_N)_{N\in\mathbb{N}}$ be two sequences of random variables in $\mathbb{R}$. Then $(X_N)_{N\in\mathbb{N}}$ is almost surely asymptotically less than $(Y_N)_{N\in\mathbb{N}}$ denoted by $X_N = o_{a.s.}(Y_N)$ if and only if*

$$\forall C > 0 \, \exists N_0 \in \mathbb{N} \, \forall N > N_0 : |X_N| < C \cdot |Y_N| \text{ almost surely.}$$

We have usually $Y_N = 1$ for $N \in \mathbb{N}$ in this thesis. This additional notation is important to differentiate the o-terms with and without stochastic components. Note that such notations are often only used for stochastic convergence and written by o_p in the literature, e.g., see van der Vaart (2000).

3.1 Integral representation for $K = 3$

In order to derive the asymptotic distributions of the K-depth, we will rewrite the statistic such that limit theorems can be used. In the first step, we use a representation of the 3-depth containing sign functions.

Lemma 3.3. *For $(x_1, x_2, x_3)^\top \in \mathbb{R}^3$, $x_i \neq 0$, $i \in \{1, 2, 3\}$, we have*

$$\begin{aligned}&\mathbb{1}\{x_1 > 0, x_2 < 0, x_3 > 0\} + \mathbb{1}\{x_1 < 0, x_2 > 0, x_3 < 0\} - \frac{1}{4}\\ =&\frac{1}{4}\left(-\psi(x_1)\psi(x_2) + \psi(x_1)\psi(x_3) - \psi(x_2)\psi(x_3)\right),\end{aligned}$$

where $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ defines the sign function.

This lemma can be proved by checking all $2^3 = 8$ cases for $x_i > 0$ or $x_i < 0$, $i \in \{1, 2, 3\}$. Note that Lemma 3.6, p. 20, will be a generalized version of Lemma 3.3 and a general proof will be provided there as well. The errors $E_1, \ldots, E_N$ are unequal to zero almost surely since they satisfy (A2) and Lemma 3.3 can be applied:

$$\begin{aligned}N\left(d_3(\mathbf{E}) - \frac{1}{4}\right) =& N\left(\frac{1}{\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \left(\mathbb{1}\{E_{n_1} > 0, E_{n_2} < 0, E_{n_3} > 0\}\right.\right.\\ &\left.\left. + \mathbb{1}\{E_{n_1} < 0, E_{n_2} > 0, E_{n_3} < 0\} - \frac{1}{4}\right)\right)\\ =&\frac{N}{4\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \left(-\psi(E_{n_1})\psi(E_{n_2}) + \psi(E_{n_1})\psi(E_{n_3}) - \psi(E_{n_2})\psi(E_{n_3})\right).\end{aligned}$$

In order to simplify the last expression, we split the sum up into three parts. Then, we omit the third summation index that is not included in each summand and obtain

$$\begin{aligned}&N\left(d_3(\mathbf{E}) - \frac{1}{4}\right)\\ =&\frac{N}{4\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \left(-\psi(E_{n_1})\psi(E_{n_2}) + \psi(E_{n_1})\psi(E_{n_3}) - \psi(E_{n_2})\psi(E_{n_3})\right)\\ =&\frac{N}{4\binom{N}{3}} \left(-\sum_{1 \leq n_1 < n_2 < n_3 \leq N} \psi(E_{n_1})\psi(E_{n_2}) + \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \psi(E_{n_1})\psi(E_{n_3})\right.\\ &\left. - \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \psi(E_{n_2})\psi(E_{n_3})\right)\\ =&\frac{N}{4\binom{N}{3}} \left(-\sum_{1 \leq n_1 < n_2 \leq N} (N - n_2)\psi(E_{n_1})\psi(E_{n_2})\right.\\ &\left. + \sum_{1 \leq n_1 < n_3 \leq N} (n_3 - n_1 - 1)\psi(E_{n_1})\psi(E_{n_3}) - \sum_{1 \leq n_2 < n_3 \leq N} (n_2 - 1)\psi(E_{n_2})\psi(E_{n_3})\right)\end{aligned} \tag{7}$$

$$
\begin{aligned}
=&\frac{N}{4\binom{N}{3}}\Bigg(- \sum_{1\le n_1<n_2\le N} (N-n_2)\psi(E_{n_1})\psi(E_{n_2}) \\
&+ \sum_{1\le n_1<n_2\le N} (n_2-n_1-1)\psi(E_{n_1})\psi(E_{n_2}) - \sum_{1\le n_1<n_2\le N} (n_1-1)\psi(E_{n_1})\psi(E_{n_2})\Bigg) \qquad (8)\\
=&\frac{N}{8\binom{N}{3}}\Bigg(- \sum_{1\le n_1\ne n_2\le N} (N-\max\{n_1,n_2\})\psi(E_{n_1})\psi(E_{n_2}) \\
&+ \sum_{1\le n_1\ne n_2\le N} (\max\{n_1,n_2\}-\min\{n_1,n_2\}-1)\psi(E_{n_1})\psi(E_{n_2}) \qquad (9)\\
&- \sum_{1\le n_1\ne n_2\le N} (\min\{n_1,n_2\}-1)\psi(E_{n_1})\psi(E_{n_2})\Bigg) \\
=&\frac{N}{8\binom{N}{3}} \sum_{1\le n_1\ne n_2\le N} \left(2\left(\max\{n_1,n_2\}-\min\{n_1,n_2\}\right)-N\right)\psi(E_{n_1})\psi(E_{n_2}) \qquad (10)\\
=&\frac{N^2}{8\binom{N}{3}} \sum_{1\le n_1\ne n_2\le N} \left(\frac{2|n_1-n_2|}{N}-1\right)\psi(E_{n_1})\psi(E_{n_2}). \qquad (11)
\end{aligned}
$$

Formula (7) is obtained by omitting the third summation index that is not included in the summands. We obtain Formula (8) by renaming all summation indices to n_1, n_2. The symmetry between n_1 and n_2 yields the sums in Formula (9). Note that n_1 and n_2 are replaced by $\min\{n_1,n_2\}$ and $\max\{n_1,n_2\}$, respectively, and the factor $\frac{1}{2}$ is necessary for compensation. By $|n_1-n_2| = \max\{n_1,n_2\} - \min\{n_1,n_2\}$ and factoring N out, we obtain Formula (11).

Remark 3.4. *Note first that* $\mathbb{E}(\psi(E_n)) = 0$ *and* $\operatorname{var}(\psi(E_n)) = 1$ *for* $n = 1,\dots,N$, *so that the Central Limit Theorem yields*

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\psi(E_n) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

A more simple version of Formula (11) without the factor $\frac{2|n_1-n_2|}{N}-1$ *would yield the asymptotic distribution directly by the Continuous Mapping Theorem:*

$$
\begin{aligned}
&\frac{1}{N}\sum_{1\le n_1\ne n_2\le N}\psi(E_{n_1})\psi(E_{n_2}) + 1 \\
=&\frac{1}{N}\sum_{n_1,n_2=1}^{N}\psi(E_{n_1})\psi(E_{n_2}) - \frac{1}{N}\sum_{n=1}^{N}\psi(E_n)^2 + 1 \\
=&\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\psi(E_n)\right)^2 \xrightarrow[N\to\infty]{\mathcal{D}} \chi_1^2
\end{aligned}
$$

since $\psi(E_n)^2 = 1$ *for* $n = 1, \ldots, N$ *almost surely. In order to handle the additional factor in Formula (11), we need to transform the derived representation even further.*

Note that the situation in Remark 3.4 occurs for the derivation of the asymptotic distribution of the 2-depth, see Leckey et al. (2020) or Malcherczyk (2018a). For $K = 3$, the weight factor $|n_1 - n_2|$ will be replaced by applying the following lemma.

Lemma 3.5. *Let* $n_1, n_2 \in \mathbb{R}$ *with* $|n_1 - n_2| \leq N$ *and* $N > 0$. *Then*

$$\frac{|n_1 - n_2|}{N} = 1 - \int_{-\infty}^{\infty} \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]^2} \left(\frac{n_1}{N} - t, \frac{n_2}{N} - t\right) \, dt. \tag{12}$$

Additionally, for $n_1, n_2 \in (0, N]$

$$\begin{aligned} &\int_{-\infty}^{\infty} \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]} \left(\frac{n_1}{N} - t\right) \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]} \left(\frac{n_2}{N} - t\right) \, dt \\ =&\int_{-\frac{1}{2}}^{\frac{3}{2}} \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]} \left(\frac{n_1}{N} - t\right) \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]} \left(\frac{n_2}{N} - t\right) \, dt. \end{aligned}$$

Lemma 3.5 describes a convolution-based representation of the absolute value function and has already been considered in Kustosz et al. (2016a) to derive the asymptotic distribution of the 3-depth. Since its proof was omitted in Kustosz et al. (2016a) and the result is crucial for the generalization of arbitrary K, the proof is given here.

Proof of Lemma 3.5.: We begin with an auxiliary calculation. For that, we modify the conditions of the indicator functions of the right side of Formula (12).

$$\begin{aligned} & -\frac{1}{2} < \frac{n_1}{N} - t \leq \frac{1}{2} \quad \text{and} \quad -\frac{1}{2} - \frac{n_2}{N} < t \leq \frac{1}{2} \\ \Leftrightarrow \quad & \frac{n_1}{N} - \frac{1}{2} \leq t < \frac{n_1}{N} + \frac{1}{2} \quad \text{and} \quad \frac{n_2}{N} - \frac{1}{2} \leq t < \frac{n_2}{N} + \frac{1}{2}. \end{aligned}$$

Therefore, we obtain an upper and lower boundary for t with

$$\frac{\max\{n_1, n_2\}}{N} - \frac{1}{2} \leq t < \frac{\min\{n_1, n_2\}}{N} + \frac{1}{2}. \tag{13}$$

The assumption $|n_1 - n_2| \leq N$ of Lemma 3.5 is crucial here. Otherwise, the upper bound could be smaller than the lower bound. The boundaries in Formula (13) will be plugged into the integral boundaries.

$$\begin{aligned} \frac{|n_1 - n_2|}{N} &= 1 - \left(\frac{\min\{n_1, n_2\}}{N} + \frac{1}{2} - \left(\frac{\max\{n_1, n_2\}}{N} - \frac{1}{2}\right)\right) \\ =1 - \int_{\frac{\max\{n_1, n_2\}}{N} - \frac{1}{2}}^{\frac{\min\{n_1, n_2\}}{N} + \frac{1}{2}} 1 \, dt &= 1 - \int_{-\infty}^{\infty} \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]^2} \left(\frac{n_1}{N} - t, \frac{n_2}{N} - t\right) \, dt. \end{aligned}$$

Note that $0 < \frac{n_1}{N} \leq 1$ and $0 < \frac{n_2}{N} \leq 1$. Thus

$$\begin{aligned} t &\geq \frac{\max\{n_1, n_2\}}{N} - \frac{1}{2} > -\frac{1}{2}, \\ t &< \frac{\min\{n_1, n_2\}}{N} + \frac{1}{2} \leq \frac{3}{2} \end{aligned}$$

and by substituting the integral boundaries, the second assertion follows. □

In the following, we illustrate the origin of the formula in Lemma 3.5 as a convolution formula of $f(t) = \mathbb{1}_{\left(-\frac{1}{2}, \frac{1}{2}\right]}(t)$. By Lemma 3.5, we have for $x, y \in \mathbb{R}$ with $|x - y| \leq 1$

$$|x - y| = 1 - \int_{-\infty}^{\infty} f(x - t) f(y - t)\, dt.$$

After substituting $v = x - t$, we obtain

$$\begin{aligned} |x - y| &= 1 - \int_{-\infty}^{\infty} f(v) f(y - x + v)\, dv \\ &= 1 - \int_{-\infty}^{\infty} f(v) f(x - y - v)\, dv = 1 - (f * f)(x - y) \end{aligned}$$

since the symmetry $f(x) = f(-x)$ holds under the integral. The last equation yields a convolution-based formula as a deeper interpretation (Aubin, 2000, p. 128):

$$1 - |x| = (f * f)(x), x \in [-1, 1].$$

Now, we continue considering Formula (11), p. 15,

$$\begin{aligned} N\left(d(\mathbf{E}) - \frac{1}{4}\right) =& \frac{N^2}{8\binom{N}{3}} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{2|n_1 - n_2|}{N} - 1\right) \psi(E_{n_1}) \psi(E_{n_2}) \\ =& \frac{N^2}{4\binom{N}{3}} \sum_{n_1, n_2 = 1}^{N} \left(\frac{|n_1 - n_2|}{N} - \frac{1}{2}\right) \psi(E_{n_1}) \psi(E_{n_2}) + \frac{N^2}{8\binom{N}{3}} \sum_{n=1}^{N} \psi(E_n)^2 \\ =& \frac{N^2}{4\binom{N}{3}} \sum_{n_1, n_2 = 1}^{N} \left(\frac{|n_1 - n_2|}{N} - \frac{1}{2}\right) \psi(E_{n_1}) \psi(E_{n_2}) + \frac{N^3}{8\binom{N}{3}} \end{aligned} \tag{14}$$

almost surely, since $\psi(E_n)^2 = 1$ holds for $n = 1, \ldots, N$ almost surely. Moreover $\frac{N^3}{8\binom{N}{3}} = \mathcal{O}(1)$, so we only need to focus on the first sum of the last formula for further asymptotic analysis.

By using Lemma 3.5, we obtain

$$\begin{aligned}
&\frac{N^2}{4\binom{N}{3}} \sum_{n_1,n_2=1}^{N} \left(\frac{|n_1 - n_2|}{N} - \frac{1}{2} \right) \psi(E_{n_1}) \psi(E_{n_2}) \\
=&\frac{N^3}{4\binom{N}{3}} \frac{1}{N} \sum_{n_1,n_2=1}^{N} \left(\frac{1}{2} - \int_{-\frac{1}{2}}^{\frac{3}{2}} \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n_1}{N} - t \right) \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n_2}{N} - t \right) dt \right) \psi(E_{n_1}) \psi(E_{n_2}) \\
=&\frac{N^3}{4\binom{N}{3}} \frac{1}{N} \Bigg(\frac{1}{2} \sum_{n_1,n_2=1}^{N} \psi(E_{n_1}) \psi(E_{n_2}) \\
&\qquad - \sum_{n_1,n_2=1}^{N} \int_{-\frac{1}{2}}^{\frac{3}{2}} \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n_1}{N} - t \right) \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n_2}{N} - t \right) \psi(E_{n_1}) \psi(E_{n_2}) \, dt \Bigg) \\
=&\frac{N^3}{4\binom{N}{3}} \left(\frac{1}{2} \left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \psi(E_n) \right)^2 - \int_{-\frac{1}{2}}^{\frac{3}{2}} \left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n}{N} - t \right) \psi(E_n) \right)^2 dt \right). \quad (15)
\end{aligned}$$

The asymptotic behavior of the first part in Formula (15) can be derived as in Remark 3.4. The second part will be simplified by considering $\mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n}{N} - t \right)$:

$$-\frac{1}{2} < \frac{n}{N} - t \leq \frac{1}{2} \quad \Leftrightarrow \quad N \left(t - \frac{1}{2} \right) < n \leq N \left(t + \frac{1}{2} \right).$$

In combination with $n \in \{1, \ldots, N\}$, this assertion is equivalent to

$$\max \left\{ \left\lfloor N \left(t - \frac{1}{2} \right) \right\rfloor, 0 \right\} + 1 \leq n \leq \min \left\{ \left\lfloor N \left(t + \frac{1}{2} \right) \right\rfloor, N \right\}.$$

For $t \in \left(-\frac{1}{2}, \frac{1}{2} \right)$, we obtain the following equivalent assertion:

$$1 \leq n \leq \left\lfloor N \left(t + \frac{1}{2} \right) \right\rfloor,$$

and for $t \in \left[\frac{1}{2}, \frac{3}{2} \right)$, it is equivalent to:

$$\left\lfloor N \left(t - \frac{1}{2} \right) \right\rfloor + 1 \leq n \leq N.$$

Then we have:

$$\sum_{n=1}^{N} \mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]} \left(\frac{n}{N} - t \right) \psi(E_n) = \begin{cases} \displaystyle\sum_{n=1}^{\lfloor N(t+\frac{1}{2}) \rfloor} \psi(E_n) & \text{for } t \in \left(-\frac{1}{2}, \frac{1}{2} \right), \\ \displaystyle\sum_{n=\lfloor N(t-\frac{1}{2}) \rfloor + 1}^{N} \psi(E_n) & \text{for } t \in \left[\frac{1}{2}, \frac{3}{2} \right), \end{cases}$$

$$= \begin{cases} \sum_{n=1}^{\lfloor N(t+\frac{1}{2})\rfloor} \psi(E_n) & \text{for } t \in \left(-\frac{1}{2}, \frac{1}{2}\right), \\ \sum_{n=1}^{N} \psi(E_n) - \sum_{n=1}^{\lfloor N(t-\frac{1}{2})\rfloor} \psi(E_n) & \text{for } t \in \left[\frac{1}{2}, \frac{3}{2}\right). \end{cases}$$

Note that the case $t \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ yields an empty sum for $t < -\frac{1}{2} + \frac{1}{N}$, which is set as zero by definition. Further, for $t \in \left[\frac{1}{2}, \frac{3}{2}\right)$, the second sum after the decomposition is empty for $\frac{1}{2} \leq t < \frac{1}{2} + \frac{1}{N}$.

Using the notation $S_t^N = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \psi(E_n)$, we simplify the following integrands:

$$\begin{aligned} &\int_{-\frac{1}{2}}^{\frac{3}{2}} \left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathbb{1}_{(-\frac{1}{2}, \frac{1}{2})} \left(\frac{n}{N} - t \right) \psi(E_n) \right)^2 dt \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(S^N_{t+\frac{1}{2}} \right)^2 dt + \int_{\frac{1}{2}}^{\frac{3}{2}} \left(S_1^N - S^N_{t-\frac{1}{2}} \right)^2 dt \\ &= \int_0^1 \left(S_t^N \right)^2 dt + \int_0^1 \left(S_1^N - S_t^N \right)^2 dt. \end{aligned} \tag{16}$$

To sum up, we obtain the following representation of the 3-depth according to Formula (14), p. 17, and Formula (15), p. 18,

$$N \left(d_3(\mathbf{E}) - \frac{1}{4} \right) = \frac{N^3}{8\binom{N}{3}} + \frac{N^3}{8\binom{N}{3}} \left(S_1^N \right)^2 - \frac{N^3}{4\binom{N}{3}} \left(\int_0^1 (S_t^N)^2 \, dt + \int_0^1 (S_1^N - S_t^N)^2 \, dt \right).$$

At the end of Chapter 3.2, we will see that the K-depth can be represented by a functional of a random walk $(S_t^N)_{t \in [0,1]}$ for all K.

3.2 Asymptotic integral representation for general K

The general derivation is structured in three steps.

1. Generalization of a product formula with the sign function ψ as in Lemma 3.3 (Chapter 3.2.1, p. 20)
2. Asymptotic negligence rule for sign products with high length (Chapter 3.2.2, p. 23). This calculation does not occur for the case $K = 3$
3. Simplification of the asymptotically relevant part - similar to the calculation after Lemma 3.3 for $K = 3$ (Chapter 3.2.3, p. 26)

3.2.1 Representation of the K-depth by products of sign functions

Analogous to $K = 3$ in Lemma 3.3, we will simplify the K-depth by a representation with products of the sign functions, see also Leckey et al. (2020). In order to avoid triple index notations, we write $n_{i(1)}, \ldots, n_{i(2L)}$ instead of $n_{i_1}, \ldots, n_{i_{2L}}$.

Lemma 3.6. *For $(x_1, \ldots, x_K)^\top$ with $x_{n_i} \neq 0$, $i = 1, ..., K$ and $K \in \mathbb{N}$ we have*

$$\begin{aligned} &\prod_{k=1}^{K} \mathbb{1}\{x_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{x_{n_k}(-1)^k < 0\} - \frac{1}{2^{K-1}} \\ =&\frac{1}{2^{K-1}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(x_{n_{i(j)}}), \end{aligned} \tag{17}$$

where $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ defines the sign function.

Formula (17) can be used to rewrite the summands of the K-depth

$$\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} - \frac{1}{2^{K-1}}$$

by a sum of products with even length up to $2\left\lfloor \frac{K}{2} \right\rfloor$ because it holds $E_{n_i} \neq 0$ almost surely for $i = 1, \ldots, K$ due to assumption (A2). E.g., we have the following representation for $K \in \{2, 3, 4\}$:

$$\begin{aligned} K = 2: \quad & -\frac{1}{2}\psi(E_{n_1})\psi(E_{n_2}), \\ K = 3: \quad & \frac{1}{4}(-\psi(E_{n_1})\psi(E_{n_2}) + \psi(E_{n_1})\psi(E_{n_3}) - \psi(E_{n_2})\psi(E_{n_3})), \\ K = 4: \quad & \frac{1}{8}\Big(\prod_{i=1}^{4} \psi(E_{n_i}) - \psi(E_{n_1})\psi(E_{n_2}) + \psi(E_{n_1})\psi(E_{n_3}) \\ & - \psi(E_{n_1})\psi(E_{n_4}) - \psi(E_{n_2})\psi(E_{n_3}) + \psi(E_{n_2})\psi(E_{n_4}) - \psi(E_{n_3})\psi(E_{n_4}) \Big). \end{aligned}$$

Note that these formulas lead to a positive or negative value if the signs of $E_{n_1}, \ldots, E_{n_K}$ are alternating or not, respectively. In order to prove Lemma 3.6, the following identity is useful (Leckey et al., 2020):

Lemma 3.7. *If $a_1, ..., a_K$ are arbitrary real numbers, then it holds for $K \in \mathbb{N}$*

$$\prod_{i=1}^{K} (a_i + 1) = \sum_{\ell=1}^{K} \sum_{1 \leq i(1) < \ldots < i(\ell) \leq K} \prod_{j=1}^{\ell} a_{i(j)} + 1. \tag{18}$$

Proof of Lemma 3.7: The assertion will be proved by induction for K. The base step for $K = 1$ is clear. If we assume the assertion holds for a fixed K, then it also holds for $K + 1$ as the following calculation shows:

$$\begin{aligned}
&\prod_{i=1}^{K+1}(a_i+1) = \left(\prod_{i=1}^{K}(a_i+1)\right)(a_{K+1}+1)\\
=&\left(\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}a_{i(j)}+1\right)(a_{K+1}+1)\\
=&\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}a_{i(j)}a_{K+1}+\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}a_{i(j)}+a_{K+1}+1\\
=&\sum_{\ell=1}^{K+1}\sum_{1\le i(1)<\ldots<i(\ell)\le K+1}\prod_{j=1}^{\ell}a_{i(j)}+1.
\end{aligned}$$

Thus, the assertion follows. □

Proof of Lemma 3.6: Note for $x \neq 0$:

$$\mathbb{1}\{x>0\} = \frac{1}{2}(\psi(x)+1), \qquad \mathbb{1}\{x<0\} = \frac{1}{2}(-\psi(x)+1). \tag{19}$$

Formula (19) and Lemma 3.7 imply:

$$\begin{aligned}
\prod_{i=1}^{K}\mathbb{1}\{x_i(-1)^i>0\} &= \frac{1}{2^K}\prod_{i=1}^{K}((-1)^i\psi(x_i)+1)\\
&= \frac{1}{2^K}\left(\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}(-1)^{i(j)}\psi(x_{i(j)})+1\right).
\end{aligned}$$

Similarly it holds:

$$\begin{aligned}
\prod_{i=1}^{K}\mathbb{1}\{x_i(-1)^i<0\} &= \frac{1}{2^K}\prod_{i=1}^{K}\left((-1)^{i+1}\psi(x_i)+1\right)\\
&= \frac{1}{2^K}\left(\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}(-1)^{\ell}\prod_{j=1}^{\ell}(-1)^{i(j)}\psi(x_{i(j)})+1\right)\\
&= \frac{1}{2^K}\left(\sum_{\substack{\ell=1\\ \ell \text{ even}}}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}(-1)^{i(j)}\psi(x_{i(j)})+1\right)\\
&\quad-\frac{1}{2^K}\sum_{\substack{\ell=1\\ \ell \text{ odd}}}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}\prod_{j=1}^{\ell}(-1)^{i(j)}\psi(x_{i(j)}).
\end{aligned}$$

That yields:

$$\begin{aligned}
&\prod_{i=1}^{K} \mathbb{1}\{x_i(-1)^i > 0\} + \prod_{i=1}^{K} \mathbb{1}\{x_i(-1)^i < 0\} \\
=&\frac{1}{2^{K-1}} \left(\sum_{\substack{\ell=1 \\ \ell \text{ even}}}^{K} \sum_{1 \le i(1) < \ldots < i(\ell) \le K} \prod_{j=1}^{\ell} (-1)^{i(j)} \psi(x_{i(j)}) + 1 \right) \\
=&\frac{1}{2^{K-1}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(x_{i(j)}) + \frac{1}{2^{K-1}}
\end{aligned}$$

and the assertion follows. □

Note that this is an alternative proof for the special case in Lemma 3.3. In the Master thesis Malcherczyk (2018a), Lemma 3.6 has already been found but proved by a less elegant induction calculation. We can apply Lemma 3.6 to obtain the following almost surely representation of the rescaled version of the K-depth.

Theorem 3.8. *For a random vector* $\mathbf{E} = (E_1, \ldots, E_N)^\top$ *satisfying* $\mathbb{P}(E_n \neq 0) = 1$ *for* $n = 1, \ldots, N$*, it holds almost surely*

$$\begin{aligned}
&N \left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}} \right) \\
=&\frac{N}{2^{K-1} \binom{N}{K}} \sum_{1 \le n_1 < \ldots < n_K \le N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(E_{n_{i(j)}}).
\end{aligned} \tag{20}$$

Proof of Theorem 3.8: By a direct consequence of Lemma 3.6, we obtain

$$\begin{aligned}
&N \left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}} \right) \\
=&\frac{N}{\binom{N}{K}} \sum_{1 \le n_1 < \ldots < n_K \le N} \left(\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} - \frac{1}{2^{K-1}} \right) \\
=&\frac{N}{2^{K-1} \binom{N}{K}} \sum_{1 \le n_1 < \ldots < n_K \le N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(E_{n_{i(j)}}).
\end{aligned}$$ □

We will exchange the sums in Formula (20) in the following by

$$\begin{aligned}
&N \left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}} \right) \\
=&\frac{N}{2^{K-1} \binom{N}{K}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \underbrace{\sum_{1 \le n_1 < \ldots < n_K \le N} \prod_{j=1}^{2L} \psi(E_{n_{i(j)}})}_{\text{asymptotic analysis in Chapter 3.2.2}}.
\end{aligned} \tag{21}$$

Formula (21) will be used to separate asymptotically negligible from asymptotically relevant parts of the K-depth in Chapter 3.2.2.

3.2.2 Asymptotic neglect of sign-products with higher length

The next theorem shows that the products of length four or higher in Formula (21) are asymptotic negligible as $N \to \infty$. In the following, m will denote the product length and M will denote the length of a tuple in the ordered sum.

Lemma 3.9. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1. Then for $m \in \{1,\dots,M\}$ and $b_N(n_1,\dots,n_M) = \mathcal{O}(N^B)$ with $B < \frac{m}{2} - M$ and arbitrary $1 \le i(1) < \dots < i(m) \le M$, it holds*

$$\sum_{1\le n_1<\dots<n_M\le N} b_N(n_1,\dots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}}) \longrightarrow 0$$

in probability as $N \to \infty$.
If $B < \frac{m-1}{2} - M$, then

$$\sum_{1\le n_1<\dots<n_M\le N} b_N(n_1,\dots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}}) \longrightarrow 0$$

almost surely as $N \to \infty$.

Lemma 3.9 describes the relation between B, the asymptotic order of the factor $b_N(n_1,\dots,n_M)$, and m, M such that the factor is able to dominate the term for large N. The condition $B < \frac{m}{2} - M$ shows that the length m of the product term should be sufficiently high comparing to the number of indices M to have this domination.

Proof of Lemma 3.9: Let m, M be fixed. Since $\mathbb{E}(\psi(E_n)) = 0$ and $E_1,\dots,E_N$ are independent, we obtain

$$\mathbb{E}\left(\sum_{1\le n_1<\dots<n_M\le N} b(n_1,\dots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\right) = 0.$$

Furthermore, $\psi(E_n)^2 = 1$ almost surely implies

$$\mathbb{E}\left(\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\psi(E_{\tilde{n}_{i(j)}})\right) = \begin{cases} 1, & \text{if } n_{i(j)} = \tilde{n}_{i(j)} \text{ for } j = 1,\dots,m, \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

Then

$$
\begin{aligned}
&\operatorname{var}\left(\sum_{1\le n_1<\ldots<n_M\le N} b_N(n_1,\ldots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\right)\\
=&\mathbb{E}\left(\left(\sum_{1\le n_1<\ldots<n_M\le N} b_N(n_1,\ldots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\right)^2\right)\\
=&\,\mathbb{E}\left(\sum_{\substack{1\le n_1<\ldots<n_M\le N\\ 1\le \tilde{n}_1<\ldots<\tilde{n}_M\le N}} b_N(n_1,\ldots,n_M)b_N(\tilde{n}_1,\ldots,\tilde{n}_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\psi(E_{\tilde{n}_{i(j)}})\right)\\
=&\,\mathcal{O}(N^{2B})\sum_{\substack{1\le n_1<\ldots<n_M\le N\\ 1\le \tilde{n}_1<\ldots<\tilde{n}_M\le N}}\mathbb{E}\left(\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\psi(E_{\tilde{n}_{i(j)}})\right)\\
\overset{(22)}{=}&\,\mathcal{O}(N^{2B})\sum_{\substack{1\le n_1<\ldots<n_M\le N\\ 1\le \tilde{n}_1<\ldots<\tilde{n}_M\le N}}\prod_{j=1}^{m}\mathbb{1}_{\mathcal{I}_{i(1),\ldots,i(m)}}(n_1,\ldots,n_M,\tilde{n}_1,\ldots,\tilde{n}_M)
\end{aligned}
\tag{23}
$$

where $\mathcal{I}_{i(1),\ldots,i(m)}$ denotes the set of $2M$-tuples $(n_1,\ldots,n_M,\tilde{n}_1,\ldots,\tilde{n}_M)$ with the property $n_{i(j)}=\tilde{n}_{i(j)}$ for $j\in\{1,\ldots,m\}$ according to Formula (22).
In the next step, we decompose the previous sum into the sum over indices containing $(n_{i(1)},\ldots,n_{i(m)})$ and not containing these indices. In order to simplify the combinatorial situation here, we also consider an upper boundary by allowing $n_{i(1)},\ldots,n_{i(m)}$ to have values from 1 to N and ignoring the restrictions between the other indices:

$$
\begin{aligned}
&\mathcal{O}(N^{2B})\sum_{\substack{1\le n_1<\ldots<n_M\le N\\ 1\le \tilde{n}_1<\ldots<\tilde{n}_M\le N}}\prod_{j=1}^{m}\mathbb{1}_{\mathcal{I}_{i(1),\ldots,i(m)}}(n_1,\ldots,n_M,\tilde{n}_1,\ldots,\tilde{n}_M)\\
\le&\,\mathcal{O}(N^{2B})\sum_{\substack{1\le n_1<\ldots<n_M\le N\\ 1\le \tilde{n}_1<\ldots<\tilde{n}_M\le N\\ \text{without } n_\ell,\tilde{n}_\ell:\,\ell\in\{i(1),\ldots,i(m)\}}}\;\sum_{1\le n_{i(1)}<\ldots<n_{i(m)}\le N}1\\
=&\,\mathcal{O}(N^{2B})\binom{N}{M-m}^2\binom{N}{m}=\mathcal{O}(N^{2B})\mathcal{O}(N^{2(M-m)})\mathcal{O}(N^m)\\
=&\,\mathcal{O}(N^{2B+m+2(M-m)})=\mathcal{O}(N^{2B+2M-m})\xrightarrow[N\to\infty]{}0,
\end{aligned}
\tag{24}
$$

since $2B+2M-m<0$ is equivalent to the assumption $B<\frac{m}{2}-M$. Note that on the one hand the factor $\binom{N}{m}=\mathcal{O}(N^m)$ is obtained by counting the number of indices $1\le n_{i(j)},\tilde{n}_{i(j)}\le N$ with $n_{i(j)}=\tilde{n}_{i(j)}$. On the other hand, the factor $\binom{N}{M-m}^2=\mathcal{O}(N^{2(M-m)})$ is obtained by counting the number of indices $1\le n_\ell,\tilde{n}_\ell\le N$ with $\ell\notin\{i(1),\ldots,i(m)\}$ and not necessarily $n_\ell=\tilde{n}_\ell$ so that we have $\binom{N}{M-m}$ to the power of two.

Chebyshev's inequality provides the convergence in probability since for all $\varepsilon > 0$:

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{1\le n_1<\ldots<n_M\le N} b_N(n_1,\ldots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\right|>\varepsilon\right)\\ \le&\frac{1}{\varepsilon^2}\operatorname{var}\left(\sum_{1\le n_1<\ldots<n_M\le N} b_N(n_1,\ldots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\right)\xrightarrow[N\to\infty]{} 0.\end{aligned}$$

The convergence in probability can be extended to convergence almost surely if $B < \frac{m-1}{2} - M$. The convergence in probability is then sufficiently quick and Borel-Cantelli's lemma can be applied. In order to see this consequence, we set for $N \ge M$

$$X_N := \sum_{1\le n_1<\ldots<n_M\le N} b_N(n_1,\ldots,n_M)\prod_{j=1}^{m}\psi(E_{n_{i(j)}})$$

and $X_N = 0$ otherwise. For arbitrary $\varepsilon > 0$, Chebyshev's inequality provides

$$\sum_{N=1}^{\infty}\mathbb{P}\left(|X_N|\ge\varepsilon\right)\le\sum_{n=1}^{\infty}\mathcal{O}(N^{2B+2M-m})=\sum_{N=1}^{\infty}\mathcal{O}(N^{-(1+\delta)})<\infty \tag{25}$$

with $B = \frac{m-1}{2} - M - \frac{\delta}{2}$ for arbitrary $\delta > 0$. The Borel-Cantelli lemma implies

$$\mathbb{P}\left(\limsup_{N\to\infty}\{|X_N|\ge\varepsilon\}\right)=0 \text{ for arbitrary } \varepsilon>0$$

which is an alternative characterization for $X_N \xrightarrow[N\to\infty]{} 0$ almost surely. □

The next theorem summarizes the asymptotics of the products with higher length.

Theorem 3.10. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1 and $\mathbf{E} = \mathbf{E}_N = (E_1,\ldots,E_N)^\top$. Then we have almost surely*

$$\begin{aligned}&N\left(d_K(\mathbf{E})-\frac{1}{2^{K-1}}\right)\\ =&\frac{N}{2^{K-1}\binom{N}{K}}\sum_{1\le i(1)<i(2)\le K}\sum_{1\le n_1<\ldots<n_K\le N}(-1)^{i(1)+i(2)}\psi(E_{n_{i(1)}})\psi(E_{n_{i(2)}})+o_{a.s.}(1).\end{aligned}$$

For $K \in \{2,3\}$, the o-term is exactly zero.

Proof of Theorem 3.10: According to Formula (21), we have

$$\begin{aligned}&N\left(d_K(\mathbf{E})-\frac{1}{2^{K-1}}\right)\\ =&\frac{N}{2^{K-1}\binom{N}{K}}\sum_{L=1}^{\lfloor\frac{K}{2}\rfloor}\sum_{1\le i(1)<\ldots<i(2L)\le K}\prod_{j=1}^{2L}(-1)^{i(j)}\sum_{1\le n_1<\ldots<n_K\le N}\prod_{j=1}^{2L}\psi(E_{n_{i(j)}})\end{aligned}$$

$$
\begin{aligned}
=&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le i(1)<i(2)\le K} \sum_{1\le n_1<\ldots<n_K\le N} (-1)^{i(1)+i(2)} \psi(E_{n_{i(1)}})\psi(E_{n_{i(2)}}) \\
+&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{L=2}^{\lfloor \frac{K}{2} \rfloor} \sum_{1\le i(1)<\ldots<i(2L)\le K} \prod_{j=2}^{2L} (-1)^{i(j)} \sum_{1\le n_1<\ldots<n_K\le N} \prod_{j=1}^{2L} \psi(E_{n_{i(j)}}) \qquad (26)
\end{aligned}
$$

almost surely. We only have to prove that Formula (26) is asymptotically negligible as $N \to \infty$. Note that for $K \in \{2,3\}$ Formula (26) is zero and the proof is trivial then. For $K \ge 4$ and $L = 2, \ldots, \lfloor \frac{K}{2} \rfloor$, we consider the expression

$$
\frac{N}{\binom{N}{K}} \sum_{1\le n_1<\ldots<n_K\le N} \prod_{j=1}^{2L} \psi(E_{n_{i(j)}})
$$

and check the conditions of Lemma 3.9. According to the notation of Lemma 3.9, we have $M = K$, $m = 2L$ and $B = 1 - K$ since

$$
b_N(n_1, \ldots, n_K) = \frac{N}{\binom{N}{K}} = \mathcal{O}(N^{1-K}).
$$

Thus, $B = 1 - K < \frac{2L-1}{2} - K$ is satisfied for $L \ge \frac{3}{2}$ and therefore, Lemma 3.9 yields

$$
\frac{N}{\binom{N}{K}} \sum_{1\le n_1<\ldots<n_K\le N} \prod_{j=1}^{2L} \psi(E_{n_i(j)}) \longrightarrow 0
$$

almost surely as $N \to \infty$. □

Theorem 3.10 implies that we can neglect the summands with product length four and higher in Formula (26) almost surely. In the following chapters, the asymptotically relevant part in Theorem 3.10 will be investigated for the further asymptotic analysis. A more detailed discussion of the asymptotically negligible part and efficient algorithms for it are given in Chapter 4.2.1. Due to the fact that this asymptotically negligible part is zero for $K \in \{2,3\}$, Lemma 3.6 and Lemma 3.9 have not been considered in Chapter 3.1.

3.2.3 Miscellaneous simplifications

After exchanging sums, Theorem 3.6 and Theorem 3.10 imply almost surely

$$
\begin{aligned}
&N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) \\
=&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le n_1<\ldots<n_K\le N} \sum_{1\le i(1)<i(2)\le K} (-1)^{i(1)+i(2)} \psi(E_{n_{i(1)}})\psi(E_{n_{i(2)}}) + o_{a.s.}(1).
\end{aligned} \qquad (27)
$$

In this chapter, we modify the ordered sum of indices $n_1 < \ldots < n_K$ to an ordinary summation of only two indices n_1, n_2 by some combinatorial arguments similar to the calculations after Lemma 3.3, p. 14, in Chapter 3.1 for $K = 3$. The reduction to two summation indices is possible because the representation in Formula (27) only depends on $n_{i(1)}$, $n_{i(2)}$ and the index of these two summation indices can be shifted arbitrarily.

Lemma 3.11. *For arbitrary* $(e_1, \ldots, e_N)^\top \in \mathbb{R}^N$ *and given* $(i(1), i(2)) \in \mathbb{N}^2$ *with* $1 \leq i(1) < i(2) \leq K$, *it holds*

$$\begin{aligned}&\sum_{1 \leq n_1 < \ldots < n_K \leq N} \psi(e_{n_{i(1)}}) \psi(e_{n_{i(2)}}) \\ =&\frac{1}{2} \sum_{1 \leq n_1 \neq n_2 \leq N} \binom{(n_1 \wedge n_2) - 1}{i(1) - 1} \binom{|n_1 - n_2| - 1}{i(2) - i(1) - 1} \binom{N - (n_1 \vee n_2)}{K - i(2)} \psi(e_{n_1}) \psi(e_{n_2})\end{aligned}$$

where we denote $n_1 \wedge n_2 := \min\{n_1, n_2\}$ *and* $n_1 \vee n_2 := \max\{n_1, n_2\}$. *Note that* $\binom{m}{n} = 0$ *for* $m < n$.

Proof of Lemma 3.11: Let $i(1), i(2)$ with $1 \leq i(1) < i(2) \leq K$ be fixed. We first determine a suitable integer $\kappa(K, i(1), i(2)) \in \mathbb{N}_0$ such that

$$\sum_{1 \leq n_1 < \ldots < n_K \leq N} \psi(e_{n_{i(1)}}) \psi(e_{n_{i(2)}}) = \sum_{1 \leq n_{i(1)} < n_{i(2)} \leq N} \kappa(K, i(1), i(2)) \psi(e_{n_{i(1)}}) \psi(e_{n_{i(2)}}).$$

Note that $\kappa(K, i(1), i(2))$ denotes the number of omitted combinations of the right sum. In order to compute $\kappa(K, i(1), i(2))$, we have to count these combinations between and around the indices $i(1)$ and $i(2)$. According to Figure 3, there are

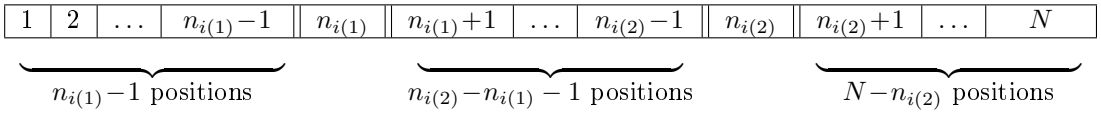

Figure 3: Number of positions between the indices $i(1)$ and $i(2)$ for general K

positions left from $n_{i(1)}$, between $n_{i(1)}$ and $n_{i(2)}$ and right from $n_{i(2)}$. We can pick $n_{i(1)} - 1$ positions from the left part, $n_{i(2)} - n_{i(1)} - 1$ positions from the middle part and $N - n_{i(2)}$ positions from the right part.
Figure 4 denotes the numbers of combinations in those three parts. Thus, we obtain

$$\kappa(K, i(1), i(2)) = \binom{n_{i(1)} - 1}{i(1) - 1} \binom{n_{i(2)} - n_{i(1)} - 1}{i(2) - i(1) - 1} \binom{N - n_{i(2)}}{K - i(2)}.$$

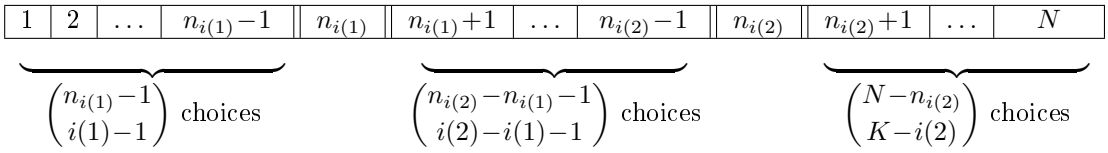

Figure 4: Number of combinations between the indices $i(1)$ and $i(2)$ for general K

By definition, $\binom{m}{n} = 0$ for $m < n$. This occurs if $n_{i(1)} < 1$, $n_{i(2)} - n_{i(1)} - 1 < i(2) - i(1) - 1$ or $n_{i(2)} > N - K + i(2)$, since it is impossible to have $n_{i(1)}, n_{i(2)}$ with such properties. The expression $\kappa(K, i(1), i(2))$ takes these cases into account. Further note that for $i(1) = 1$, we have no indices on the left side of $i(1)$, so the first binomial coefficient would be one (with $\binom{m}{0} = 1$ for $m \in \mathbb{N}_0$). For $i(1) + 1 = i(2)$, the fixed indices are next to each other so there are no indices between them and the second binomial coefficient would be one. For $i(2) = K$, there are no indices right from $i(2)$, thus the third binomial coefficient is one then. For $K = 3$, only one of these three binomial coefficients is not one and for $K = 4$, one of these binomial coefficients is always one. For general K, we need this representation with three binomial coefficients:

$$
\begin{aligned}
&\sum_{1\le n_1<\ldots<n_K\le N} \psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}}) \\
=&\sum_{1\le n_{i(1)}<n_{i(2)}\le N} \binom{n_{i(1)}-1}{i(1)-1}\binom{n_{i(2)}-n_{i(1)}-1}{i(2)-i(1)-1}\binom{N-n_{i(2)}}{K-i(2)}\psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}}) \\
=&\sum_{1\le n_1<n_2\le N} \binom{n_1-1}{i(1)-1}\binom{n_2-n_1-1}{i(2)-i(1)-1}\binom{N-n_2}{K-i(2)}\psi(e_{n_1})\psi(e_{n_2}) \\
=&\frac{1}{2}\sum_{1\le n_1\ne n_2\le N} \binom{(n_1\wedge n_2)-1}{i(1)-1}\binom{|n_1-n_2|-1}{i(2)-i(1)-1}\binom{N-(n_1\vee n_2)}{K-i(2)}\psi(e_{n_1})\psi(e_{n_2}).
\end{aligned}
$$

Note that the last equation holds by exchanging n_1 and n_2 with $n_1 \wedge n_2$ and $n_1 \vee n_2$ and by $|n_1 - n_2| = (n_1 \vee n_2) - (n_1 \wedge n_2)$. □

We use this auxiliary calculation to prove the next lemma.

Lemma 3.12. *For arbitrary $(e_1, \ldots, e_N)^\top \in \mathbb{R}^N$ and given $(i(1), i(2)) \in \mathbb{N}^2$ with $1 \le i(1) < i(2) \le K$, it holds*

$$
\begin{aligned}
&\sum_{1\le i(1)<i(2)\le K}\ \sum_{1\le n_1<\ldots<n_K\le N} (-1)^{i(1)+i(2)}\psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}}) \\
=&\frac{1}{2}\sum_{1\le n_1\ne n_2\le N}\left(-\sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}\right)\psi(e_{n_1})\psi(e_{n_2}).
\end{aligned}
$$

This combinatorial derivation can be proved more formal with Vandermonde's convolution (Gould and Srivastava, 1997):

Lemma 3.13 (Vandermonde's convolution)**.** *For integers $m, n, r \geq 0$, we have*

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}.$$

Proof of Lemma 3.12: We consider $J = i(2) - i(1) \in \{1, \ldots, K-1\}$. Note that

$$(-1)^{i(1)+i(2)} = (-1)^{2\cdot i(1)+i(2)-i(1)} = (-1)^J. \tag{28}$$

Using Lemma 3.11 and Formula (28), we have

$$\sum_{1\leq n_1<\ldots<n_K\leq N} (-1)^{i(1)+i(2)}\psi(E_{n_{i(1)}}, E_{n_{i(2)}}) = (-1)^J \sum_{1\leq n_1<\ldots<n_K\leq N} \psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}})$$
$$= \frac{(-1)^J}{2} \sum_{1\leq n_1\neq n_2\leq N} \binom{(n_1\wedge n_2)-1}{i(1)-1}\binom{|n_1-n_2|-1}{J-1}\binom{N-(n_1\vee n_2)}{K-i(1)-J}\psi(e_{n_1})\psi(e_{n_2}).$$

Our aim is to sum up over all combinations for $i(1)$ and $i(2)$, see Figure 5. We fix

	$i(1)=1$	$i(1)=2$	$\ldots$	$i(1)=K-2$	$i(1)=K-1$
$J=1$	(1,2)	(2,3)	$\ldots$	$(K-2,K-1)$	$(K-1,K)$
$J=2$	(1,3)	(2,4)	$\ldots$	$(K-2,K)$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$		
$J=K-2$	$(1,K-1)$	$(2,K)$			
$J=K-1$	$(1,K)$				

Figure 5: Illustration for the summation process of $(i(1), i(2))$ via $i(1)$ and J in the current calculation

a row index that is represented by J. We can sum up $i(1) \in \{1, \ldots, K-J\}$ for any fixed row J:

$$2(-1)^J \sum_{i(1)=1}^{K-J} \sum_{1\leq n_1<\ldots<n_K\leq N} (-1)^{i(1)+i(2)}\psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}})$$
$$= \sum_{i(1)=1}^{K-J} \sum_{1\leq n_1\neq n_2\leq N} \binom{(n_1\wedge n_2)-1}{i(1)-1}\binom{|n_1-n_2|-1}{J-1}\binom{N-(n_1\vee n_2)}{K-i(1)-J}\psi(e_{n_1})\psi(e_{n_2})$$
$$= \sum_{1\leq n_1\neq n_2\leq N} \binom{|n_1-n_2|-1}{J-1} \sum_{i(1)=0}^{K-J-1} \binom{(n_1\wedge n_2)-1}{i(1)}\binom{N-(n_1\vee n_2)}{K-J-1-i(1)}\psi(e_{n_1})\psi(e_{n_2})$$

$$= \sum_{1\le n_1\neq n_2\le N} \binom{|n_1-n_2|-1}{J-1}\binom{N-|n_1-n_2|-1}{K-J-1}\psi(e_{n_1})\psi(e_{n_2}),$$

where Vandermonde's convolution (Lemma 3.13) yields the last equality. If we sum up over $J \in \{1, ..., K-1\}$, we obtain all $\binom{K}{2}$ combinations for $1 \le i(1) < i(2) \le K$ so that

$$\begin{aligned}
&\sum_{1\le i(1)<i(2)\le K} \sum_{1\le n_1<\ldots<n_K\le N} (-1)^{i(1)+i(2)}\psi(e_{n_{i(1)}})\psi(e_{n_{i(2)}}) \\
=&\sum_{J=1}^{K-1} \frac{(-1)^J}{2} \sum_{1\le n_1\neq n_2\le N} \binom{|n_1-n_2|-1}{J-1}\binom{N-|n_1-n_2|-1}{K-J-1}\psi(e_{n_1})\psi(e_{n_2}) \\
=&\frac{1}{2}\sum_{1\le n_1\neq n_2\le N} \left(-\sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}\right)\psi(e_{n_1})\psi(e_{n_2}).
\end{aligned}$$

Thus, the assertion follows. □

Applying Lemma 3.12, we can continue the simplification of the K-depth.

Theorem 3.14. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1 and let $\mathbf{E} = \mathbf{E}_N = (E_1, \ldots, E_N)^\top$. Then we have almost surely*

$$N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) = -\frac{N}{2^K\binom{N}{K}} \sum_{1\le n_1\neq n_2\le N} \alpha_{K,N}(n_1,n_2)\psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1) \quad (29)$$

with $\alpha_{K,N}(n_1,n_2) = \sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}$.
For $K \in \{2,3\}$, the o-term is zero in Formula (29).

Note that the result in Formula (29) coincides with the situation in Formula (10), p. 15, for $K = 3$ in Chapter 3.1, since $\alpha_{3,N}(n_1,n_2) = N - 2|n_1-n_2|$.

Proof of Theorem 3.14: According to Theorem 3.10, it holds almost surely

$$\begin{aligned}
&N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) \\
=&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le i(1)<i(2)\le K} \sum_{1\le n_1<\ldots<n_K\le N} (-1)^{i(1)+i(2)}\psi(E_{n_{i(1)}})\psi(E_{n_{i(2)}}) + o_{a.s.}(1).
\end{aligned}$$

Lemma 3.12 then yields

$$\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le i(1)<i(2)\le K} \sum_{1\le n_1<\ldots<n_K\le N} (-1)^{i(1)+i(2)}\psi(E_{n_{i(1)}})\psi(E_{n_{i(2)}}) + o_{a.s.}(1)$$

$$= -\frac{N}{2^K\binom{N}{K}} \sum_{1\leq n_1\neq n_2\leq N} \alpha_{K,N}(n_1,n_2)\psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1)$$

and the assertion follows. □

Another asymptotic neglect

In the following, the inner factor $\alpha_{K,N}(n_1,n_2)$ will be decomposed into two parts for further simplifications of Formula (29). We will use the notation of the falling factorial $\langle N\rangle_K := \frac{N!}{(N-K)!}$.

Lemma 3.15. *For arbitrary $n_1,n_2\in\mathbb{N}$ with $1\leq n_1\neq n_2\leq N$, it holds*

$$\begin{aligned}\frac{N}{2^K\binom{N}{K}}\alpha_{K,N}(n_1,n_2) =& \frac{N}{2^K\binom{N}{K}}\sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}\\ =& \frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2} \qquad (30)\\ &+\frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{0\leq M_1<M_2\leq K-2} a_K(M_1,M_2)\frac{|n_1-n_2|^{M_1}}{N^{M_2}} \qquad (31)\end{aligned}$$

for suitable constants $a_K(M_1,M_2)\in\mathbb{R}$.

The first part of the decomposition in Formula (30) will be investigated in the further asymptotic analysis. The second part in Formula (31) will be neglected asymptotically.

Proof of Lemma 3.15: Note that $\binom{N}{K} = \frac{N^K}{K!}+\mathcal{O}(N^{K-1})$ for a fixed $K\in\{0,\dots,N\}$. For $K=0$, we have $\binom{N}{0}=1$ where $\mathcal{O}(N^{-1})$ corresponds to zero then. Further we have $|n_1-n_2|=\mathcal{O}(N)$ due to the fact that $|n_1-n_2|\leq N$. For a more convenient notation, we set $x := |n_1-n_2|$. We obtain:

$$\begin{aligned}\binom{x-1}{J} &= \frac{(x-1)^J}{J!}+\mathcal{O}(N^{J-1})\\ \binom{N-x-1}{K-2-J} &= \frac{(N-x-1)^{K-2-J}}{(K-2-J)!}+\mathcal{O}(N^{K-3-J})\end{aligned} \qquad (32)$$

for arbitrary values for $x\in\{0,\dots,N\}$ and $J\in\{0,\dots,K-2\}$. Formula (32) yields:

$$\begin{aligned}&\binom{x-1}{J}\binom{N-x-1}{K-2-J}\\ =&\frac{(x-1)^J}{J!}\frac{(N-x-1)^{K-2-J}}{(K-2-J)!}+\mathcal{O}(N^{K-3})\end{aligned}$$

$$=\frac{\binom{K-2}{J}}{(K-2)!}(x-1)^J(N-x-1)^{K-2-J}+\mathcal{O}(N^{K-3}).$$

This will be plugged into the sum over J. Then

$$\begin{aligned}
&\frac{N}{2^K\binom{N}{K}}\sum_{J=0}^{K-2}(-1)^J\binom{x-1}{J}\binom{N-x-1}{K-2-J}\\
=&\frac{N}{2^K\binom{N}{K}}\left(\sum_{J=0}^{K-2}\frac{\binom{K-2}{J}}{(K-2)!}(1-x)^J(N-x-1)^{K-2-J}+\mathcal{O}(N^{K-3})\right)\\
=&\frac{NK(K-1)}{2^K\langle N\rangle_K}(N-2x)^{K-2}+\mathcal{O}(N^{K-3})\\
=&\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\left(\frac{1}{2}-\frac{x}{N}\right)^{K-2}+\mathcal{O}(N^{K-3}),
\end{aligned}\tag{33}$$

where in the second last equation the Binomial Theorem is applied. The left summand corresponds exactly to Formula (30). Now, we have to identify the $\mathcal{O}(N^{K-3})$-term and show that it can be represented as in Formula (31). At first, we consider

$$\sum_{J=0}^{K-2}(-1)^J\binom{x-1}{J}\binom{N-x-1}{K-2-J}$$

and represent this sum as a polynomial in x and N. We do not need to specify the coefficients in detail, but we have to determine the possible degrees of the polynomials. In the second step, we apply the previous calculation from Formula (33). First note that for each $J=0,\dots,K-2$

$$\binom{x-1}{J}=\sum_{r=0}^{J}k_r(J)x^r \text{ and } \binom{N-x-1}{K-2-J}=\sum_{q_1=0}^{K-2-J}\sum_{q_2=0}^{K-2-J-q_1}k_{q_1,q_2}(J)N^{q_1}x^{q_2}$$

for suitable constants $k_r(J)$, $k_{q_1,q_2}(J)$. Multiplying these polynomials in x and N yields

$$\binom{x-1}{J}\binom{N-x-1}{K-2-J}=\sum_{q_1=0}^{K-2-J}\sum_{q_2=0}^{K-2-J-q_1}\sum_{r=0}^{J}k_r(J)k_{q_1,q_2}(J)x^{r+q_2}N^{q_1}.$$

Then, we substitute $r+q_2$ by p and q_1 by q. Further, we also sum up the constants $k_r(J)$ and $k_{q_1,q_2}(J)$ to one constant $\tilde{k}_{p,q}(J)$:

$$\sum_{q_1=0}^{K-2-J}\sum_{q_2=0}^{K-2-J-q_1}\sum_{r=0}^{J}k_r(J)k_{q_1,q_2}(J)x^{r+q_2}N^{q_1}=\sum_{q=0}^{K-2-J}\sum_{p=0}^{K-2-q}\tilde{k}_{p,q}(J)x^pN^q$$

for suitable constants $\tilde{k}_{p,q}(J)$. If we sum over all $J = 0, \ldots, K-2$, we obtain

$$\begin{aligned}
&\sum_{J=0}^{K-2}(-1)^J\binom{x-1}{J}\binom{N-x-1}{K-2-J} = \sum_{J=0}^{K-2}(-1)^J \sum_{q=0}^{K-2-J}\sum_{p=0}^{K-2-q} \tilde{k}_{p,q}(J)x^pN^q \\
=&\sum_{q=0}^{K-2}\sum_{p=0}^{K-2-q} c_K(p,q)x^pN^q = \sum_{\substack{0\le q\le K-2\\ 0\le p\le K-2-q}} c_K(p,q)x^pN^q
\end{aligned}$$

for suitable constants $c_K(p,q)$. According to the first calculation of this proof, see Formula (33), we obtain the following decomposition:

$$\begin{aligned}
&\frac{N}{2^K\binom{N}{K}}\sum_{J=0}^{K-2}(-1)^J\binom{x-1}{J}\binom{N-x-1}{K-2-J} \\
=&\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\left(\frac{1}{2}-\frac{x}{N}\right)^{K-2} + \frac{N}{2^K\binom{N}{K}}\sum_{\substack{0\le q\le K-3\\ 0\le p\le K-3-q}} c_K(p,q)x^pN^q,
\end{aligned}$$

since the case $q = K-2$ cannot appear in the second term or otherwise it would be a contradiction to the order $\mathcal{O}(N^{K-3})$ that we derived in Formula (33). Analogously, $p > K-3-q$ would be a contradiction to the order $\mathcal{O}(N^{K-3})$ as well. Further

$$\frac{N}{2^K\binom{N}{K}}\sum_{\substack{0\le q\le K-3\\ 0\le p\le K-3-q}} c_K(p,q)x^pN^q = \frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{\substack{0\le q\le K-3\\ 0\le p\le K-3-q}} c_K(p,q)\frac{x^p}{N^{K-2-q}}.$$

We substitute with $M_1 := p$ and $M_2 := K-2-q$. Since $0 \le q \le K-3$, it holds

$$1 \le M_2 \le K-2$$

and since $0 \le p \le K-3-q$, we have

$$0 \le M_1 \le K-3-q = M_2 - 1 < M_2.$$

Further, we also substitute $c_K(p, K-2-M_2) = a_K(M_1, M_2)$ and then obtain

$$\frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{\substack{0\le q\le K-3\\ 0\le p\le K-3-q}} c_K(p,q)\frac{x^p}{N^{K-2-q}} = \frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{0\le M_1<M_2\le K-2} a_K(M_1,M_2)\frac{x^{M_1}}{N^{M_2}}$$

and the assertion follows. □

Note that Formula (30) can be included into the sum of Formula (31) for the case

$M_1 = M_2$ according to the binomial theorem:

$$\begin{aligned}
&\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\\
=&\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\sum_{M=0}^{K-2}\binom{K-2}{M}\frac{|n_1-n_2|^M}{2^{K-2+M}N^M}(-1)^M\\
=&\frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{M=0}^{K-2}a_K(M,M)\frac{|n_1-n_2|^M}{N^M}.
\end{aligned}$$

Formula (30) denotes the asymptotically relevant part which will be analyzed in the following. The constants $a_K(M_1, M_2)$ of Formula (31) do not need to be stated explicitly for arbitrary K. Nevertheless for given K, they can be computed once and then be used for computations. Note that the second part does not occur in the derivation for $K = 3$, since $a_3(0, 1) = 0$ (cf. Chapter 4.2.2, for more details). For the asymptotic analysis, we only need to show that Formula (31) is asymptotically negligible. In order to show this, we apply Lemma 3.9, p. 23, in Chapter 3.2.2 again.

Corollary 3.16. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1. For integers M_1, M_2 with $M_1 < M_2 \leq K-2$, it holds almost surely*

$$\frac{1}{N}\sum_{1\leq n_1\neq n_2\leq N}\frac{|n_1-n_2|^{M_1}}{N^{M_2}}\psi(E_{n_1})\psi(E_{n_2}) \xrightarrow[N\to\infty]{} 0.$$

Proof of Corollary 3.16: With the notation of Lemma 3.9, p. 23, we have $M = 2$ for the product length and $m = 2$ for the number of indices in the sum. The choice of B corresponds to the bound:

$$b_N(n_1, n_2) = \frac{1}{N}\frac{|n_1-n_2|^{M_1}}{N^{M_2}} = \mathcal{O}(N^{M_1-M_2-1})$$

so that $B = M_1 - M_2 - 1$. The following inequalities are equivalent:

$$B < \frac{m}{2} - M - \frac{1}{2} \quad\Leftrightarrow\quad M_1 - M_2 - 1 < 1 - 2 - \frac{1}{2} \quad\Leftrightarrow\quad M_1 < M_2 - \frac{1}{2}.$$

As assumed, $M_1 < M_2$ holds. Since M_1 and M_2 are integers, $M_1 < M_2 - \frac{1}{2}$ holds as well. Thus, the condition of Lemma 3.9 is satisfied and the assertion follows. □

By Theorem 3.14, p. 30, Lemma 3.15, p. 31, and Corollary 3.16, we obtain the next theorem which is the extension of Formula (11) of the case $K = 3$.

Theorem 3.17. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1 and let $\mathbf{E} = \mathbf{E}_N = (E_1, \dots, E_N)^\top$. Then we have almost surely*

$$\begin{aligned} &N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) \\ =&-\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K} \sum_{1\le n_1\neq n_2\le N} \left(\frac{1}{2} - \frac{|n_1-n_2|}{N}\right)^{K-2} \psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1). \end{aligned} \tag{34}$$

For $K \in \{2,3\}$, the o-term is zero in Formula (34).

Note that in Malcherczyk et al. (2021), the asymptotic derivation starts to be different from this thesis at this point. The alternative resulting representation in this paper will be given in a remark at the end of this chapter.

Last steps to the integral representation for general K

At first, we introduce notations for some terms that will appear in the following.

Notation 3.18.

(i) For $n_1, n_2 \in \{1, \dots, N\}$ and $J \in \mathbb{N}_0$, we will have a similar integral representation to Lemma 3.5, p. 16:

$$\mathcal{I}(J, n_1, n_2) := \begin{cases} \displaystyle\int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \prod_{j=1}^{J} \mathbb{1}_{\left(-\frac{1}{2},\frac{1}{2}\right]^2}\left(\frac{n_1}{N} - t_j, \frac{n_2}{N} - t_j\right) d\mathbf{t}, & \text{for } J \in \mathbb{N}, \\ 1, & \text{for } J = 0. \end{cases}$$

(ii) For $t \in [0,1]$ and arbitrary $x_1, \dots, x_N \in \mathbb{R}$, we introduce the random walk:

$$S_t^N := S_t^N(x_1, \dots, x_N) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt\rfloor} \psi(x_n).$$

For expressing the whole path of a random walk, e.g. for all $t \in [0,1]$ simultaneously, we replace t by a bullet point to denote that a function is considered:

$$S_\bullet^N := \left(S_t^N\right)_{t\in[0,1]}.$$

(iii) For $\mathbf{t} = (t_1, \dots, t_J)^\top \in \mathbb{R}^J$ and $J \in \mathbb{N}$, we have:

$$W_J(\mathbf{t}) := (t_1 \wedge \dots \wedge t_J) + \frac{1}{2}, V_J(\mathbf{t}) := (t_1 \vee \dots \vee t_J) - \frac{1}{2}. \tag{35}$$

Recall that $x \wedge y$ denotes the minimum and $x \vee y$ the maximum of $x, y \in \mathbb{R}$.

(iv) For $J, N \in \mathbb{N}$, we will later modify the integration area by:

$$\mathcal{C}_{J,N} := \left\{ \mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J ; \lfloor N \cdot (V_J(\mathbf{t}) \vee 0) \rfloor + 1 \leq \lfloor N \cdot (W_J(\mathbf{t}) \wedge 1) \rfloor \right\}$$

We simplify the factor $\left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2}$ in Formula (34) in the next lemma.

Lemma 3.19. *Let $(e_1, \ldots, e_N)^\top \in \mathbb{R}^N$ with $e_n \neq 0$ for $n = 1, \ldots, N$. Then we have*

$$\begin{aligned} & -\frac{1}{N} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} \psi(e_{n_1})\psi(e_{n_2}) \\ = & -\sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \left(\frac{1}{N} \sum_{n_1, n_2 = 1}^{N} \mathcal{I}(J, n_1, n_2)\psi(e_{n_1})\psi(e_{n_2})\right) \\ & - \left(-\frac{1}{2}\right)^{K-2} \left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \psi(e_n)\right)^2 + \frac{1}{2^{K-2}} \end{aligned}$$

where the Notation 3.18 (i), p. 35, is used.

Proof of Lemma 3.19: By adding zero, we have

$$\begin{aligned} & -\frac{1}{N} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} \psi(e_{n_1})\psi(e_{n_2}) \\ = & -\frac{1}{N} \sum_{n_1, n_2 = 1}^{N} \left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} \psi(e_{n_1})\psi(e_{n_2}) + \frac{1}{2^{K-2}} \end{aligned} \tag{36}$$

since $\psi(e_n)^2 = 1$ by the fact that $e_n \neq 0$ for $n = 1, \ldots, N$. We will simplify the inner factor of the sum in Formula (36) by applying Lemma 3.5, p. 16, on the inner factor

$$\frac{1}{2} - \frac{|n_1 - n_2|}{N} = \mathcal{I}(1, n_1, n_2) - \frac{1}{2}.$$

Note that $\mathcal{I}(1, n_1, n_2)^J = \mathcal{I}(J, n_1, n_2)$ for $J \in \mathbb{N}_0$ due to Fubini's Theorem. Then, we obtain with the Binomial Theorem

$$\begin{aligned} & -\left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} = -\left(\mathcal{I}(1, n_1, n_2) - \frac{1}{2}\right)^{K-2} \\ = & -\sum_{J=0}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \mathcal{I}(1, n_1, n_2)^J = -\sum_{J=0}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \mathcal{I}(J, n_1, n_2). \end{aligned}$$

Furthermore, we will simplify Formula (36):

$$\begin{aligned}
&-\frac{1}{N}\sum_{n_1,n_2=1}^{N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\psi(e_{n_1})\psi(e_{n_2})+\frac{1}{2^{K-2}}\\
=&-\frac{1}{N}\sum_{J=0}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\left(\sum_{n_1,n_2=1}^{N}\mathcal{I}(J,n_1,n_2)\psi(e_{n_1})\psi(e_{n_2})\right)+\frac{1}{2^{K-2}}\\
=&-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\left(\frac{1}{N}\sum_{n_1,n_2=1}^{N}\mathcal{I}(J,n_1,n_2)\psi(e_{n_1})\psi(e_{n_2})\right)\\
&-\left(-\frac{1}{2}\right)^{K-2}\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\psi(e_n)\right)^2+\frac{1}{2^{K-2}}.
\end{aligned}\tag{37}$$

Recall that $\mathcal{I}(0,n_1,n_2)=1$ to obtain Formula (37). □

The next lemma shows how the integrals $\mathcal{I}(J,n_1,n_2)$ in Lemma 3.19 can be simplified in order to obtain a final representation that can be handled with the Functional Central Limit Theorem.

Lemma 3.20. *For arbitrary* $(e_1,\dots,e_N)^\top\in\mathbb{R}^N$ *and* $J\in\mathbb{N}$

$$\frac{1}{N}\sum_{n_1,n_2=1}^{N}\mathcal{I}(J,n_1,n_2)\psi(e_{n_1})\psi(e_{n_2})=\int_{\mathcal{C}_{J,N}}\left(S^N_{W_J(\mathbf{t})\wedge 1}-S^N_{V_J(\mathbf{t})\vee 0}\right)^2 d\mathbf{t}$$

where the Notation 3.18, p. 35, is used.

Proof of Lemma 3.20: First note that

$$\begin{aligned}
&\frac{1}{N}\sum_{n_1,n_2=1}^{N}\mathcal{I}(J,n_1,n_2)\psi(e_{n_1})\psi(e_{n_2})\\
=&\int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J}\frac{1}{N}\sum_{n_1,n_2=1}^{N}\prod_{j=1}^{J}\mathbb{1}_{\left(-\frac{1}{2},\frac{1}{2}\right]^2}\left(\frac{n_1}{N}-t_j,\frac{n_2}{N}-t_j\right)\psi(e_{n_1})\psi(e_{n_2})\,d\mathbf{t}\\
=&\int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J}\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\prod_{j=1}^{J}\mathbb{1}_{\left(-\frac{1}{2},\frac{1}{2}\right]}\left(\frac{n}{N}-t_j\right)\psi(e_n)\right)^2 d\mathbf{t}.
\end{aligned}$$

It remains to simplify expressions of the form

$$\sum_{n=1}^{N}\prod_{j=1}^{J}\mathbb{1}_{\left(-\frac{1}{2},\frac{1}{2}\right]}\left(\frac{n}{N}-t_j\right)\psi(e_n)$$

for $J\in\mathbb{N}$ and $\mathbf{t}=(t_1,...,t_J)^\top\in\left(-\frac{1}{2},\frac{3}{2}\right)^J$. The condition of the indicator function

can be rewritten by

$$\begin{aligned} & -\frac{1}{2} < \frac{n}{N} - t_j \leq \frac{1}{2} \text{ for all } j = 1, \ldots, J. \\ \Leftrightarrow \; & N\left(t_j - \frac{1}{2}\right) < n \leq N\left(t_j + \frac{1}{2}\right) \text{ for all } j = 1, \ldots, J. \end{aligned}$$

Since $n = 1, ..., N$, the last proposition is equivalent to

$$\left\lfloor N\left(t_j - \frac{1}{2}\right)\right\rfloor + 1 \leq n \leq \left\lfloor N\left(t_j + \frac{1}{2}\right)\right\rfloor \text{ for all } j = 1, \ldots, J.$$

Then, this is equivalent to

$$\left\lfloor N\left((t_1 \vee \ldots \vee t_J) - \frac{1}{2}\right)\right\rfloor + 1 \leq n \leq \left\lfloor N\left((t_1 \wedge \ldots \wedge t_J) + \frac{1}{2}\right)\right\rfloor$$

and according to Notation 3.18 (iii), p. 35, we can rewrite this assertion by

$$\lfloor N \cdot V_J(\mathbf{t})\rfloor + 1 \leq n \leq \lfloor N \cdot W_J(\mathbf{t})\rfloor .$$

Now, we can simplify the original expression by:

$$\sum_{n=1}^{N}\prod_{j=1}^{J}\mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]}\left(\frac{n}{N} - t_j\right)\psi(e_n) = \sum_{n=\lfloor N\cdot(V_J(\mathbf{t})\vee 0)\rfloor+1}^{\lfloor N\cdot(W_J(\mathbf{t})\wedge 1)\rfloor}\psi(e_n). \tag{38}$$

The sum in Formula (38) can be empty if $\lfloor N \cdot (V_J(\mathbf{t}) \vee 0)\rfloor + 1 > \lfloor N \cdot (W_J(\mathbf{t}) \wedge 1)\rfloor$. To take care of this case, we only consider $\mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J$ that are in the set $\mathcal{C}_{J,N}$ from Notation 3.18 (iv), p. 35. Thus, we have:

$$\sum_{n=\lfloor N\cdot(V_J(\mathbf{t})\vee 0)\rfloor+1}^{\lfloor N\cdot(W_J(\mathbf{t})\wedge 1)\rfloor}\psi(e_n) = \begin{cases} \sum_{n=1}^{\lfloor N\cdot(W_J(\mathbf{t})\wedge 1)\rfloor}\psi(e_n) - \sum_{n=1}^{\lfloor N\cdot(V_J(\mathbf{t})\vee 0)\rfloor}\psi(e_n), & \mathbf{t} \in \mathcal{C}_{J,N}, \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

The results in Formula (38) and (39) provide:

$$\begin{aligned} & \frac{1}{N}\sum_{n_1,n_2=2}^{N}\mathcal{I}(J, n_1, n_2)\psi(e_{n_1})\psi(e_{n_2}) \\ = & \int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J}\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\prod_{j=1}^{J}\mathbb{1}_{(-\frac{1}{2},\frac{1}{2}]}\left(\frac{n}{N} - t_j\right)\psi(e_n)\right)^2 d\mathbf{t} \\ = & \int_{\mathcal{C}_{J,N}}\left(S^N_{W_J(\mathbf{t})\wedge 1} - S^N_{V_J(\mathbf{t})\vee 0}\right)^2 d\mathbf{t} \text{ for } J \in \mathbb{N} \end{aligned} \tag{40}$$

and the assertion follows. □

The resulting integral in Formula (40) will be shortly denoted by:

Notation 3.21. *For $J, N \in \mathbb{N}$, we introduce the functional:*

$$\Upsilon_{J,N}(S^N_\bullet) := \int_{\mathcal{C}_{J,N}} \left(S^N_{W_J(\mathbf{t})\wedge 1} - S^N_{V_J(\mathbf{t})\vee 0}\right)^2 d\mathbf{t}. \tag{41}$$

where the Notation 3.18 (ii) - (iv), p. 35, is used.

The functional $\Upsilon_{J,N}$ maps càdlàg functions $f : [0,1] \to \mathbb{R}$ (i.e., f is right continuous and its left limits exist) to $\mathbb{R}$. Note that the complete path $(S^N_t)_{t\in[0,1]}$ is mapped by $\Upsilon_{J,N}$ so that we write $\Upsilon_{J,N}(S^N_\bullet)$. The domain set of this functional contains indeed functions in $[0,1]$, since $W_J(\mathbf{t}) \wedge 1 \in [0,1]$ and $V_J(\mathbf{t}) \vee 0 \in [0,1]$ for $\mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J$. We will introduce and discuss the function space of the càdlàg-functions and its topology in Chapter 3.3 in detail. Lemma 3.19 and Lemma 3.20 yield the final asymptotic integral representation of the K-depth:

Theorem 3.22. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1. Then there exists a functional $\mathbf{\Psi}_{K,N}$ such that almost surely*

$$N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) = \mathbf{\Psi}_{K,N}(S^N_\bullet) + o_{a.s.}(1) \tag{42}$$

where $\mathbf{\Psi}_{K,N}$ is defined by

$$\mathbf{\Psi}_{K,N}(S^N_\bullet) = \beta_{K,N}\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(S^N_\bullet) - \left(-\frac{1}{2}\right)^{K-2}\left(S^N_1\right)^2 + \frac{1}{2^{K-2}}\right)$$

with $\beta_{K,N} = \frac{N^K K(K-1)}{4\langle N\rangle_K}$ and Notation 3.18 (ii), p. 35, and Notation 3.21, p. 39. For $K \in \{2,3\}$, the o-term is zero in Formula (42).

Proof of Theorem 3.22: The assertion is a direct consequence of Theorem 3.17, p. 35, Lemma 3.19, p. 36, and Lemma 3.20, p. 37:

$$\begin{aligned}
&N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right)\\
&\overset{3.17}{=} -\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\sum_{1\le n_1\neq n_2\le N}\left(\frac{1}{2} - \frac{|n_1-n_2|}{N}\right)^{K-2}\psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1)\\
&= -\frac{\beta_{K,N}}{N}\sum_{1\le n_1\neq n_2\le N}\left(\frac{1}{2} - \frac{|n_1-n_2|}{N}\right)^{K-2}\psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1)
\end{aligned}$$

$$\overset{3.19}{=} -\beta_{K,N}\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\left(\frac{1}{N}\sum_{n_1,n_2=1}^{N}\mathcal{I}(J,n_1,n_2)\psi(E_{n_1})\psi(E_{n_2})\right)$$
$$-\beta_{K,N}\left(-\frac{1}{2}\right)^{K-2}\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\psi(E_n)\right)^2+\frac{\beta_{K,N}}{2^{K-2}}+o_{a.s.}(1)$$
$$\overset{3.20}{=}\beta_{K,N}\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(S^N_\bullet)-\left(-\frac{1}{2}\right)^{K-2}\left(S^N_1\right)^2+\frac{1}{2^{K-2}}\right)+o_{a.s.}(1)$$

and the assertion follows. □

E.g., Theorem 3.22 yields the following asymptotically equivalent representations for $K\in\{2,3\}$ almost surely:

$$K=2:\frac{1}{2}\left(1-\left(S^N_1\right)^2\right)\frac{N^2}{\langle N\rangle_2},$$
$$K=3:\frac{3}{2}\left(\frac{1}{2}+\frac{1}{2}\left(S^N_1\right)^2-\Upsilon_{1,N}(S^N_\bullet)\right)\frac{N^3}{\langle N\rangle_3}.$$

There, we have $\Upsilon_{1,N}(S^N_\bullet)=\int_{-\frac{1}{2}+\frac{1}{N}}^{\frac{3}{2}}\left(S^N_{(t+\frac{1}{2})\wedge 1}-S^N_{(t-\frac{1}{2})\vee 0}\right)^2dt$ due the fact that $\mathcal{C}_{1,N}=\left[-\frac{1}{2}+\frac{1}{N},\frac{3}{2}\right)$, compare with the derivation on page 18 for details. We can also rewrite this integral by:

$$\begin{aligned}\Upsilon_{1,N}(S^N_\bullet)&=\int_{-\frac{1}{2}+\frac{1}{N}}^{\frac{1}{2}}\left(S^N_{t+\frac{1}{2}}\right)^2dt+\int_{\frac{1}{2}}^{\frac{3}{2}}\left(S^N_1-S^N_{t-\frac{1}{2}}\right)^2dt\\&=\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(S^N_{t+\frac{1}{2}}\right)^2dt+\int_{\frac{1}{2}}^{\frac{3}{2}}\left(S^N_1-S^N_{t-\frac{1}{2}}\right)^2dt.\end{aligned}$$

The last equality holds by $S^N_{t+\frac{1}{2}}=0$ for $t<-\frac{1}{2}+\frac{1}{N}$. Note that this integral is equal to the result in Formula (16), p. 19, at the end of Chapter 3.1, where this expression has already appeared and been simplified. In this case, we can replace $\mathcal{C}_{1,N}$ by $\mathcal{C}_1:=\left(-\frac{1}{2},\frac{3}{2}\right)$ under the integral. Thus, we can define $\Upsilon_{1,N}=\Upsilon_1$, since the region of integration does not depend on N.
For $K\in\{4,5\}$, we have:

$$K=4:3\left(\frac{1}{4}-\frac{1}{4}\left(S^N_1\right)^2+\Upsilon_1(S^N_\bullet)-\Upsilon_{2,N}(S^N_\bullet)\right)\frac{N^4}{\langle N\rangle_4}+o_{a.s.}(1),$$
$$K=5:5\left(\frac{1}{8}+\frac{1}{8}\left(S^N_1\right)^2-\frac{3}{4}\Upsilon_1(S^N_\bullet)+\frac{3}{2}\Upsilon_{2,N}(S^N_\bullet)-\Upsilon_{3,N}(S^N_\bullet)\right)\frac{N^5}{\langle N\rangle_5}+o_{a.s.}(1).$$

The higher K, the more summands of the form $\Upsilon_{k,N}$ with $k=1,\dots,K-2$ appear

with another scaling. This characteristic of the K-depth will also reappear when we discuss efficient computations in Chapter 4.1 and Chapter 5.2.
The derivation in Chapter 3.2 is based on residuals under the true parameter. Note that the whole derivation can be redone for residuals that are not zero almost surely since only the limit theorems require more assumptions on the residuals. Then, we only have to replace the o-terms by rest terms which are not necessarily negligible.

Remark 3.23. *Let* $\mathbf{R}(\boldsymbol{\theta}) = (R_1(\boldsymbol{\theta}), \dots, R_N(\boldsymbol{\theta}))^\top$ *be a residual vector under an arbitrary parameter* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *of the model in Formula (1), p. 4, i.e.*

$$R_n(\boldsymbol{\theta}) = Y_n - g(\mathbf{X}_n, \boldsymbol{\theta}) = g(\mathbf{X}_n, \boldsymbol{\theta}^*) - g(\mathbf{X}_n, \boldsymbol{\theta}) + E_n \text{ for } n \in \{1, \dots, N\}$$

satisfying the condition

$$\mathbb{P}(R_n(\boldsymbol{\theta}) \neq 0) = 1 \text{ for } n = 1, \dots, N. \tag{43}$$

Then analogously to the derivation in Chapter 3.2, we can obtain the following representation

$$\begin{aligned}
&N\left(d_K(\mathbf{R}(\boldsymbol{\theta})) - \frac{1}{2^{K-1}}\right) \\
=&\beta_{K,N}\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(S^N_{\bullet,\boldsymbol{\theta}}) - \left(-\frac{1}{2}\right)^{K-2}\left(S^N_{1,\boldsymbol{\theta}}\right)^2 + \frac{1}{2^{K-2}}\right) + \mathcal{R}^N_{\boldsymbol{\theta}} \\
&\text{with } S^N_{t,\boldsymbol{\theta}} := \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nt \rfloor}\psi(R_n(\boldsymbol{\theta})), S^N_{\bullet,\boldsymbol{\theta}} := \left(S^N_{t,\boldsymbol{\theta}}\right)_{t\in[0,1]}.
\end{aligned}$$

The term $\mathcal{R}^N_{\boldsymbol{\theta}}$ *denotes the rest terms that can be neglected under Assumption 2.1, cf. Theorem 3.10, p. 25, and Theorem 3.17, p. 35. Note in this context that all results from Chapter 3.2 except of those two theorems only require the assumption in Formula (43) for deriving the given representation.*

The representation in Remark 3.23 will be used in Chapter 4 to construct an algorithm for the computation of the K-depth under an arbitrary $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. A sufficient condition of the required assumption on the residuals can be derived by

$$\begin{aligned}
\mathbb{P}(R_n(\boldsymbol{\theta}) \neq 0) =& 1 - \mathbb{P}(g(\mathbf{X}_n, \boldsymbol{\theta}^*) - g(X_n, \boldsymbol{\theta}) + E_n = 0) \\
=& 1 - \mathbb{P}(E_n = g(\mathbf{X}_n, \boldsymbol{\theta}) - g(\mathbf{X}_n, \boldsymbol{\theta}^*)) \overset{!}{=} 1, \\
\Leftrightarrow \quad & \mathbb{P}(E_n = g(\mathbf{X}_n, \boldsymbol{\theta}) - g(\mathbf{X}_n, \boldsymbol{\theta}^*)) \overset{!}{=} 0.
\end{aligned}$$

This condition is satisfied for $n = 1, \dots, N$ and arbitrary $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ if the errors $E_1, \dots, E_N$ have a continuous distribution, but it does not necessarily hold if errors have a discrete distribution. E.g., we consider errors with $\mathbb{P}(E_n = 1) = \frac{1}{2} = \mathbb{P}(E_n = -1)$ and the model function $g(\theta) = \theta$ with $\theta \in \mathbb{R}$ and true parameter θ^*. Then for $\theta := \theta^* \pm 1$, we can obtain residuals equal to zero with positive probability so that the equality in Remark 3.23 is not correct in this case.
Note also that we can obtain residuals equal to zero under continuous distributed errors in practice. For $g(\theta) = \theta$ with $\theta \in \mathbb{R}$ again, we consider continuous distributed errors in $\mathbb{R}$ this time. Then for realizations of the errors $e_1, \dots, e_N \in \mathbb{R}$, we obtain $r_n(\theta) = 0$ for $\theta = e_n$. Thus, at least one residual can be exactly zero. In practice, this will often not occur, since the machine grid has to be sufficiently fine. We neglect this case in the following.

Remark 3.24. *In Malcherczyk et al. (2021), the derivation of the asymptotic distribution is identical to this thesis up to Theorem 3.17 and then starts to differ. Instead of applying Lemma 3.19 and Lemma 3.20, another integral identity similar to Lemma 3.5, p. 16, is formulated there. For $J \geq 2$ and $a, b \in [c, d]$, it holds*

$$\begin{aligned}&(b-a)^J\\ =&J(J-1)\int_c^d\int_c^t (t-s)^{J-2}\mathbb{1}\{s \geq a\}\mathbb{1}\{t < b\} + (s-t)^{J-2}\mathbb{1}\{s \geq b\}\mathbb{1}\{t < a\}dsdt.\end{aligned}$$

It leads to the following representation of the rescaled K-depth for $K \geq 4$:

$$T_K(\mathbf{E}) = \frac{N^K(N-K)!}{N!}\widetilde{\boldsymbol{\Psi}}_K(S_\bullet^N) + o_{a.s.}(1)$$

almost surely, where $\widetilde{\boldsymbol{\Psi}}$ is defined by

$$\begin{aligned}\widetilde{\boldsymbol{\Psi}}_K(S_\bullet^N) =& -\frac{K!}{2(K-4)!}\int_{\frac{1}{2}}^1\int_0^{t-\frac{1}{2}}\left(\frac{1}{2}+s-t\right)^{K-4} S_s^N\left(S_1^N - S_t^N\right)dsdt\\ &-\frac{K!}{4(K-4)!}\int_{-\frac{1}{2}}^1\int_{t\vee 0}^{t+\frac{1}{2}}\left(\frac{1}{2}+t-s\right)^{K-4}\left((S_{s\wedge 1}^N - S_{t\vee 0}^N)^2 - ((s\wedge 1)-(t\vee 0))\right)dsdt,\end{aligned}$$

almost surely. A mathematical connection between the representations for $K = 3$ and $K \geq 4$ is not given in contrast to the representation in Theorem 3.22, p. 39.

3.3 Derivation of the asymptotic distribution

In Theorem 3.22, the rescaled K-depth is rewritten as

$$T_K(\mathbf{E}) = \mathbf{\Psi}_{K,N}(S_\bullet^N) + o_{a.s.}(1)$$

for a suitable chosen functional $\mathbf{\Psi}_{K,N}$. The càdlàg-process $S_\bullet^N = (S_t^N)_{t\in[0,1]}$ (i.e., the paths are right continuous and the left limits exist almost surely) has increments $(\psi(E_N))_{N\in\mathbb{N}}$ satisfying $\mathbb{E}(\psi(E_N)) = 0$ and $\mathrm{var}(\psi(E_N)) = 1$ for $N \in \mathbb{N}$. In Chapter 3.3.1, we will introduce a Functional Central Limit Theorem also known as Donsker's invariance principle which describes the convergence in distribution of such càdlàg-processes to the standard Brownian motion $B_\bullet = (B_t)_{t\in[0,1]}$ with respect to the Skorokhod topology:

$$S_\bullet^N = (S_t^N)_{t\in[0,1]} \xrightarrow[N\to\infty]{\mathcal{D}} (B_t)_{t\in[0,1]} = B_\bullet.$$

In Chapter 3.3.2, we will apply Donsker's invariance principle to prove

$$T_K(\mathbf{E}) = \mathbf{\Psi}_{K,N}(S_\bullet^N) + o_{a.s.}(1) \xrightarrow[N\to\infty]{} \mathbf{\Psi}_K(B_\bullet)$$

with an additional continuity argument by an extended version of the Continuous Mapping Theorem. In this extended Continuous Mapping Theorem, the composed functional $\mathbf{\Psi}_{K,N}$ is allowed to depend on N in contrast to the standard known version.

3.3.1 Functional Central Limit Theorem in the Skorokhod space

The Skorokhod space $\big(D[0,1], d_{D[0,1]}\big)$ is a metric space with

$$D[0,1] := \{f : [0,1] \to \mathbb{R}\,;\, f \text{ is right-continuous and left limits exist}\}$$

and equipped with the Skorokhod metric $d_{D[0,1]}$. In order to define the metric $d_{D[0,1]}$, we introduce the set of all homeomorphisms from $[0,1]$ to $[0,1]$:

$$\Lambda := \{\lambda : [0,1] \to [0,1]\,;\, \lambda \text{ is monotonically increasing and bijective}\}.$$

Further for any homeomorphism $\lambda \in \Lambda$, the term $\|\lambda\|^\circ$ measures the similarity of λ to the identity mapping $\mathrm{id}_{[0,1]}$ from $[0,1]$ to $[0,1]$ by

$$\|\lambda\|^\circ := \sup_{s\neq t\in[0,1]} \left|\log\left(\frac{\lambda(t)-\lambda(s)}{t-s}\right)\right|. \tag{44}$$

The quotient in Formula (44) can also be understood as measuring the slope of λ and how much it varies from 1. The absolute value of the logarithm can also increase arbitrarily for slopes smaller than 1. Therefore, too small and too large slopes indicate differences between λ and $\mathrm{id}_{[0,1]}$. This leads to some topological advantages as the completeness of the Skorokhod space comparing to other definitions of Skorokhod (1956) also introduced originally (Billingsley, 1999, p. 125). Note that $\|\lambda\|^\circ = 0$ if and only if λ is the identity mapping. Then for $f, g \in D[0,1]$, we define

$$d_{D[0,1]}(f,g) := \inf_{\lambda \in \Lambda} \left\{ \|\lambda\|^\circ \vee \sup_{t \in [0,1]} |f(t) - g(\lambda(t))| \right\}.$$

We can understand $g(\lambda(t))$ as a reparametrized version of $g \in D[0,1]$ in time by the homeomorphism $\lambda \in \Lambda$. This reparametrization is important in order to have an appropriate concept of convergence for functions in $D[0,1]$ with discontinuities. E.g., the sequence of functions $\mathbb{1}_{\left[0, \frac{1}{2} + \frac{1}{2N}\right)}$ for $N \in \mathbb{N}$ converges to $\mathbb{1}_{\left[0, \frac{1}{2}\right)}$ with respect to the Skorokhod topology. This convergence would fail if we only considered the uniform distance by the supremum (Billingsley, 1999, p. 124) since

$$\sup_{t \in [0,1]} |\mathbb{1}_{\left[0, \frac{1}{2} + \frac{1}{2N}\right)} - \mathbb{1}_{\left[0, \frac{1}{2}\right)}| = 1$$

for arbitrary N due to the discontinuities. For a better comparison, some time distortions by λ are allowed so that all time points where discontinuities appear can be matched. In our example, the time distortions have to shift the point $\frac{1}{2} + \frac{1}{2N}$ to $\frac{1}{2}$ such that a comparison by the supremum makes sense. For increasing N, these time distortions have to become smaller. The intensity of time distortions is measured by $\|\lambda\|^\circ$. Thus, the Skorokhod metric quantifies the uniform distance between two functions with this additional sufficiently small parametrization λ. If $d_{[0,1]}(f,g) \leq \varepsilon$ for $\varepsilon > 0$ then there is a function λ with

$$\|\lambda\|^\circ \leq \varepsilon \text{ and } \sup_{t \in [0,1]} |f(t) - g(\lambda(t))| \leq \varepsilon,$$

so the deviation of λ from $\mathrm{id}_{[0,1]}$ has to be simultaneously sufficiently small. Moreover,

$$f_N \xrightarrow[N \to \infty]{} f \text{ in } D[0,1] \text{ if and only if}$$

there exists a sequence $(\lambda_N)_{N \in \mathbb{N}} \subseteq \Lambda$ with $\sup_{t \in [0,1]} |\lambda_N(t) - t| \xrightarrow[N \to \infty]{} 0$ such that

$$\sup_{t \in [0,1]} |f_N(\lambda_N(t)) - f(t)| \xrightarrow[N \to \infty]{} 0.$$

We define the Borel-σ-field $\mathfrak{D}$ on $(D[0,1], d_{[0,1]})$ by the generated σ-field from the open sets. In order to introduce Donsker's invariance principle for càdlàg-processes, we need to clarify the definition of the Wiener measure in the Skorokhod space. Let $C[0,1] := \{f : [0,1] \to \mathbb{R}; f \text{ is continuous}\}$ with the uniform metric

$$d_{C[0,1]}(f,g) := \sup_{t\in[0,1]} |f(t) - g(t)| \text{ for } f, g \in C[0,1].$$

E.g. in Billingsley (1999), the Wiener measure W on $(C[0,1], d_{C[0,1]})$ with σ-field $\mathfrak{C}$, the Borel-σ-field based on the uniform topology, is constructed. The Wiener measure on $(D[0,1], d_{D[0,1]})$ can be constructed by the pushforward measure of the mapping

$$H : C[0,1] \to D[0,1], H(f) = f$$

since the mapping H is continuous and thus $\mathfrak{C}$-$\mathfrak{D}$-measurable. The continuity of H is a consequence that the Skorokhod topology coincides with the uniform topology on $C[0,1]$. Let $(f_N)_{N\in\mathbb{N}} \subseteq C[0,1]$ with $f_N \xrightarrow[N\to\infty]{} f \in C[0,1]$. Then

$$d_{D[0,1]}(H(f_N), H(f)) = d_{D[0,1]}(f_N, f) = d_{C[0,1]}(f_N, f) \xrightarrow[N\to\infty]{} 0$$

which implies the continuity of H. We will denote the associated pushward measure by W_H. The associated stochastic processes of W and W_H have the same finite-dimensional distributions. (Two stochastic processes $(X_t)_{t\in[0,1]}$ and $(Y_t)_{t\in[0,1]}$ have the same finite-dimensional distributions if for arbitrary $0 \le t_1 < \ldots < t_k \le 1$ and $k \in \mathbb{N}$, the joint distributions $(X_{t_1}, \ldots, X_{t_k})$ and $(Y_{t_1}, \ldots, Y_{t_k})$ are equal). Therefore, it makes sense to call W_H a Wiener measure (Billingsley, 1999, p. 146). Now, we can formulate Donsker's invariance principle in the Skorokhod space:

Theorem 3.25. *If $(X_N)_{N\in\mathbb{N}}$ is a sequence of i.i.d. random variables with $\mathbb{E}(X_1) = \mu$ and $\operatorname{var}(X_1) = \sigma^2 > 0$ then for*

$$S^N : [0,1] \to \mathbb{R},\ S^N(t) := S_t^N := \frac{1}{\sigma\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} (X_n - \mu)$$

the following convergence in distribution holds in $(D[0,1], \mathfrak{D})$ with respect to the Skorokhod topology

$$S_\bullet^N = (S_t^N)_{t\in[0,1]} \xrightarrow[N\to\infty]{\mathcal{D}} B_\bullet$$

where $B_\bullet = (B_t)_{t\in[0,1]}$ is the standard Brownian motion on $D[0,1]$ with $B_\bullet \sim W_H$.

3.3.2 Application of the extended Continuous Mapping Theorem

According to Theorem 3.22, the functional $\mathbf{\Psi}_{K,N}$ contains a sum of the $\Upsilon_{J,N}$

$$
\begin{aligned}
&N\left(d_K(\mathbf{E}) - \left(\frac{1}{2}\right)^{K-1}\right) = \mathbf{\Psi}_{K,N}(S_\bullet^N) + o_{a.s.}(1) \\
=&\beta_{K,N}\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(S_\bullet^N) - \left(-\frac{1}{2}\right)^{K-2}\left(S_1^N\right)^2 + \frac{1}{2^{K-2}}\right) + o_{a.s.}(1)
\end{aligned}
$$

with $\Upsilon_{J,N} : D[0,1] \to \mathbb{R}$, $\Upsilon_{J,N}(f) := \int_{\mathcal{C}_{J,N}} \left(f(W_J(\mathbf{t}) \wedge 1) - f(V_J(\mathbf{t}) \vee 0)\right)^2 d\mathbf{t}$ for $J = 1, \ldots, K-2$. Note that we can also rewrite $(S_1^N)^2 = \Upsilon_0(S_\bullet^N)$ with $\Upsilon_0(f) = f(1)^2$ by a functional evaluating $(S_t^N)_{t\in[0,1]}$. Applying the Continuous Mapping Theorem, we obtain immediately

$$
\Upsilon_0(S_\bullet^N) \xrightarrow[N\to\infty]{\mathcal{D}} \Upsilon_0(B_\bullet) = B_1^2,
$$

since Υ_0 is continuous as the square of a projection. It is more complicated to deal with $\Upsilon_{J,N}$, because the original Continuous Mapping Theorem cannot be applied due to the dependence on N. For this problem, we will use an extended version of the Continuous Mapping Theorem (Beutner and Zähle, 2016, Theorem C.1.).

Theorem 3.26 (Extended Continuous Mapping Theorem). *Let (M, d_M) be a separable[1] metric space with Borel-σ-field $\mathfrak{M}$. Let $(X_N)_{N\in\mathbb{N}}$ be a sequence of random variables and X a random variable in $(M, \mathcal{M})$ such that*

$$
X_N \xrightarrow[N\to\infty]{\mathcal{D}} X.
$$

Further, let $(\tilde{M}, d_{\tilde{M}})$ be a metric space with Borel-σ-field $\widetilde{\mathfrak{M}}$. Let $\Upsilon_N : M \to \tilde{M}$ for $N \in \mathbb{N}$ and $\Upsilon : M \to \tilde{M}$ be measurable mappings[2]. Assume that

$$
\begin{aligned}
&\text{for every sequence } (x_N)_{N\in\mathbb{N}} \subseteq M \text{ with } d_M(x_N, x) \xrightarrow[N\to\infty]{} 0 \text{ with } x \in M, \\
&\text{it holds } d_{\tilde{M}}(\Upsilon_N(x_N), \Upsilon(x)) \xrightarrow[N\to\infty]{} 0.
\end{aligned}
$$

Then $\Upsilon_N(X_N) \xrightarrow[N\to\infty]{\mathcal{D}} \Upsilon(X)$.

The separability of $(D[0,1], d_{D[0,1]})$ is proved in Billingsley (1999), p. 128. Therefore,

[1] A metric space (M, d_M) is called separable if a countable subset $M_0 \subseteq M$ exists that is dense in M with respect to the topology induced by d_M. (Bogachev and Smolyanov, 2020, p. 7)

[2] In Beutner and Zähle (2016), a weaker condition on the measurebility on the Υ_N is given which follows from the given condition here in Theorem 3.26.

Theorem 3.26 can be applied for our situation.
Now, we will introduce some preparations so that we can use the extended Continuous Mapping Theorem. First of all, we need to clarify how the set $\mathcal{C}_{J,N}$ behaves asymptotically. The limit of a set can be expressed by the pointwise limit of its indicator function.

Lemma 3.27. *Let*

$$\mathcal{C}_{J,N} = \left\{ \mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J ; \lfloor N \cdot (V_J(\mathbf{t}) \vee 0) \rfloor + 1 \leq \lfloor N \cdot (W_J(\mathbf{t}) \wedge 1) \rfloor \right\} \text{ and}$$

$$\mathcal{C}_J = \left\{ \mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J ; \max\{t_1, \ldots, t_J\} - \min\{t_1, \ldots, t_J\} < 1 \right\}.$$

Then

$$\mathbb{1}_{\mathcal{C}_{J,N}}(\mathbf{t}) \xrightarrow[N \to \infty]{} \mathbb{1}_{\mathcal{C}_J}(\mathbf{t}) \text{ for all } \mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J.$$

In Figure 6, the convergence of the indicator functions in Lemma 3.27 for $J = 2$ is shown for increasing sample sizes of $N \in \{5, 10, 25\}$. The white color represents the areas where the indicator function is 1. The following three cases $\left(-\frac{1}{2}, \frac{1}{2}\right)^2$, $\left[\frac{1}{2}, \frac{3}{2}\right)^2$

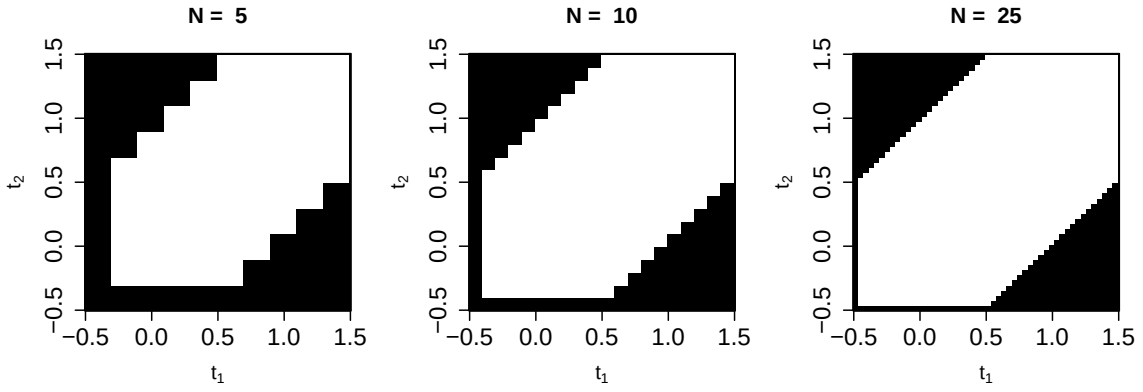

Figure 6: Illustration of the function $\mathbb{1}_{\mathcal{C}_{J,N}}$ for $J = 2$. White indicates 1 and black 0.

and $\left(-\frac{1}{2}, \frac{1}{2}\right) \times \left[\frac{1}{2}, \frac{3}{2}\right)$ (the reversed cartesian product is symmetric as the figure shows) will be investigated in the following proof. Note also that the characteristic form of those areas are determined by the position of the maximal and minimal value of the vector $\mathbf{t} = (t_1, \ldots, t_J)^\top$.

Proof of Lemma 3.27: Note that for $J = 1$, the assertion is straightforward to prove since $\mathcal{C}_{1,N} = \left[-\frac{1}{2} + \frac{1}{N}, \frac{3}{2}\right)$ for all $N \in \mathbb{N}$. For $J \geq 2$, we consider for each $\mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J$ the maximal and minimal value

$$t_{\max} := \max\{t_1, \ldots, t_J\} \text{ and } t_{\min} := \min\{t_1, \ldots, t_J\}.$$

The other entries in $\mathbf{t}$ can be ignored due to the fact that $\mathcal{C}_{J,N}$ and $\mathcal{C}_J$ only depend on $t_{\max}$ and $t_{\min}$. Then, we consider several cases for $t_{\max}$ and $t_{\min}$. In order to prove $\mathbb{1}_{\mathcal{C}_{J,N}}(\mathbf{t}) \xrightarrow[N\to\infty]{} \mathbb{1}_{\mathcal{C}_J}(\mathbf{t})$ for arbitrary $\mathbf{t} \in \left(-\frac{1}{2}, \frac{3}{2}\right)^J$ with $\mathbf{t} \in \mathcal{C}_J$, we have to show that an integer N^* exists such that the condition

$$\lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor + 1 \leq \lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor \tag{45}$$

is true for all $N \geq N^*$. For $\mathbf{t} \notin \mathcal{C}_J$, we will show that $\mathbf{t} \notin \mathcal{C}_{J,N}$ for all $N \in \mathbb{N}$.

Case 1 ($t_{\max}, t_{\min} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$):
It holds $\mathbf{t} \in \mathcal{C}_J$ for Case 1. We simplify the left and right side of Formula (45):

$$\begin{aligned}
\lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor &= \left\lfloor N\left(\left(t_{\max} - \frac{1}{2}\right) \vee 0\right)\right\rfloor = 0, \\
\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor &= \left\lfloor N\left(\left(t_{\min} + \frac{1}{2}\right) \wedge 1\right)\right\rfloor = \left\lfloor N\left(\left(t_{\min} + \frac{1}{2}\right)\right)\right\rfloor.
\end{aligned}$$

Thus, Formula (45) can be rewritten as

$$1 \leq \left\lfloor N\left(t_{\min} + \frac{1}{2}\right)\right\rfloor.$$

This inequality holds for all $N \geq \left\lceil \frac{1}{t_{\min}+\frac{1}{2}} \right\rceil =: N^*$ since $N^* \geq \frac{1}{t_{\min}+\frac{1}{2}}$.

Case 2 ($t_{\max} \in \left[\frac{1}{2}, \frac{3}{2}\right), t_{\min} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$):
Once again, we simplify

$$\begin{aligned}
\lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor &= \left\lfloor N\left(\left(t_{\max} - \frac{1}{2}\right) \vee 0\right)\right\rfloor = \left\lfloor N\left(t_{\max} - \frac{1}{2}\right)\right\rfloor, \\
\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor &= \left\lfloor N\left(\left(t_{\min} + \frac{1}{2}\right)\right)\right\rfloor.
\end{aligned}$$

Note that both situations $\mathbf{t} \in \mathcal{C}_J$ and $\mathbf{t} \notin \mathcal{C}_J$ can occur in Case 2.

Case 2.1 ($t_{\max} \in \left[\frac{1}{2}, \frac{3}{2}\right), t_{\min} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{t} \in \mathcal{C}_J$):
$\mathbf{t} \in \mathcal{C}_J$ implies that $-1 < t_{\min} - t_{\max} < 0$. According to Formula (45), we will show

that

$$\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor - \lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor \geq 1$$

for $N \geq N^* := \left\lceil \frac{2}{\delta} \right\rceil$ with $\delta := t_{\min} - t_{\max} + 1 \in (0,1)$. We have

$$\begin{aligned}
&\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor - \lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor = \left\lfloor N\left(t_{\min} + \frac{1}{2}\right) \right\rfloor - \left\lfloor N\left(t_{\max} - \frac{1}{2}\right) \right\rfloor \\
\geq &\left\lfloor N\left(t_{\min} + \frac{1}{2}\right) \right\rfloor - N\left(t_{\max} - \frac{1}{2}\right) > N\left(t_{\min} + \frac{1}{2}\right) - 1 - N\left(t_{\max} - \frac{1}{2}\right) \\
= &N\left(t_{\min} - t_{\max} + 1\right) - 1 = N\delta - 1 \geq \frac{2}{\delta}\delta - 1 = 1.
\end{aligned}$$

Note that this is a strict inequality.

Case 2.2 ($t_{\max} \in \left[\frac{1}{2}, \frac{3}{2}\right), t_{\min} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{t} \notin \mathcal{C}_J$):
$\mathbf{t} \notin \mathcal{C}_J$ yields $t_{\min} - t_{\max} \leq -1$. Here, we will prove that for all $N \in \mathbb{N}$ the negation of Formula (45) is correct. Note that $\delta := t_{\min} - t_{\max} + 1 \in (-1, 0]$ in Case 2.2.

$$\begin{aligned}
&\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor - \lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor = \left\lfloor N\left(t_{\min} + \frac{1}{2}\right) \right\rfloor - \left\lfloor N\left(t_{\max} - \frac{1}{2}\right) \right\rfloor \\
\leq &N\left(t_{\min} + \frac{1}{2}\right) - \left\lfloor N\left(t_{\max} - \frac{1}{2}\right) \right\rfloor < N\left(t_{\min} + \frac{1}{2}\right) - N\left(t_{\max} - \frac{1}{2}\right) + 1 \\
= &N\delta + 1 \leq 1,
\end{aligned}$$

since $N\delta \leq 0$. Thus, the negation of Formula (45) holds for arbitrary N.

Case 3 ($t_{\max}, t_{\min} \in \left[\frac{1}{2}, \frac{3}{2}\right)$):
It holds $\mathbf{t} \in \mathcal{C}_J$ in general for the Case 3 (see also Figure 6). Simplifying Formula (45) delivers

$$\begin{aligned}
\lfloor N(V_J(\mathbf{t}) \vee 0) \rfloor &= \left\lfloor N\left(t_{\max} - \frac{1}{2}\right) \right\rfloor, \\
\lfloor N(W_J(\mathbf{t}) \wedge 1) \rfloor &= \left\lfloor N\left(\left(t_{\min} + \frac{1}{2}\right) \wedge 1\right) \right\rfloor = N.
\end{aligned}$$

We consider $N \geq N^* = \left\lceil \frac{1}{1-\delta} \right\rceil$ with $\delta := t_{\max} - \frac{1}{2} \in (0,1)$. Note that this also implies $\delta \leq 1 - \frac{1}{N}$. Then, we rewrite Formula (45) as

$$\left\lfloor N\left(t_{\max} - \frac{1}{2}\right) \right\rfloor + 1 \leq N\left(t_{\max} - \frac{1}{2}\right) + 1 = N\delta + 1 \leq N\left(1 - \frac{1}{N}\right) + 1 = N$$

and the assertion follows. □

Further, we will need some properties and tools to deal with càdlàg-functions and their convergence in Skorokhod space (Billingsley, 1999, p. 122-124).

Lemma 3.28 (Properties of càdlàg-functions)**.** *Let $f \in D[0,1]$. Then*

(i) f is bounded, that is $\sup_{t\in[0,1]} |f(t)| < \infty$.

(ii) The set $U_f := \{t \in [0,1]; f \text{ is discontinuous at } t\}$ is countable.

(iii) If $(f_N)_{N\in\mathbb{N}} \subset D[0,1]$ is a sequence with $f_N \xrightarrow[N\to\infty]{} f$, then

$$\text{for } t \in [0,1] \setminus U_f : f_N(t) \xrightarrow[N\to\infty]{} f(t).$$

The first statement will be needed in order to apply the Dominated Convergence Theorem. The third assertion yields that the convergence with respect to the Skorokhod metric implies the pointwise convergence at continuity points of the limit f. The main idea of the following proof will be to neglect the points of discontinuity under the integral in the functional since they are countable and thus a null set with respect to the Lebesgue measure on $\mathbb{R}^J$.

Theorem 3.29. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 2.1. Then*

$$N\left(d(\mathbf{E}) - \frac{1}{2^{K-1}}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathbf{\Psi}_K(B_\bullet)$$

where $\mathbf{\Psi}_K$ is defined as

$$\mathbf{\Psi}_K(B_\bullet) = \beta_K \left(-\sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \Upsilon_J(B_\bullet) - \left(-\frac{1}{2}\right)^{K-2} (B_1)^2 + \frac{1}{2^{K-2}} \right)$$

with

$$\beta_K := \frac{K(K-1)}{4} \text{ and } \Upsilon_J(B_\bullet) := \int_{\mathcal{C}_J} \left(B_{W_J(\mathbf{t})\wedge 1} - B_{V_J(\mathbf{t})\vee 0}\right)^2 d\mathbf{t}$$

with $\mathcal{C}_J$ from Lemma 3.27, p. 47.

Proof of Theorem 3.29: Let $(f_N)_{N\in\mathbb{N}} \subseteq D[0,1]$ and $f \in D[0,1]$ be such that $f_N \xrightarrow[N\to\infty]{} f$ with respect to the Skorokhod topology. According to the extended Continuous Mapping Theorem 3.26, p. 46, it is sufficient to show that

$$\Upsilon_{J,N}(f_N) \xrightarrow[N\to\infty]{} \Upsilon_J(f)$$

with respect to the canonical topology on $\mathbb{R}$. We have

$$\begin{aligned}\lim_{N\to\infty} \Upsilon_{J,N}(f_N) &= \lim_{N\to\infty} \int_{\mathcal{C}_{J,N}} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\ &= \lim_{N\to\infty} \int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \mathbb{1}_{\mathcal{C}_{J,N}}(\mathbf{t})(f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t}.\end{aligned}$$

Due to Lemma 3.28 (i), there exists a constant $C > 0$ such that

$$\sup_{t\in[0,1]} |f(t)| < C \text{ and } \sup_{t\in[0,1]} |f_N(t)| < C$$

for all sufficiently large $N \in \mathbb{N}$. Moreover, $|\mathbb{1}_{\mathcal{C}_{J,N}}| \leq 1$ for all $N \in \mathbb{N}$. Thus, the integrand can be bounded and the Dominated Convergence theorem can be applied. This and Lemma 3.27, p. 47, imply

$$\begin{aligned}&\lim_{N\to\infty} \int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \mathbb{1}_{\mathcal{C}_{J,N}}(\mathbf{t})(f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\ &= \int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \mathbb{1}_{\mathcal{C}_J}(\mathbf{t}) \lim_{N\to\infty} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\ &= \int_{\mathcal{C}_J} \lim_{N\to\infty} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t}.\end{aligned}$$

The pointwise convergence of the integrand with

$$\lim_{N\to\infty} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 = (f(W_J(\mathbf{t}) \wedge 1) - f(V_J(\mathbf{t}) \vee 0))^2$$

holds for $\mathbf{t} \in \mathcal{C}_J$ if $W_J(\mathbf{t}) \wedge 1$ and $V_J(\mathbf{t}) \vee 0$ are continuous points of f, cf. Lemma 3.28 (iii). Our aim is to remove the part of the integration area $\mathcal{C}_J$ which leads to $W_J(\mathbf{t}) \wedge 1 \in U_f$ or $V_J(\mathbf{t}) \vee 0 \in U_f$ where

$$U_f := \{t \in [0,1]; f \text{ is discontinuous at } t\}.$$

Such an integration area which contains all resulting discountinuous points of f under the integral can be expressed by

$$\mathcal{U}_f = \{\mathbf{t} \in \mathcal{C}_J; W_J(\mathbf{t}) \wedge 1 \in U_f \text{ or } V_J(\mathbf{t}) \vee 0 \in U_f\}.$$

If we show that $\mathcal{U}_f$ is null set with respect to the Lebesgue measure on $\mathbb{R}^J$ then the

assertion follows from the following derivation:

$$\begin{aligned}
&\int_{\mathcal{C}_J} \lim_{N\to\infty} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\
&= \int_{\mathcal{C}_J \setminus \mathcal{U}_f} \lim_{N\to\infty} (f_N(W_J(\mathbf{t}) \wedge 1) - f_N(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\
&= \int_{\mathcal{C}_J \setminus \mathcal{U}_f} (f(W_J(\mathbf{t}) \wedge 1) - f(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} \\
&= \int_{\mathcal{C}_J} (f(W_J(\mathbf{t}) \wedge 1) - f(V_J(\mathbf{t}) \vee 0))^2 d\mathbf{t} = \Upsilon_J(f).
\end{aligned}$$

In order to show that $\mathcal{U}_f$ is a null set with respect to the Lebesgue measure on $\mathbb{R}^J$, we give a characterization of $\mathcal{U}_f$ by the following countable union

$$\mathcal{U}_f = \bigcup_{s \in U_f} \{\mathbf{t} \in \mathcal{C}_J; W_J(\mathbf{t}) \wedge 1 = s \text{ or } V_J(\mathbf{t}) \vee 0 = s\}.$$

If we show for arbitrary $s \in U_f$ that $\{\mathbf{t} \in \mathcal{C}_J; W_J(\mathbf{t}) \wedge 1 = s \text{ or } V_J(\mathbf{t}) \vee 0 = s\}$ is a null set with respect to the Lebesgue on $\mathbb{R}^J$, then the proof is finished since the countable union of null sets yields a null set as well. These sets can be characterized as follows:

$$\begin{aligned}
& W_J(\mathbf{t}) \wedge 1 = s \\
\Leftrightarrow\ & \mathbf{t} \in \mathcal{W}_{J,s} := \bigcup_{j=1}^{J} \left\{ \mathbf{t} \in \mathcal{C}_J; t_j + \frac{1}{2} = s, t_i + \frac{1}{2} \geq s \text{ for } i = 1, \ldots, J, i \neq j \right\}, \\
& V_J(\mathbf{t}) \vee 0 = s \\
\Leftrightarrow\ & \mathbf{t} \in \mathcal{V}_{J,s} := \bigcup_{j=1}^{J} \left\{ \mathbf{t} \in \mathcal{C}_J; t_j - \frac{1}{2} = s, t_i - \frac{1}{2} \leq s \text{ for } i = 1, \ldots, J, i \neq j \right\}.
\end{aligned}$$

This yields

$$\mathcal{U}_f = \bigcup_{s \in U_f} (\mathcal{W}_{J,s} \cup \mathcal{V}_{J,s}).$$

Note that $\mathcal{W}_{J,s}$ and $\mathcal{V}_{J,s}$ are null sets with respect to the Lebesgue measure on $\mathbb{R}^J$ since at least one coordinate is kept fixed on s. This proves that $\mathcal{U}_f$ is also a null set with respect to the Lebesgue measure and the assertion follows. □

3.4 Visualization of the asymptotic distribution

This paragraph presents the shape of the asymptotic distribution and gives a short overview about its properties. In Figure 7, the estimated densities of the K-depth for $K \in \{3, 4, 5, 6, 7\}$ are presented. In order to obtain this figure, we can consider a large

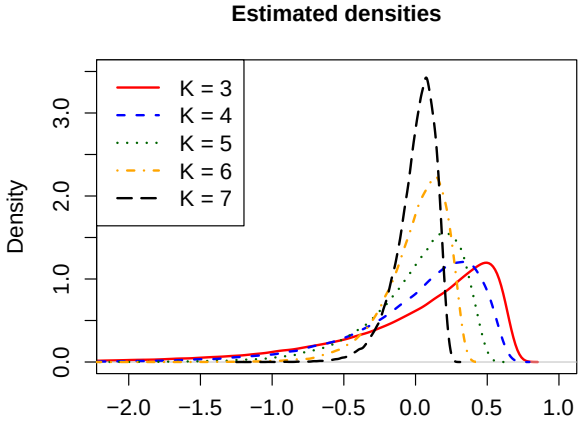

Figure 7: Estimated densities of the asymptotic distribution of the K-depth

number of realizations of the K-depth with a high sample size N. In detail, $N = 1{,}000$ and 100,000 repititions of the K-depth's asymptotically equivalent version (given in Theorem 3.22, p. 39) have been computed for this sample size. The default settings of the `density()`-function in `R` have been used to compute these estimated densities from the computed K-depths. Note that an efficient algorithm of the asymptotically equivalent version has been used here and will be presented in Chapter 4.1.
In Figure 7, we see that the asymptotic distributions of the K-depth are not symmetric. The upper boundary is $\frac{K(K-1)}{2^K}$ which is the limit of the rescaled K-depth T_K as $N \to \infty$ if the residuals $r_1, \ldots, r_N$ have alternating signs (Leckey et al., 2020):

$$\lim_{N\to\infty} T_K(r_1, \ldots, r_N) = \lim_{N\to\infty} N\left(d_K(r_1, \ldots, r_N) - \frac{1}{2^{K-1}}\right) = \frac{K(K-1)}{2^K}.$$

Moreover, the asymptotic distribution is unbounded on the other side. Since the minimal value of the K-depth is zero, we obtain for the rescaled K-depth T_K:

$$-\frac{N}{2^{K-1}} \xrightarrow[N\to\infty]{} -\infty.$$

Further, it is noticeable that the density mass becomes more concentrated around zero for higher K is. This effect is highlighted for $K \in \{3, 7, 10\}$ in Figure 8. Note that the axis for the density has a different scaling than in Figure 7. This effect is important to know because choosing large K can result in numeric problems. Due

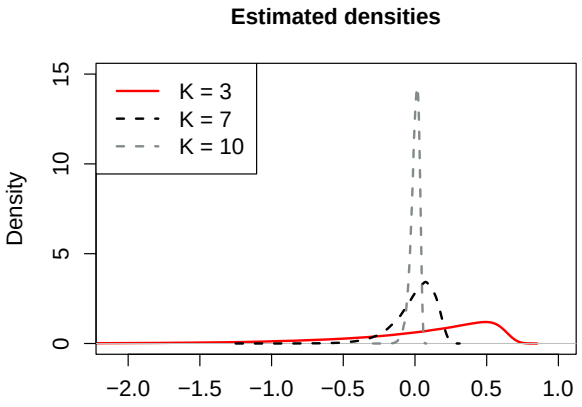

Figure 8: Estimated densities of the asymptotic distribution for higher K

to the stronger concentration around zero, the realizations of these K-depths vary by smaller distances the larger K. For sufficiently large K, these distances can be below the accuracy of the floating-point arithmetic and numeric errors can occur.

4 Efficient computation of the K-sign depth based on its asymptotic derivation

The derivation of the asymptotic distribution yields an (asymptotically equivalent) integral representation of the K-depth in Theorem 3.22, p. 39. Based on the derived formulas from Chapter 3.2, we will construct algorithms to compute the K-depth much more efficiently than a naive implementation according to Definition 2.2, p. 7. Such a *naive* implementation based on the definition has undesirable computational costs of $\Theta(N^K)$ since we have to compute K nested loops in order to compute the ordered sum over the indices $1 \leq n_1 < \ldots < n_K \leq N$.

In Chapter 4.1, we will show that the *asymptotic* representation of the K-depth can be computed in linear time. Moreover, this representation is exact for $K \in \{2, 3\}$. In Chapter 4.2, we will derive *exact* representations of the K-depth for $K \in \{4, 5\}$ in linear time and for $K = 6$ in quadratic time by analyzing the asymptotically neglected parts from Theorem 3.10, p. 25, and Theorem 3.17, p. 35.

In the following, we will use the notation from Remark 3.23, p. 41. For the decomposition of the K-depth, we assume for the residual vector $\mathbf{R} = (R_1, \ldots, R_N)^\top$ that $\mathbb{P}(R_n \neq 0) = 1$ for $n = 1, \ldots, N$ (the parameter $\boldsymbol{\theta}$ will be omitted in the notation):

$$N\left(d_K(\mathbf{R}) - \frac{1}{2^{K-1}}\right) = \boldsymbol{\Psi}_{K,N} + \mathcal{R}_1^N + \mathcal{R}_2^N,$$

where $\boldsymbol{\Psi}_{K,N}$ is the asymptotically relevant part of the K-depth in Theorem 3.22, p. 39, and $\mathcal{R}_1^N$ and $\mathcal{R}_2^N$ are the asymptotically neglected rest components. $\mathcal{R}_1^N$ describes the neglected products of length four or higher in Theorem 3.10, p. 25,

$$\mathcal{R}_1^N := \frac{N}{2^{K-1}\binom{N}{K}}\left(\sum_{1 \leq n_1 < \ldots < n_K \leq N} \sum_{L=2}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(R_{n_{i(j)}})\right)$$

which will be discussed in Chapter 4.2.1. $\mathcal{R}_2^N$ denotes the rest sum part in Theorem 3.17, p. 35,

$$\mathcal{R}_2^N := \frac{N^{K-1}}{2^K\binom{N}{K}} \sum_{1 \leq n_1 \neq n_2 \leq N} \sum_{0 \leq M_1 < M_2 \leq K-2} a_K(M_1, M_2)\psi(R_{n_1})\psi(R_{n_2}).$$

which will be discussed in Chapter 4.2.2. An implementation of the naive algorithm will be compared with the other implementations based on simulated runtimes for se-

veral sample sizes N in Chapter 4.3. All algorithms with linear time from Chapter 4 are included in the `R`-package `GSignTest` (Horn, 2021a) available on GitHub and implemented by the author of this thesis. All algorithms are based on `C++` implementations with the package `Rcpp` (Eddelbuettel and Francois, 2011).

4.1 Approximate computation

The approximate computation only considers the part $\mathbf{\Psi}_{K,N}$. The following lemma shows that this part can be computed in linear time for all K. Instead of considering the expression used for deriving the asymptotic distribution in Theorem 3.22, we investigate and simplify the representation of the K-depth in Theorem 3.17, p. 35.

Lemma 4.1. *Let* $S_{n,\alpha}^N = \sum_{k=1}^n \left(\frac{k}{N}\right)^\alpha \psi(R_n)$ *for* $n \in \{1,\dots,N\}$, $\alpha \geq 0$ *and let* $\mathbf{R} = (R_1,\dots,R_N)^\top$ *such that* $\mathbb{P}(R_n \neq 0) = 1$ *for* $n \in \{1,\dots,N\}$. *Then*

$$\begin{aligned} &N\left(d_K(\mathbf{R}) - \frac{1}{2^{K-1}}\right) \\ =&-\frac{N^{K-1}K(K-1)}{2\langle N\rangle_K}\sum_{j=0}^{K-2}\binom{K-2}{j}\sum_{n=2}^{N}\left(\frac{1}{2}-\frac{n}{N}\right)^{K-2-j}\psi(R_n)S_{n-1,j}^N + \mathcal{R}^N \end{aligned}$$

almost surely with $\mathcal{R}^N = \mathcal{R}_1^N + \mathcal{R}_2^N$. *For* $K \in \{2,3\}$, *the rest term* $\mathcal{R}^N$ *is zero.*

Proof of Lemma 4.1: According to Theorem 3.17, p. 35, and Remark 3.23, p. 41, a representation of the K-depth is given by

$$\begin{aligned} &N\left(d_K(\mathbf{R}) - \frac{1}{2^{K-1}}\right) \\ =&-\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\sum_{1\leq n_1\neq n_2\leq N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\psi(R_{n_1})\psi(R_{n_2}) + \mathcal{R}^N \end{aligned} \tag{46}$$

almost surely. Then

$$\begin{aligned} &\sum_{1\leq n_1\neq n_2\leq N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\psi(R_{n_1})\psi(R_{n_2}) \\ =&2\sum_{n_2=2}^{N}\sum_{n_1=1}^{n_2-1}\left(\frac{1}{2}-\frac{n_2-n_1}{N}\right)^{K-2}\psi(R_{n_1})\psi(R_{n_2}) \\ =&2\sum_{j=0}^{K-2}\binom{K-2}{j}\sum_{n_2=2}^{N}\left(\frac{1}{2}-\frac{n_2}{N}\right)^{K-2-j}\psi(R_{n_2})\sum_{n_1=1}^{n_2-1}\left(\frac{n_1}{N}\right)^{j}\psi(R_{n_1}) \\ =&2\sum_{j=0}^{K-2}\binom{K-2}{j}\sum_{n_2=2}^{N}\left(\frac{1}{2}-\frac{n_2}{N}\right)^{K-2-j}\psi(R_{n_2})S_{n_2-1,j}^N \end{aligned}$$

and the assertion follows from the recent calculation and Formula (46). □

For an efficient computation, we have to compute the following matrix in advance:

$$\mathtt{S} := \left(S^N_{\bullet,0}, \ldots, S^N_{\bullet,K-2}\right),$$

where $S^N_{\bullet,\alpha}$ denotes the column vector $(S^N_{1,\alpha}, \ldots, S^N_{N-1,\alpha})^\top$ for $\alpha = 0, \ldots, K-2$ which can be computed in linear time for all $\alpha \geq 0$. By memorizing the values of $S^N_{n-1,j}$ only once in a matrix, we avoid to compute them multiple times and obtain a more efficient computation. This algorithm is fairly easy to implement for general K.

4.2 Exact computation

In this chapter, we will derive formulas for implementations of the two asymptotically negligible parts $\mathcal{R}_1^N$ (in Chapter 4.2.1) and $\mathcal{R}_2^N$ (in Chapter 4.2.2). We will see that $\mathcal{R}_1^N$ is more complicated to compute efficiently. For $K \leq 5$, computations in linear time are presented and for $K \geq 6$, the introduced algorithm has time complexity $\Theta(N^{K-4})$. In contrast, $\mathcal{R}_2^N$ can be computed for arbitrary K in linear time.

4.2.1 Efficient computation of the sign products of higher length

We will analyze $\mathcal{R}_1^N$ step by step for $K \in \{4, 5, 6\}$. Note that $\mathcal{R}_1^N = 0$ for $K \in \{2, 3\}$ so that we start with the 4-depth.

4-sign depth

By Theorem 3.8, p. 22, we obtain for $K = 4$ the following decomposition in sign-products:

$$N\left(d_4(\mathbf{R}) - \frac{1}{8}\right) = \frac{N}{8\binom{N}{4}}\left(\sum_{1\leq n_1<\ldots<n_4\leq N}\ \sum_{1\leq i(1)<i(2)\leq 4}\ \prod_{j=1}^{2}(-1)^{i(j)}\psi(R_{n_{i(j)}}) + \sum_{1\leq n_1<\ldots<n_4\leq N}\ \prod_{j=1}^{4}\psi(R_{n_j})\right). \tag{47}$$

An efficient computation of the first sum can be done as mentioned in the proof of Lemma 4.1, p. 56 in combination with Theorem 3.17, p. 35. The second summand in Formula (47) is neglected in the derivation in Chapter 3.2.2 and corresponds to $\mathcal{R}_1^N$. It can be computed in linear time since it only depends on the number of positive or negative signs of the residual vector as the next lemma shows.

Lemma 4.2. *Let* $(x_1, \ldots, x_N)^\top \in \mathbb{R}^n$ *with* $x_n \neq 0$, $n \in \{1, \ldots, N\}$ *and let* $M := \sum_{n=1}^N \mathbb{1}\{x_n < 0\}$. *Then*

$$\sum_{1 \leq n_1 < \ldots < n_K \leq N} \prod_{j=1}^K \psi(x_{n_j}) = \sum_{k=0}^K (-1)^k \binom{M}{k} \binom{N-M}{K-k}. \tag{48}$$

M is the number of negative signs in the residual vector and can be computed in linear time. The case $\binom{n}{m} = 0$ for $n < m$ can appear in the formulas of Lemma 4.2 and describes impossible cases.

Proof of Lemma 4.2: Note that $\sum_{i=1}^K \mathbb{1}\{x_{n_i} < 0\} \in \{0, 1, \ldots, K\}$ for all ordered indices $1 \leq n_1 < \ldots < n_K \leq N$ for arbitrary K-tuples $(x_{n_1}, \ldots, x_{n_K})^\top$. Then, we can split the ordered sum as follows:

$$\begin{aligned}
\sum_{1 \leq n_1 < \ldots < n_K \leq N} \prod_{j=1}^K \psi(x_{n_j}) &= \sum_{k=0}^K \sum_{\substack{1 \leq n_1 < \ldots < n_K \leq N \\ \sum_{i=1}^K \mathbb{1}\{x_{n_i} < 0\} = k}} \prod_{j=1}^K \psi(x_{n_j}) \\
&= \sum_{k=0}^K \sum_{\substack{1 \leq n_1 < \ldots < n_K \leq N \\ \sum_{i=1}^K \mathbb{1}\{x_{n_i} < 0\} = k}} (-1)^k \cdot 1^{K-k} = \sum_{k=0}^K (-1)^k \binom{M}{k} \binom{N-M}{K-k}.
\end{aligned}$$

The last equality holds by counting all combinations of the given sum for each k. □

Formula (48) uses the fact that the product term only depends on the number of positive or negative signs. Thus, we have to count how many K-tuples of the residual vector exists with $0, \ldots, K$ negative signs by binomial coefficients. If the number of negative signs is even, we have a positive resulting value and vice versa for odd numbers. The first factor $\binom{M}{k}$ counts the number of all K-tuples with k negative indices and the second factor $\binom{N-M}{K-k}$ counts the complementary number $K-k$ of positive signs in all K-tuples. Note that Lemma 4.2 is formulated for arbitrary K and will be also used for the computation of the expressions for $K \geq 5$.

5-sign depth

By Theorem 3.8, p. 22, we obtain for $K = 5$

$$\begin{aligned}
N\left(d_5(\mathbf{R}) - \frac{1}{16}\right) = \frac{N}{16\binom{N}{5}} \Bigg(\sum_{1 \leq n_1 < \ldots < n_5 \leq N} \Bigg(\sum_{1 \leq i(1) < i(2) \leq 5} \prod_{j=1}^2 (-1)^{i(j)} \psi(R_{n_{i(j)}}) \\
+ \sum_{1 \leq i(1) < \ldots i(4) \leq 5} \prod_{j=1}^4 (-1)^{i(j)} \psi(R_{n_{i(j)}}) \Bigg) \Bigg).
\end{aligned} \tag{49}$$

Analogously to $K = 4$, we only have to simplify the second part of the sum in Formula (49). We can simplify the ordered sum over all 5-tuples $(n_1, \dots, n_5)^\top$ since the summands only contain four of five indices. Similarly to the calculations in Chapter 3.1 after Lemma 3.3, p. 14, the missing index can be omitted by compensating a multiplying factor corresponding to the number of the omitted combinations. Table 1 gives an overview off all possible situations.

$\psi(R_{n_1})$	$\psi(R_{n_2})$	$\psi(R_{n_3})$	$\psi(R_{n_4})$	$\psi(R_{n_5})$	**sign**	**omitted combinations**
✓	✓	✓	✓		$+$	$(N - n_4)$
✓	✓	✓		✓	$-$	$(n_5 - n_3 - 1)$
✓	✓		✓	✓	$+$	$(n_4 - n_2 - 1)$
✓		✓	✓	✓	$-$	$(n_3 - n_1 - 1)$
	✓	✓	✓	✓	$+$	$(n_2 - 1)$

Table 1: Omitted combinations of the products of length four for 5-sign depth

Thus, we can expand Formula (49) as follows:

$$
\begin{aligned}
&\sum_{1 \le n_1 < \dots < n_5 \le N} \sum_{1 \le i(1) < \dots < i(4) \le 5} \prod_{j=1}^{4} (-1)^{i(j)} \psi(R_{n_{i(j)}}) \\
&= \sum_{1 \le n_1 < \dots < n_4 \le N} (N - n_4) \prod_{\substack{i=1 \\ i \neq 5}}^{5} \psi(R_{n_i}) - \sum_{1 \le n_1 < n_2 < n_3 < n_5 \le N} (n_5 - n_3 - 1) \prod_{\substack{i=1 \\ i \neq 4}}^{5} \psi(R_{n_i}) \\
&+ \sum_{1 \le n_1 < n_2 < n_4 < n_5 \le N} (n_4 - n_2 - 1) \prod_{\substack{i=1 \\ i \neq 3}}^{5} \psi(R_{n_i}) - \sum_{1 \le n_1 < n_3 < n_4 < n_5 \le N} (n_3 - n_1 - 1) \prod_{\substack{i=1 \\ i \neq 2}}^{5} \psi(R_{n_i}) \\
&+ \sum_{1 \le n_2 < \dots < n_5 \le N} (n_2 - 1) \prod_{i=2}^{5} \psi(R_{n_i}).
\end{aligned}
$$

We can rename the indices to n_1, n_2, n_3, n_4 for all sums. Then, we obtain

$$
\begin{aligned}
&\sum_{1 \le n_1 < \dots < n_4 \le N} (N - n_4) \prod_{i=1}^{4} \psi(R_{n_i}) - \sum_{1 \le n_1 < \dots < n_4 \le N} (n_4 - n_3 - 1) \prod_{i=1}^{4} \psi(R_{n_i}) \\
&+ \sum_{1 \le n_1 < \dots < n_4 \le N} (n_3 - n_2 - 1) \prod_{i=1}^{4} \psi(R_{n_i}) - \sum_{1 \le n_1 < \dots < n_4 \le N} (n_2 - n_1 - 1) \prod_{i=1}^{4} \psi(R_{n_i}) \\
&+ \sum_{1 \le n_1 < \dots < n_4 \le N} (n_1 - 1) \prod_{i=1}^{4} \psi(R_{n_i}) \\
&= \sum_{1 \le n_1 < n_2 < n_3 < n_4 \le N} (N - 2n_4 + 2n_3 - 2n_2 + 2n_1) \prod_{i=1}^{4} \psi(R_{n_i}).
\end{aligned}
$$

Finally, we have the following form of the 5-depth:

$$N\left(d_5(\mathbf{R})-\frac{1}{16}\right)=\frac{N}{16\binom{N}{5}}\Bigg(\sum_{1\le n_1<\ldots<n_5\le N}\sum_{1\le i(1)<i(2)\le 5}\prod_{j=1}^{2}(-1)^{i(j)}\psi(R_{n_{i(j)}})$$
$$+\sum_{1\le n_1<n_2<n_3<n_4\le N}(N-2n_4+2n_3-2n_2+2n_1)\prod_{i=1}^{4}\psi(R_{n_i})\Bigg).$$

Hence it only remains to compute the following term efficiently:

$$\sum_{1\le n_1<n_2<n_3<n_4\le N}(N-2n_4+2n_3-2n_2+2n_1)\prod_{i=1}^{4}\psi(R_{n_i})$$
$$=N\sum_{1\le n_1<n_2<n_3<n_4\le N}\prod_{i=1}^{4}\psi(R_{n_i}) \tag{50}$$
$$-2\sum_{L=1}^{4}(-1)^L\sum_{1\le n_1<n_2<n_3<n_4\le N}n_L\prod_{i=1}^{4}\psi(R_{n_i}). \tag{51}$$

The sum in Formula (50) can be computed in linear time by Lemma 4.2. In Formula (51), we have the additional weight-factor n_L. Thus, the sum also depends on the position of the signs so that we cannot count them directly as in Lemma 4.2. Nevertheless, we can simplify the problem with an analogous idea by summing up over all values of n_L. The next lemma will yield a possibility to compute expressions of the form in Formula (51) in linear time.

Lemma 4.3. *Let $(x_1,\ldots,x_N)^\top\in\mathbb{R}^n$ be with $x_n\neq 0$, $n\in\{1,\ldots,N\}$. Further, we introduce the following notations for $j=1,\ldots,N+1$:*

$$\mathcal{M}_\ell(j):=\sum_{n=1}^{j-1}\mathbb{1}\{x_n<0\},\mathcal{M}_r(j):=\sum_{n=j+1}^{N}\mathbb{1}\{x_n<0\}.$$

Then for arbitrary mappings $g:\mathbb{R}\to\mathbb{R}$, it holds for $L\in\{1,\ldots,K\}$

$$(a)\sum_{1\le n_1<\ldots<n_K\le N}g(n_L)\prod_{i=1}^{K}\psi(x_{n_i})$$
$$=\sum_{j=L}^{N-K+L}\left(g(j)\psi(x_j)\left(\sum_{k_1=0}^{L-1}(-1)^{k_1}\iota_{k_1}^j\right)\left(\sum_{k_2=0}^{K-L}(-1)^{k_2}\rho_{k_2}^j\right)\right)$$
$$\text{with }\iota_{k_1}^j=\binom{\mathcal{M}_\ell(j)}{k_1}\binom{j-\mathcal{M}_\ell(j)-1}{L-1-k_1},\rho_{k_2}^j=\binom{\mathcal{M}_r(j)}{k_2}\binom{N-j-\mathcal{M}_r(j)}{K-L-k_2}.$$

Especially, the cases for $L = 1$ and $L = K$ can be simplified as follows:

$$(b) \sum_{1 \le n_1 < \ldots < n_K \le N} g(n_1) \prod_{i=1}^{K} \psi(x_{n_i}) = \sum_{j=1}^{N-K+1} g(j)\psi(x_j) \left(\sum_{k=0}^{K-1} (-1)^k \rho_k^j \right),$$

$$(c) \sum_{1 \le n_1 < \ldots < n_K \le N} g(n_K) \prod_{i=1}^{K} \psi(x_{n_i}) = \sum_{j=K}^{N} g(j)\psi(x_j) \left(\sum_{k=0}^{K-1} (-1)^k \iota_k^j \right).$$

Note that the case $\binom{n}{m} = 0$ for $n < m$ can appear in the formulas of Lemma 4.3 and describes impossible cases. We set $\binom{0}{0} := 1$. According to Formula (51), $g : \mathbb{R} \to \mathbb{R}$ will be chosen as the identity mapping. Due to further computations of the K-depths for $K \ge 6$, it is useful to formulate Lemma 4.3 for arbitrary g. $\mathcal{M}_\ell(j)$ is the number of negative signs on the left side from the index j. At the beginning

$$(\mathcal{M}_\ell(2), \ldots, \mathcal{M}_\ell(N+1))^\top$$

can be simply computed in linear time in N where $\mathcal{M}_\ell(1) := 0$. The number of negative signs on the right side from the index can be computed with

$$\mathcal{M}_r(j) = \mathcal{M}_\ell(N+1) - \mathcal{M}_\ell(j+1) \text{ for } j = 1, \ldots, N.$$

The formulas in Lemma 4.3 can be computed in linear time as well and thus we can reduce the time complexity from $\Theta(N^K)$ to $\Theta(N)$.

Proof of Lemma 4.3: We begin with (a). For every K-tuple $(n_1, \ldots, n_K)^\top$, it holds $1 \le n_1 < \ldots < n_K \le N$. This yields that $L \le n_L \le N - K + L$ so that n_L can be replaced by a sum over $\{L, \ldots, N - K + L\}$ as the following equation shows:

$$\sum_{1 \le n_1 < \ldots < n_K \le N} g(n_L) \prod_{i=1}^{K} \psi(x_{n_i}) = \sum_{j=L}^{N-K+L} g(j)\psi(x_j) \sum_{(n_i) \in \mathcal{B}_L^j} \prod_{\substack{i=1 \\ i \ne j}}^{K} \psi(x_{n_i})$$

where we use the following notation:

$$\mathcal{B}_L^j := \{(n_i)_{i \in \{1,\ldots,K\} \setminus \{L\}}; (n_1, \ldots, n_K)^\top \in \{1, \ldots, N\}^K \\ \text{with } n_{i(1)} < n_{i(2)} \text{ for all } i(1) < i(2) \text{ and } n_L = j\}.$$

For the next step, we consider the following subsets of $\mathcal{B}_L^j$:

$$\mathcal{B}_L^j(k_1, k_2) := \left\{ (n_i) \in \mathcal{B}_L^j; \sum_{i=1}^{L-1} \mathbb{1}\{x_{n_i} < 0\} = k_1 \text{ and } \sum_{i=L+1}^{K} \mathbb{1}\{x_{n_i} < 0\} = k_2 \right\}.$$

$\mathcal{B}_L^j(k_1,k_2)$ describes the set of vectors with indices $(n_1,\dots,n_{L-1},n_{L+1},\dots,n_K)^\top$ which have k_1 negative signs in the subvector $(n_1,\dots,n_{L-1})^\top$ and k_2 negative sign in the subvector $(n_{L+1},\dots,n_K)^\top$. Note that

$$\sum_{i=1}^{L-1}\mathbb{1}\{x_{n_i}<0\}\in\{0,\dots,L-1\} \text{ and } \sum_{i=L+1}^{K}\mathbb{1}\{x_{n_i}<0\}\in\{0,\dots,K-L\}.$$

Then, we obtain

$$\begin{aligned}
&\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\sum_{(n_i)\in\mathcal{B}_L^j}\prod_{\substack{i=1\\ i\neq j}}^{K}\psi(x_{n_i})\\
&=\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\sum_{k_1=0}^{L-1}\sum_{k_2=0}^{K-L}\sum_{(n_i)\in\mathcal{B}_L^j(k_1,k_2)}\prod_{\substack{i=1\\ i\neq j}}^{K}\psi(x_{n_i})\\
&=\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\sum_{k_1=0}^{L-1}\sum_{k_2=0}^{K-L}\sum_{(n_i)\in\mathcal{B}_L^j(k_1,k_2)}(-1)^{k_1+k_2}\\
&=\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\sum_{k_1=0}^{L-1}\sum_{k_2=0}^{K-L}(-1)^{k_1+k_2}\kappa^j_{(k_1,k_2)}
\end{aligned}$$

where the cardinality of the set $\mathcal{B}_L^j(k_1,k_2)$ is denoted as $\kappa^j_{(k_1,k_2)}$ and can be stated explicitly for given $\mathcal{M}_\ell(j)$ and $\mathcal{M}_r(j)$ by the following formula

$$\kappa^j_{(k_1,k_2)}=\binom{\mathcal{M}_\ell(j)}{k_1}\binom{j-\mathcal{M}_\ell(j)-1}{L-1-k_1}\binom{\mathcal{M}_r(j)}{k_2}\binom{N-j-\mathcal{M}_r(j)}{K-L-k_2}.$$

We denote $\iota^j_{k_1}=\binom{\mathcal{M}_\ell(j)}{k_1}\binom{j-\mathcal{M}_\ell(j)-1}{L-1-k_1}$ and $\rho^j_{k_2}=\binom{\mathcal{M}_r(j)}{k_2}\binom{N-j-\mathcal{M}_r(j)}{K-L-k_2}$. Then the assertion follows by splitting $\kappa^j_{(k_1,k_2)}$ to $\iota^j_{k_1}$ and $\rho^j_{k_2}$:

$$\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\sum_{k_1=0}^{L-1}\sum_{k_2=0}^{K-L}(-1)^{k_1+k_2}\kappa^j_{(k_1,k_2)} \tag{52}$$

$$=\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\left(\sum_{k_1=0}^{L-1}(-1)^{k_1}\iota^j_{k_1}\right)\left(\sum_{k_2=0}^{K-L}(-1)^{k_2}\rho^j_{k_2}\right). \tag{53}$$

(b) follows from (a) for $L=1$

$$\begin{aligned}
\sum_{1\le n_1<\dots<n_K\le N} g(n_L)\prod_{i=1}^{K}\psi(x_{n_i}) &= \sum_{j=1}^{N-K+1} g(j)\psi(x_j)\left(\sum_{k_1=0}^{0}(-1)^{k_1}\iota^j_{k_1}\right)\left(\sum_{k_2=0}^{K-L}(-1)^{k_2}\rho^j_{k_2}\right)\\
&=\sum_{j=L}^{N-K+L} g(j)\psi(x_j)\left(\sum_{k_2=0}^{K-L}(-1)^{k_2}\rho^j_{k_2}\right)
\end{aligned}$$

since $\iota_0^j = 1$ for $L = 1$.

(c) follows from (a) for $L = K$ analogously to (b). □

Splitting $\kappa^j_{(k_1,k_2)}$ to $\iota^j_{k_1}$ and $\rho^j_{k_2}$ is useful for the computation since the number of arithmetic iterations can thereby be reduced from Formula (52) with $(L-1)(K-L)$ summands to Formula (53) with $K-1$ summands.

6-sign depth

By Theorem 3.8, p. 22, we obtain for $K = 6$:

$$N\left(d_6(\mathbf{R}) - \frac{1}{32}\right) = \frac{N}{32\binom{N}{6}}\Bigg(\sum_{1\le n_1<\ldots<n_6\le N}\Bigg(\sum_{1\le i(1)<i(2)\le 6}\prod_{j=1}^{2}(-1)^{i(j)}\psi(R_{n_j}) + \sum_{1\le i(1)<\ldots<i(4)\le 6}\prod_{j=1}^{4}(-1)^{i(j)}\psi(R_{n_{i(j)}}) \tag{54}$$

$$- \prod_{j=1}^{6}\psi(R_{n_j})\Bigg)\Bigg). \tag{55}$$

We have to simplify the second and third part of the sum in Formula (54) and (55). The term in Formula (55) can be computed by Lemma 4.2 in linear time since this lemma has been proved for general product lengths. However, the products of length 4 in Formula (54) are more challenging to compute. At first, we reduce the number of summation indices to four instead of six with an appropriate compensation factor for the omitted combinations since each summand only depends on four indices.

$\psi(R_{n_1})$	$\psi(R_{n_2})$	$\psi(R_{n_3})$	$\psi(R_{n_4})$	$\psi(R_{n_5})$	$\psi(R_{n_6})$	**sign**	**omitted combinations**
✓	✓	✓	✓			+	$\binom{N-n_4}{2}$
✓	✓	✓		✓		−	$\binom{N-n_5}{1}\binom{n_5-n_3-1}{1}$
✓	✓	✓			✓	+	$\binom{n_6-n_3-1}{2}$
✓	✓		✓	✓		+	$\binom{N-n_5}{1}\binom{n_4-n_2-1}{1}$
✓	✓		✓		✓	−	$\binom{n_6-n_4-1}{1}\binom{n_4-n_2-1}{1}$
✓	✓			✓	✓	+	$\binom{n_5-n_2-1}{2}$
✓		✓	✓	✓		−	$\binom{N-n_5}{1}\binom{n_3-n_1-1}{1}$
✓		✓	✓		✓	+	$\binom{n_6-n_4-1}{1}\binom{n_3-n_1-1}{1}$
✓		✓		✓	✓	−	$\binom{n_5-n_3-1}{1}\binom{n_3-n_1-1}{1}$
✓			✓	✓	✓	+	$\binom{n_4-n_1-1}{2}$
	✓	✓	✓	✓		+	$\binom{N-n_5}{1}\binom{n_2-1}{1}$
	✓	✓	✓		✓	−	$\binom{n_6-n_4-1}{1}\binom{n_2-1}{1}$
	✓	✓		✓	✓	+	$\binom{n_5-n_3-1}{1}\binom{n_2-1}{1}$
	✓		✓	✓	✓	−	$\binom{n_4-n_2-1}{1}\binom{n_2-1}{1}$
		✓	✓	✓	✓	+	$\binom{n_3-1}{2}$

Table 2: Omitted combinations of the products of length four for 6-sign depth

Table 2 gives an overview of all possible situations. If we simplify the binomial coefficients of the last column and order them, Formula (54) can be represented by:

$$\begin{aligned}
&\sum_{1\le n_1<\ldots<n_6\le N}\ \sum_{1\le i(1)<\ldots<i(4)\le 6}\ \prod_{j=1}^{4}(-1)^{i(j)}\psi(R_{n_{i(j)}})\\
&=\left(\frac{1}{2}\left(N^2-3N\right)+2\right)\sum_{1\le n_1<\ldots<n_4\le N}\ \prod_{j=1}^{4}\psi(R_{n_j})\\
&\quad-2N\sum_{L=1}^{4}(-1)^L\sum_{1\le n_1<\ldots n_4\le N}\ n_L\prod_{j=1}^{4}\psi(R_{n_j})\\
&+4\sum_{1\le i(1)<i(2)\le 4}(-1)^{i(1)+i(2)}\sum_{1\le n_1<\ldots<n_4\le N}\ n_{i(1)}n_{i(2)}\prod_{j=1}^{4}\psi(R_{n_j}) \qquad (56)\\
&\quad+2\sum_{L=1}^{4}\ \sum_{1\le n_1<\ldots<n_4\le N}\ n_L^2\prod_{j=1}^{4}\psi(R_{n_j}). \qquad (57)
\end{aligned}$$

It is recommended to expand this by computational algebraic tools as `Mathematica` (Wolfram Research, 2019). Formula (57) can be computed by Lemma 4.3 in linear time by choosing $g(x) = x^2$. Beside terms of the form given in Lemma 4.2 and Lemma 4.3, there are sums with two factors $n_{i(1)}n_{i(2)}$, cf. Formula (56), for which we do not have provided efficient computational tools currently. In the following theorem, we will generalize Lemma 4.3 for multiple weighting factors. Before that, we want to insert the final representation of the 6-depth which can be computed as efficiently as possible with our tools:

$$\begin{aligned}
N\left(d_6(\mathbf{R})-\frac{1}{32}\right)=\frac{N}{32\binom{N}{6}}\Bigg(&\sum_{1\le n_1<\ldots<n_6\le N}\ \sum_{1\le i(1)<i(2)\le 6}\ \prod_{j=1}^{2}(-1)^{i(j)}\psi(R_{n_j})\\
&+\left(\frac{1}{2}\left(N^2-3N\right)+2\right)\sum_{1\le n_1<\ldots<n_4\le N}\ \prod_{j=1}^{4}\psi(R_{n_j})\\
&-2N\sum_{L=1}^{4}(-1)^L\sum_{1\le n_1<\ldots n_4\le N}\ n_L\prod_{j=1}^{4}\psi(R_{n_j})\\
&+4\sum_{1\le i(1)<i(2)\le 4}(-1)^{i(1)+i(2)}\sum_{1\le n_1<\ldots<n_4\le N}\ n_{i(1)}n_{i(2)}\prod_{j=1}^{4}\psi(R_{n_j})\\
&+2\sum_{L=1}^{4}\ \sum_{1\le n_1<\ldots<n_4\le N}\ n_L^2\prod_{j=1}^{4}\psi(R_{n_j})\\
&-\sum_{1\le n_1<\ldots<n_6\le N}\ \prod_{j=1}^{6}\psi(R_{n_j})\Bigg).
\end{aligned}$$

Theorem 4.4. *Let* $(x_1, \dots, x_N)^\top$ *be real numbers with* $x_n \neq 0$, $n = 1, \dots, N$ *and* $\mathbf{L} = (L_1, \dots, L_M)^\top \in \{1, \dots, K\}^M$ *with* $L_{i(1)} < L_{i(2)}$ *for* $i(1) < i(2)$. *Furthermore, we introduce the following notations:*

$$\mathcal{J}(\mathbf{L}) = \{(j(1), \dots, j(M))^\top \in \{1, \dots, N\}^M; L_1 \leq j(1),$$
$$j(m)+(L_{m+1} - L_m - 1) < j(m+1) \text{ for } m \leq M-1, j(M) \leq N - K + L_M\}$$
$$\text{and } \mathcal{M}_{m-1}^m(\mathbf{j}) := \sum_{i=j(m-1)+1}^{j(m)-1} \mathbb{1}\{x_{n_i} < 0\},\ m \in \{1, \dots, M+1\}$$

and $\mathbf{j} = (j(1), \dots, j(M))^\top \in \mathcal{J}(\mathbf{L})$ *and* $j(0) := 0$, $j(M+1) := N+1$. *Then for arbitrary mappings* $g : \mathbb{R}^M \to \mathbb{R}$, *it holds*

$$\sum_{1 \leq n_1 < \dots < n_K \leq N} g(n_{L_1}, \dots, n_{L_M}) \prod_{k=1}^{K} \psi(x_{n_k})$$
$$= \sum_{\substack{\mathbf{j} \in \mathcal{J}(\mathbf{L}) \\ \mathbf{j} = (j(1), \dots, j(M))^\top}} g(\mathbf{j}) \prod_{m=1}^{M} \left(\psi(x_{j(m)}) \sum_{k_m=0}^{L_m - L_{m-1} - 1} (-1)^{k_m} \kappa_{k_m}^{\mathbf{j}} \right) \left(\sum_{k_{M+1}=0}^{K - L_M} (-1)^{k_{M+1}} \kappa_{k_{M+1}}^{\mathbf{j}} \right)$$
$$\text{with } \kappa_{k_m}^{\mathbf{j}}(m) = \kappa_{k_m}^{\mathbf{j}} = \binom{\mathcal{M}_{m-1}^m(\mathbf{j})}{k_m} \binom{j(m) - j(m-1) - 1 - \mathcal{M}_{m-1}^m(\mathbf{j})}{L_m - L_{m-1} - 1 - k_m}$$
$$\text{and } L_0 = 0, L_{M+1} = K+1.$$

The expression $\mathcal{M}_{m-1}^m(\mathbf{j})$ describes the number of negative signs of the residuals with indices between $j(m-1)$ and $j(m)$. By computing and memorizing $\mathcal{M}(n) := \sum_{i=1}^{n} \mathbb{1}\{x_{n_i} < 0\}$ for $n = 1, \dots, N$, we obtain all $\mathcal{M}_{m-1}^m(\mathbf{j})$ by using

$$\mathcal{M}_{m-1}^m(\mathbf{j}) = \mathcal{M}(j(m)) - \mathcal{M}(j(m-1)).$$

Note that the expressions in Theorem 4.4 can be computed in polynomial time with degree M since an ordered sum with M indices has to be computed. Therefore, the expressions of the 6-depth with mixed factors can be computed in quadratic time by using Theorem 4.4. For simplicity, we will consider an example with $M = 4$ and $\mathbf{L} = (2, 4)^\top$. Then, we can compute the following expression in quadratic time:

$$\sum_{1 \leq n_1 < \dots < n_4 \leq N} n_2 n_4 \prod_{k=1}^{4} \psi(x_{n_k})$$
$$= \sum_{\substack{2 \leq j(1), j(2) \leq N \\ j(1)+1 < j(2)}} j(1) j(2) \prod_{m=1}^{2} \psi(x_{j(m)}) \left(\sum_{k_1=0}^{1} \sum_{k_2=0}^{1} \sum_{k_3=0}^{0} (-1)^{k_1+k_2+k_3} \kappa_{k_1}^{\mathbf{j}} \kappa_{k_2}^{\mathbf{j}} \kappa_{k_3}^{\mathbf{j}} \right)$$

$$= \sum_{\substack{2\le j(1),j(2)\le N\\ j(1)+1<j(2)}} j(1)j(2)\prod_{m=1}^{2}\psi(x_{j(m)})\left(\sum_{k_1=0}^{1}\sum_{k_2=0}^{1}(-1)^{k_1+k_2}\kappa_{k_1}^{\mathbf{j}}\kappa_{k_2}^{\mathbf{j}}\right).$$

Proof of Theorem 4.4: Note that $(n_{L_1},\dots,n_{L_M})^\top$ have values equal to $\mathbf{j}\in\mathcal{J}(\mathbf{L})$ so that we can filter out this vector by a sum over $\mathcal{J}(\mathbf{L})$. Further, we introduce the notation

$$\mathcal{B}_{\mathbf{L}}^{\mathbf{j}} = \{(n_i)_{i\in\{1,\dots,K\}\setminus\{L_1,\dots,L_M\}}; (n_1,\dots,n_K)^\top\in\{1,\dots,N\}^K \\ \text{with } n_{i(1)}<n_{i(2)} \text{ for all } i(1)<i(2) \text{ and } (n_{L_1},\dots,n_{L_M})^\top=\mathbf{j}\}.$$

and for $\mathbf{k}=(k_1,\dots,k_{M+1})^\top\in\mathbb{N}^{M+1}$,

$$\mathcal{B}_{\mathbf{L}}^{\mathbf{j}}(\mathbf{k}) = \left\{(n_i)\in\mathcal{B}_{\mathbf{L}}^{\mathbf{j}}; \sum_{i=L_{m-1}+1}^{L_m-1}\mathbb{1}\{x_{n_i}<0\}=k_m,\ m=1,\dots,M+1\right\}.$$

Then

$$\begin{aligned}
&\sum_{1\le n_1<\dots<n_K\le N} g(n_{L_1},\dots,n_{L_M})\prod_{k=1}^{K}\psi(x_{n_k})\\
&= \sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\dots,j(M))^\top}} g(\mathbf{j})\prod_{m=1}^{M}\psi(x_{j(m)})\sum_{(n_i)\in\mathcal{B}_{\mathbf{L}}^{\mathbf{j}}}\prod_{\substack{i=1\\ i\notin\{j(1),\dots,j(M)\}}}^{K}\psi(x_{n_i})\\
&= \sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\dots,j(M))^\top}} g(\mathbf{j})\prod_{m=1}^{M}\psi(x_{j(m)})\sum_{k_1=0}^{L_1-1}\sum_{k_2=0}^{L_2-L_1-1}\cdots\sum_{k_{M+1}=0}^{K-L_M}\sum_{(n_i)\in\mathcal{B}_{\mathbf{L}}^{\mathbf{j}}(\mathbf{k})}\prod_{\substack{i=1\\ i\notin\{j(1),\dots,j(M)\}}}^{K}\psi(x_{n_i})\\
&= \sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\dots,j(M))^\top}} g(\mathbf{j})\prod_{m=1}^{M}\psi(x_{j(m)})\sum_{k_1=0}^{L_1-1}\sum_{k_2=0}^{L_2-L_1-1}\cdots\sum_{k_{M+1}=0}^{K-L_M}\sum_{(n_i)\in\mathcal{B}_{\mathbf{L}}^{\mathbf{j}}(\mathbf{k})}\prod_{i=1}^{M+1}(-1)^{k_i}\\
&= \sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\dots,j(M))^\top}} g(\mathbf{j})\prod_{m=1}^{M}\psi(x_{j(m)})\sum_{k_1=0}^{L_1-1}\sum_{k_2=0}^{L_2-L_1-1}\cdots\sum_{k_{M+1}=0}^{K-L_M}\prod_{i=1}^{M+1}(-1)^{k_i}\times\\
&\prod_{m=1}^{M+1}\binom{\mathcal{M}_{m-1}^m(\mathbf{j})}{k_m}\binom{j(m)-j(m-1)-1-\mathcal{M}_{m-1}^m(\mathbf{j})}{L_m-L_{m-1}-1-k_m}
\end{aligned}$$

where $L_0=0$ and $L_{M+1}=K+1$. The $M+1$ sums can be reordered since they do not depend on each other. Moreover, we use the notation $\kappa_{k_m}^{\mathbf{j}}$ given in the assumptions

of Theorem 4.4:

$$\sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\ldots,j(M))^\top}} g(\mathbf{j}) \prod_{m=1}^{M} \psi(x_{j(m)}) \sum_{k_1=0}^{L_1-1} \sum_{k_2=0}^{L_2-L_1-1} \cdots \sum_{k_{M+1}=0}^{K-L_M} \prod_{i=1}^{M+1} (-1)^{k_i} \times$$
$$\prod_{m=1}^{M+1} \binom{\mathcal{M}_{m-1}^m(\mathbf{j})}{k_m} \binom{j(m)-j(m-1)-1-\mathcal{M}_{m-1}^m(\mathbf{j})}{L_m-L_{m-1}-1-k_m}$$
$$= \sum_{\substack{\mathbf{j}\in\mathcal{J}(\mathbf{L})\\ \mathbf{j}=(j(1),\ldots,j(M))^\top}} g(\mathbf{j}) \left(\prod_{m=1}^{M} \psi(x_{j(m)}) \sum_{k_m=0}^{L_m-L_{m-1}-1} (-1)^{k_m} \kappa_{k_m}^{\mathbf{j}} \right) \left(\sum_{k_{M+1}=0}^{K-L_M} (-1)^{k_{M+1}} \kappa_{k_{M+1}}^{\mathbf{j}} \right)$$

and the assertion follows. □

The recent results in Chapter 4.2.1 show that we are able to compute the product parts for $K = 4$ and $K = 5$ in linear time. For $K \geq 6$, the time complexity will increase undesirably.

Algorithms for higher K

To sum up, an explicit representation of the product formula of the K-depth for $K \in \{4, 5, 6\}$ has been derived which can be computed more efficiently than the naive implementation. If we consider K-depths for higher K, we will obtain products with even lengths up to K (e.g., products of length eight for the 8-depth). If the product length is two, we can apply the algorithm in Chapter 4.1 which has linear time. The product length of K or $K-1$ (depending on whether K is odd or even) can also be computed in linear time by the algorithms in Chapter 4.2.1. For product lengths higher than two and less than $K-1$, the computational situation is difficult as the following discussion shows.
Let $\tilde{K} \in \{4, \ldots, K\}$ for even K or $\tilde{K} \in \{4, \ldots, K-1\}$ for odd K be the product length. Analogously to the previous derivations from this chapter, the expression

$$\sum_{1\leq n_1<\ldots<n_K\leq N} \prod_{j=1}^{\tilde{K}} \psi(R_{n_{i(j)}})$$

leads for particular $1 \leq i(1) < \ldots < i(\tilde{K}) \leq K$ to terms of the form

$$\sum_{1\leq n_1<\ldots<n_{\tilde{K}}\leq N} g(n_{L_1}, \ldots, n_{L_{\min\{\tilde{K},K-\tilde{K}\}}}) \prod_{j=1}^{\tilde{K}} \psi(R_{n_j}) \tag{58}$$

for suitable $1 \leq L_1 < \ldots < L_{\min\{\tilde{K},K-\tilde{K}\}} \leq \tilde{K}$ and a suitable function g. We obtain

Formula (58) by reducing the number of indices of the sum from K to $\tilde{K}$ and need $K - \tilde{K}$ factors for compensation. Note that the maximal number of compensating factors is $\tilde{K}$ since the sum only depend on $\tilde{K}$ indices so that we take the minimum of $\tilde{K}$ and $K - \tilde{K}$ here. According to Theorem 4.4, Formula (58) can be computed with time complexity of $\Theta(N^{\max\{\min\{\tilde{K}, K-\tilde{K}\},1\}})$. Note that $\min\{\tilde{K}, K - \tilde{K}\}$ can be zero if $K = \tilde{K}$. Since we always need to compute the sign structure in advance in general, we always have linear time complexity at least.
The time complexity gets worse the higher the distance between the numbers K and $\tilde{K}$ is. Moreover, the derivation of the corresponding representations gets more tedious. E.g. for $K = 7$, $\binom{7}{4} = 35$ and $\binom{7}{6} = 7$ terms need to be calculated in order to obtain a formula that can be computed by the methods of Chapter 4.2.1. Algebraic computational tools for expanding polynomials as `Mathematica` (Wolfram Research, 2019) exist and can be used for more comfort. However, for general $K \geq 7$, we obtain exploding numbers of combinations to derive a more useful representation.
To sum up, this approach is not appropriate for higher K due to the rising time complexity and the immense effort to derive more efficient representations. Table 3 gives an overview of the behavior of the product terms for $K \in \{4, \ldots, 12\}$. We can

K	**product lengths $\tilde{K}$**	**respective runtime (referring to Chapter 4.2.1)**
4	4	$\Theta(N)$
5	4	$\Theta(N)$
6	4,6	$\Theta(N^2), \Theta(N)$
7	4,6	$\Theta(N^3), \Theta(N)$
8	4,6,8	$\Theta(N^4), \Theta(N^2), \Theta(N)$
9	4,6,8	$\Theta(N^4), \Theta(N^3), \Theta(N)$
10	4,6,8,10	$\Theta(N^4), \Theta(N^4), \Theta(N^2), \Theta(N)$
11	4,6,8,10	$\Theta(N^4), \Theta(N^5), \Theta(N^3), \Theta(N)$
12	4,6,8,10,12	$\Theta(N^4), \Theta(N^6), \Theta(N^4), \Theta(N^2), \Theta(N)$

Table 3: Overlook of the runtimes for various K to compute different product lengths

conclude that the computational effort of all product terms is $\mathcal{O}(N^{K^*})$ with

$$K^* = \max_{\substack{4 \leq \tilde{K} \leq K \\ \tilde{K} \in 2\mathbb{N}}} \{1, \min\{\tilde{K}, K - \tilde{K}\}\}$$

in worst case for given K, when using the given tools in Chapter 4.2.1. There may exist other approaches to simplify these products terms similar to the derivations from Chapter 3 which we have not found.

4.2.2 Linear computation of the other rest terms

The other neglected part of the K-depth is given in Lemma 3.15, p. 31:

$$
\begin{aligned}
&-\frac{N}{2^K\binom{N}{K}}\sum_{1\le n_1\neq n_2\le N}\sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}\\
=&-\frac{N^{K-1}K(K-1)}{4\langle N\rangle_K}\sum_{1\le n_1\neq n_2\le N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2} \qquad (59)\\
&-\frac{N^{K-1}}{2^K\binom{N}{K}}\sum_{1\le n_1\neq n_2\le N}\sum_{0\le M_1<M_2\le K-2}a_K(M_1,M_2)\frac{|n_1-n_2|^{M_1}}{N^{M_2}} \qquad (60)
\end{aligned}
$$

Chapter 4.1 provides an algorithm in linear time for Formula (59). For Formula (60), we will also obtain a linear algorithm in the following. The coefficients $a_K(M_1, M_2)$ do not depend on the given residuals and only need to be calculated in advance once theoretically to obtain an exact representation for the K-depth.

K = 3
$(0,1)$

K = 4	
$(0,1)$	$(1,2)$
$(0,2)$	

K = 5		
$(0,1)$	$(1,2)$	$(2,3)$
$(0,2)$	$(1,3)$	
$(0,3)$		

K = 6			
$(0,1)$	$(1,2)$	$(2,3)$	$(3,4)$
$(0,2)$	$(1,3)$	$(2,4)$	
$(0,3)$	$(1,4)$		
$(0,4)$			

Figure 9: Situations for (M_1, M_2) in Formula (60)

Figure 9 shows for $K = 3, 4, 5, 6$ the possible situations of $0 \le M_1 < M_2 \le K-2$ we have to compute. This can be simply done by expanding the sum

$$
f_K^N(a) := \frac{1}{N^{K-2}}\sum_{J=0}^{K-2}(-1)^J\binom{a-1}{J}\binom{N-a-1}{K-2-J} \qquad (61)
$$

with $a := |n_1 - n_2|$ and by ordering the coefficients in a and N such that:

$$
f_K^N(a) = \sum_{0\le M_1\le M_2\le K-2}a_K(M_1,M_2)\frac{a^{M_1}}{N^{M_2}}
$$

for suitable $a_K(M_1, M_2) \in \mathbb{R}$. The coefficients $\frac{a^{M_1}}{N^{M_2}}$ with $M_1 = M_2$ correspond to the asymptotically relevant part in Formula (59) and are not mentioned in Figure 9. It is recommended to use a computational algebraic system as `Mathematica` (Wolfram Research, 2019) to expand the terms.

3-depth

For $K = 3$, we obtain in Formula (61)

$$f_3^N(a) = \frac{1}{N}\sum_{J=0}^{1}(-1)^J\binom{a-1}{J}\binom{N-a-1}{1-J} = 1 - 2\frac{a}{N}.$$

No factor of the form $\frac{a^0}{N}$ appears and $a_3(0,1) = 0$ therefore.

4-depth

For $K = 4$, we obtain in Formula (61)

$$f_4^N(a) = \frac{1}{N^2}\sum_{J=0}^{2}(-1)^J\binom{a-1}{J}\binom{N-a-1}{2-J} = 2\frac{a^2}{N^2} - 2\frac{a}{N} + \frac{1}{N^2} - \frac{1}{2}\frac{1}{N} + \frac{1}{2}.$$

We obtain $a_4(0,1) = -\frac{1}{2}$, $a_4(0,2) = 1$ and $a_4(1,2) = 0$.

5-depth

For $K = 5$, we obtain in Formula (61)

$$\begin{aligned} f_5^N(a) =& \frac{1}{N^3}\sum_{J=0}^{3}(-1)^J\binom{a-1}{J}\binom{N-a-1}{3-J} \\ =& -\frac{3}{4}\frac{a^3}{N^3} + 2\frac{a^2}{N^2} - \frac{a}{N} + \frac{a}{N^2} - \frac{8}{3}\frac{a}{N^3} + \frac{1}{6} - \frac{1}{2}\frac{1}{N} + \frac{4}{3}\frac{1}{N^2}. \end{aligned}$$

We obtain $a_5(0,1) = -\frac{1}{2}$, $a_5(0,2) = \frac{4}{3}$, $a_5(0,3) = 0$, $a_5(1,2) = 1$, $a_5(1,3) = -\frac{8}{3}$ and $a_5(2,3) = 0$.

6-depth

For $K = 6$, we obtain in Formula (61)

$$\begin{aligned} f_6^N(a) =& \frac{1}{N^4}\sum_{J=0}^{4}(-1)^J\binom{a-1}{J}\binom{N-a-1}{4-J} \\ =& \frac{2}{3}\frac{a^4}{N^4} - \frac{4}{3}\frac{a^3}{N^3} + \frac{10}{3}\frac{a^2}{N^4} - \frac{a^2}{N^3} + \frac{a^2}{N^2} - \frac{10}{3}\frac{a}{N^3} \\ &+ \frac{a}{N^2} - \frac{1}{3}\frac{a}{N} + \frac{1}{N^4} - \frac{3}{4}\frac{1}{N^3} + \frac{23}{24}\frac{1}{N^2} - \frac{1}{4}\frac{1}{N} + \frac{1}{24}. \end{aligned}$$

Thus, we obtain $a_6(0,1) = -\frac{1}{4}$, $a_6(0,2) = \frac{23}{24}$, $a_6(0,3) = -\frac{3}{4}$, $a_6(0,4) = 1$, $a_6(1,2) = 1$, $a_6(1,3) = -\frac{10}{3}$, $a_6(1,4) = 0$, $a_6(2,3) = -1$, $a_6(2,4) = \frac{10}{3}$ and $a_6(3,4) = 0$.

$\mathbf{K=3}$
0

$\mathbf{K=4}$	
$-1/2$	0
1	

$\mathbf{K=5}$		
$-1/2$	1	0
$4/3$	$-8/3$	
0		

$\mathbf{K=6}$			
$-1/4$	1	-1	0
$23/24$	$-10/3$	$10/3$	
$-3/4$	0		
1			

Figure 10: Values for $a_K(M_1, M_2)$ for $K \in \{3, 4, 5, 6\}$ and $0 \leq M_1 < M_2 \leq K-2$

Figure 10 summarizes the values for the coefficients $a_K(M_1, M_2)$. The rest sums can be computed similarly to the approximate computation of the K-depth in Lemma 4.1, p. 56. According to Theorem 3.12, p. 28 and Lemma 3.15, p. 31, we have to derive a simplifcation of the following formula:

$$
\begin{aligned}
&-\frac{N^{K-1}}{2^K\binom{N}{K}} \sum_{1\leq n_1\neq n_2\leq N} \left(\sum_{0\leq M_1<M_2\leq K-2} a_K(M_1,M_2) \frac{|n_1-n_2|^{M_1}}{N^{M_2}} \psi(E_{n_1})\psi(E_{n_2}) \right) \\
=&-\frac{N^{K-1}}{2^K\binom{N}{K}} \sum_{0\leq M_1<M_2\leq K-2} a_K(M_1,M_2) \left(\sum_{1\leq n_1\neq n_2\leq N} \frac{|n_1-n_2|^{M_1}}{N^{M_2}} \psi(E_{n_1})\psi(E_{n_2}) \right) \\
=&-\frac{N^{K-1}}{2^K\binom{N}{K}} \sum_{0\leq M_1<M_2\leq K-2} \frac{a_K(M_1,M_2)}{N^{M_2-M_1}} \left(\sum_{1\leq n_1\neq n_2\leq N} \left(\frac{|n_1-n_2|}{N}\right)^{M_1} \psi(E_{n_1})\psi(E_{n_2}) \right).
\end{aligned}
$$

Then analogous to the proof of Lemma 4.1, we obtain

$$
\begin{aligned}
&\sum_{1\leq n_1\neq n_2\leq N} \left(\frac{|n_1-n_2|}{N}\right)^{M_1} \psi(E_{n_1})\psi(E_{n_2}) = 2\sum_{n_1=1}^{N-1}\sum_{n_2=n_1+1}^{N} \left(\frac{n_2-n_1}{N}\right)^{M_1} \psi(E_{n_1})\psi(E_{n_2}) \\
=&2\sum_{m=0}^{M_1}\binom{M_1}{m}(-1)^m \sum_{n_2=2}^{N} \left(\frac{-n_2}{N}\right)^{M_1-m} \psi(E_{n_2}) \sum_{n_1=1}^{n_2-1} \left(\frac{n_1}{N}\right)^m \psi(E_{n_1}) \\
=&2\sum_{m=0}^{M_1}\binom{M_1}{m}(-1)^m \sum_{n_2=2}^{N} \left(\frac{-n_2}{N}\right)^{M_1-m} \psi(E_{n_2}) S_{n_2-1,m}.
\end{aligned}
$$

During the computation of the asymptotic part in Lemma 4.1, p. 56, the matrix $(S_{\bullet,0}, \ldots, S_{\bullet,M_1})$ is already computed since $M_1 < K-2$. In order to avoid multiple computations, this stored matrix should be used. The following equation summarizes the efficient representation for Formula (60):

$$
\begin{aligned}
&-\frac{N^{K-1}}{2^K\binom{N}{K}} \sum_{1\leq n_1\neq n_2\leq N} \left(\sum_{0\leq M_1<M_2\leq K-2} a_K(M_1,M_2) \frac{|n_1-n_2|^{M_1}}{N^{M_2}} \psi(E_{n_1})\psi(E_{n_2}) \right) \\
=&-\frac{N^{K-1}}{2^{K-1}\binom{N}{K}} \sum_{0\leq M_1<M_2\leq K-2} \frac{a_K(M_1,M_2)}{N^{M_2-M-1}} \times \\
&\qquad \left(\sum_{m=0}^{M_1}\binom{M_1}{m}(-1)^m \sum_{n_2=2}^{N} \left(\frac{-n_2}{N}\right)^{M_1-m} \psi(E_{n_2}) S_{n_2-1,m} \right).
\end{aligned}
$$

Although this representation can be computed in linear time, the computational costs can be high due to the large number of summands for increasing K.

4.3 Runtime comparison

The implementations introduced in Chapter 4.1 and 4.2 will be compared based on their runtimes. The asymptotic implementations in Chapter 4.1 are obviously faster than the exact implementations in Chapter 4.2 since the exact implementations have to compute additional terms. The interesting question is how much faster the asymptotic implementations are compared to the exact implementations for different K. Additionally, we compare the naive implementation based on the definition to underline the requirement of the improved implementations. The runtimes will be visualized for different sample sizes N by boxplots that contain 100 repeated runtimes given in milliseconds (ms). The default settings of the R-function `boxplot()` are used with `outline = FALSE` in order to omit outliers for clearer presentation. If the algorithms are linear, we consider samples sizes $N \in \{100, 200, \ldots, 10000\}$ and for the implementation with quadratic or higher complexities, we consider $N \in \{10, 20, \ldots, 300\}$. For the naive implementation of the 5-depth and 6-depth, we stopped the computations at $N = 200$ or $N = 150$, respectively, due to long computation times.

3-depth

Figure 11 presents the runtimes of the (exact) efficient implementation in Chapter 4.1 and the naive implementation based on the definition. Further, a simple linear re-

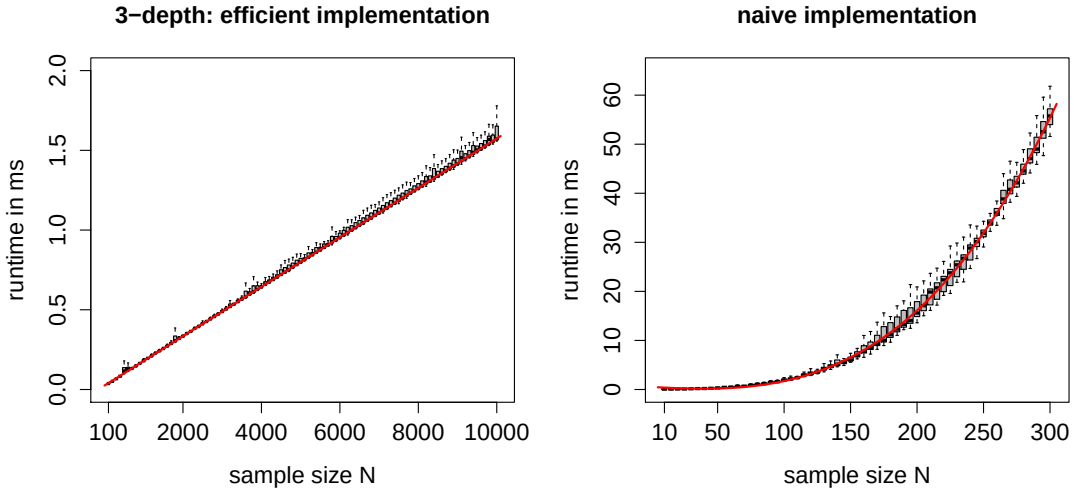

Figure 11: Simulated runtimes of the efficient implementation (cf. Chapter 4.1) and the naive implementation.

gression line for the efficient implementation and a cubic regression line for the naive implementation are calculated by the medians from each boxplot based on their theoretical time complexities of $\Theta(N)$ and $\Theta(N^3)$. The difference of the runtimes from the naive implementation to efficient implementation is so immense that plotting them for the same sample sizes N does not make sense. We can conclude that there is no reason to use the naive implementation.

4-depth

Figure 12 presents the runtimes of the efficient implementations in Chapter 4.1 and 4.2 and the naive implementation. It also contains the simple linear regression line or a line of a polynomial regression from degree four, respectively. We conclude

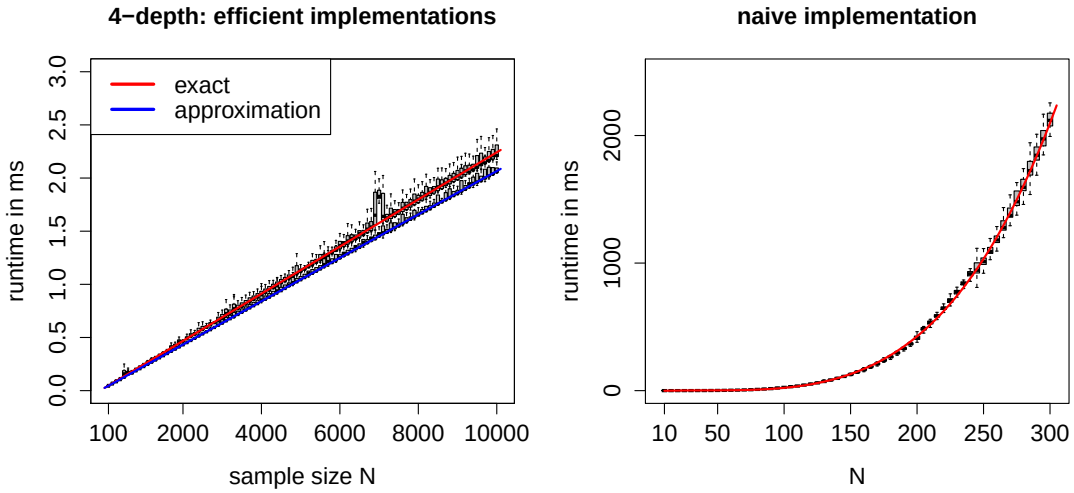

Figure 12: Simulated runtimes of both efficient implementations (cf. Chapter 4.1 and 4.2) and the naive implementation

that the differences between the runtime of the exact and approximate version are negligibly small. For $N = 10000$, both algorithms take about 2ms. This can be explained by the minor differences of both implementations since the additional rest term in Chapter 4.2 can be computed in constant time. We only need to compute the number of negative signs in linear time in advance, see Lemma 4.2, p. 58. The runtimes of the naive implementation show that the higher K, the more efficient algorithms are needed.

5-depth

Figure 13 is analogously set up to Figure 12 for $K = 5$. In contrast, we have here a much higher difference between the exact and approximate implementation.

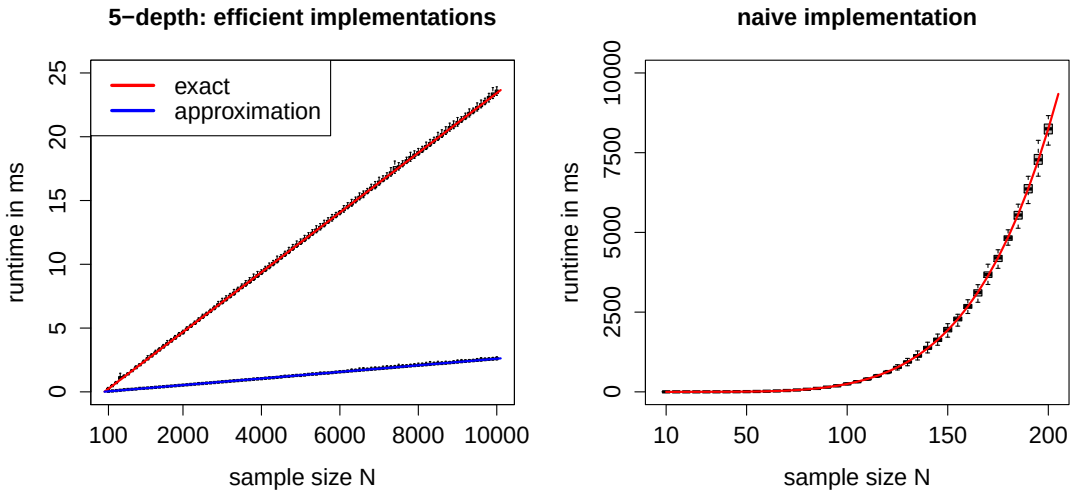

Figure 13: Simulated runtimes of both efficient implementations (cf. Chapter 4.1 and 4.2) and the naive implementation

For $N = 10000$, the runtimes are above 25ms and 3ms, respectively. For larger simulation studies or situations where the 5-depth has to be computed for multiple parameters, the approximate versions should be preferred due to the computational costs. Nevertheless, the runtime of the exact implementation also increases linearly. The runtimes of the naive implementation is only computed until $N = 200$ since the runtimes increase more and more. The differences between the exact and approximate implementation of the 5-depth can be explained by the large number of summations in Lemma 4.3, p. 60 and from the the resulting formula in Chapter 4.2.

6-depth and more

In Figure 14, the runtimes of the other algorithms based on the 6-depths are presented. The exact algorithm has a quadratic time complexity (cf. Chapter 4.2), so the exact computation is difficult for higher sample size N. Comparing to the naive implementation, we have a large improvement nevertheless. Figure 15 is a summary of the runtimes of the approximate algorithms for $K \in \{3, 4, 5, 6\}$. We can conclude that choosing higher K will not cause computational problems as we had with with the naive implementation. For $N = 10000$, we need approximately 0,5ms more time per K. The result in Chapter 4.1 is a strong improvement for further research since simulation studies can now be done efficiently, in contrast to past papers, see Kustosz and Müller (2014), Kustosz et al. (2016a), Kustosz et al. (2016b) or Falkenau (2016), where naive implementations have been used. These papers only consider $K = 3$, probably due to the increasing computational complexity for larger K.

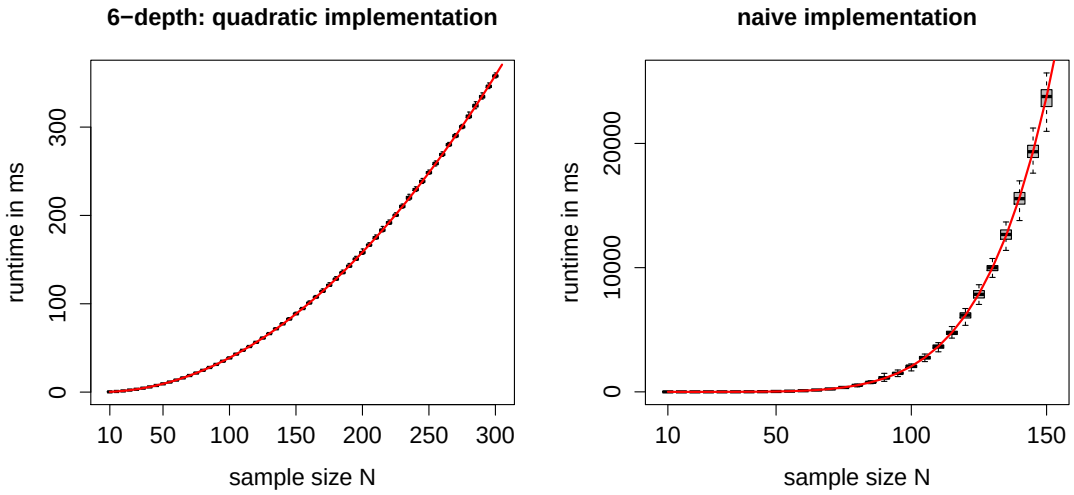

Figure 14: Simulated runtimes of the (exact) quadratic implementation (cf. Chapter 4.2) and the naive implementation for $K = 6$

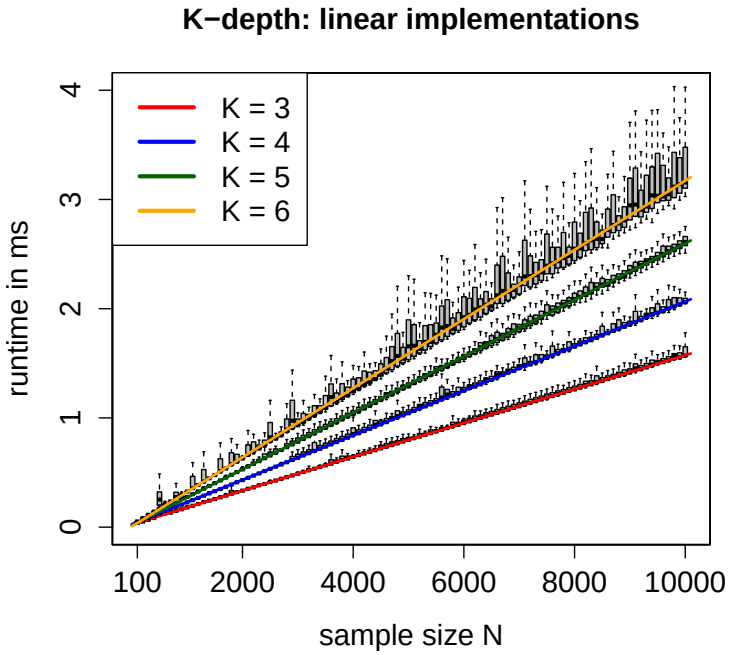

Figure 15: Simulated runtimes of the linear implementations (cf. Chapter 4.1) for $K \in \{3, 4, 5, 6\}$

Based on the approaches of Chapter 4, we can compute the K-depth efficiently for all K if an approximation is sufficient. For the exact K-depth, we have only found efficient computational tools for $K \leq 5$.

5 Efficient computation of the K-sign depth based on block structures

This chapter introduces another efficient implementation of the K-depth based on the definition. As we have seen in Chapter 4, a naive implementation based on the definition provides an algorithm with undesirable time complexity of $\Theta(N^K)$ since the alternation property of the signs is checked for all K-tuples in K nested loops. Some improved approaches of the naive algorithm are considered in Horn (2021b): A first idea to reduce computational costs is to stop checking the condition whether a K-tuple has alternating signs if two consecutive signs of the residual vectors are already equal. Although this method can reduce the computational costs, the time complexity is still of order $\Theta(N^K)$ in best and worst case due to the number of K-tuples we have to check. Horn (2021b) shows that this procedure can be improved by computing the K-tuples recursively and quadratic time complexity can be obtained in the best case. Nevertheless, the worst case still has a time complexity of $\mathcal{O}(N^K)$. In the following, we will take up a similar idea. We extract only K-tuples where alternating signs could be possible in advance and count them. This can be done by analyzing the block structure of the signs of the residual vector. A block is a run of consecutive entries in the residuals with the same signs. Example 5.1 gives an introduction for this approach:

Example 5.1. *Let* $\mathbf{r} = (1, 1, -1, -1, -1, 1, 1, 1, -1, -1, 1, 1, 1)^\top$ *be the signs of a residual vector. Then, there are the following five blocks:*

$$(\underbrace{+,+}_{\text{block 1}},\underbrace{-,-,-}_{\text{block 2}},\underbrace{+,+,+}_{\text{block 3}},\underbrace{-,-}_{\text{block 4}},\underbrace{+,+,+}_{\text{block 5}}).$$

Let $K = 3$ *as an example. A 3-tuple has alternating signs if and only if its entries are from one of the following block index combinations:*

$$123,\ 125,\ 145,\ 234,\ 345.$$

The parities (parity is the property of whether an integer is even or odd) have to alternate, i.e., (even, odd, even) or (odd, even, odd), in order to have alternating signs. In this example, odd block indices correspond to positive signs and even block indices to negative signs. Therefore, the other block combinations cannot have alternating signs. For a given residual vector and the numbers of elements in each block,

we can calculate the number of alternating signs simply and quickly:

$$\begin{aligned} 123: &\ \ 2 \cdot 3 \cdot 3 = 18 \\ 125: &\ \ 2 \cdot 3 \cdot 3 = 18 \\ 145: &\ \ 2 \cdot 2 \cdot 3 = 12 \\ 234: &\ \ 3 \cdot 3 \cdot 2 = 18 \\ 345: &\ \ 3 \cdot 2 \cdot 3 = 18 \end{aligned}$$

$$d_3(\mathbf{r}) = \frac{1}{\binom{13}{3}}(18 + 18 + 12 + 18 + 18) = \frac{84}{286} \approx 0.2937.$$

This approach will be formalized for general K in Chapter 5.1. If $B(\mathbf{r})$ denotes the number of blocks of the residual vector $\mathbf{r}$, then the exact time complexity of this algorithm is $\Theta(B(\mathbf{r})^K)$. However, this procedure is still too naive, since it leads to a time complexity of $\mathcal{O}(N^K)$ in the worst case. Therefore, we will develop improvements of this idea based on some strategically used memory storage in Chapter 5.2 that avoids multiple computation of the same terms. This idea yields an algorithm that can compute the K-depth with linear complexity in the number of blocks $B(\mathbf{r})$. This linear algorithm is implemented by the author of this thesis and included in the `R`-package `GSignTest` Horn (2021a) based on an `Rcpp`-implementation. In Chapter 5.3, the runtimes of the methods in the previous chapters will be compared. In Chapter 5.4, a conjecture concerning a maximality property of the K-depth is discussed. The formulation of this conjecture is based on the block implementation. It shows that further theoretical understanding of the K-depth (e.g., its performance in test scenarios) can be developed by considering the blocks of the residuals.

5.1 Introduction to the block notation and the naive algorithm based on blocks

The terms from Example 5.1 will be formalized by mathematical definitions at first.

Definition 5.2. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top \in \mathbb{R}^N$ *be a residual vector with* $r_n \neq 0$, $n = 1, \ldots, N$.

(a) The ***number of blocks*** $B(\mathbf{r})$ *is defined as*

$$B(\mathbf{r}) := 1 + \sum_{n=2}^{N} \mathbb{1}\{\psi(r_n) \neq \psi(r_{n-1})\} \tag{62}$$

where $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ *defines the sign function.*

(b) The ***starting points*** *of a new block* $s_1(\mathbf{r}), \ldots, s_{B(\mathbf{r})+1}(\mathbf{r})$ *are defined as*

$$s_1(\mathbf{r}) := 1, s_b(\mathbf{r}) := \min\{\ell > s_{b-1}(\mathbf{r}); \psi(\mathbf{r}_\ell) \neq \psi(\mathbf{r}_{\ell-1})\} \text{ for } b = 2, \ldots, B(\mathbf{r}),$$
$$s_{B(\mathbf{r})+1} := N + 1.$$

(c) The ***block lengths*** $q_1(\mathbf{r}), \ldots, q_{B(\mathbf{r})}(\mathbf{r})$ *are defined as*

$$q_b(\mathbf{r}) := s_{b+1}(\mathbf{r}) - s_b(\mathbf{r}) \text{ for } b = 1, \ldots, B(\mathbf{r}). \tag{63}$$

We will also write $B, s_1, \ldots, s_{B+1}$ and $q_1, \ldots, q_B$ instead of $B(\mathbf{r}), s_1(\mathbf{r}), \ldots, s_{B(\mathbf{r})+1}(\mathbf{r})$ and $q_1(\mathbf{r}), \ldots, q_{B(\mathbf{r})}(\mathbf{r})$ for a more convenient notation if the residual vector $\mathbf{r}$ is known from the context. Note that $B, s_1, \ldots, s_{B+1}$ and $q_1 \ldots, q_B$ can be computed in linear time in N. Further, we will use the following notations:

Definition 5.3. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top \in \mathbb{R}^N$ *be a residual vector with* $r_n \neq 0$, $n = 1, \ldots, N$. *Let* B *be the number of blocks and let* s_b *be the starting points from Definition 5.2.*

(a) For an element r_n, $n \in \{1, \ldots, N\}$, *of* $\mathbf{r} \in \mathbb{R}^N$ *and some* $b \in \{1, \ldots, B\}$, *we say that* r_n ***belongs to the block number*** b *if* $s_b \leq n < s_{b+1}$.

(b) For arbitrary $1 \leq i_1 < \ldots < i_K \leq N$, *we say that* $(r_{i_1}, \ldots, r_{i_K})$ ***belongs to the vector of blocks*** $(b_{j_1}, \ldots, b_{j_K})$ *if each* r_{i_k} *belongs to the block number* b_{j_k} *for* $k = 1, \ldots, K$ *and suitable* $1 \leq j_1 \leq \ldots \leq j_K \leq B$.

In order to illustrate the previous definitions, we give the following example.

Example 5.4. *For* $\mathbf{r} = (1, 1, -1, -1, -1, 1, 1, 1, -1, -1, 1, 1, 1)^\top$, *we have*

(a) number of blocks: $B = 5$,
(b) starting points of a new block: $(s_1, \ldots, s_5)^\top = (1, 3, 6, 9, 11)^\top$,
(c) block lengths: $(q_1, \ldots, q_5)^\top = (2, 3, 3, 2, 3)^\top$.

E.g., the fifth entry r_5 *of the residual vector (highlighted in red) belongs to block 2 and* (r_1, r_2, r_9) *(highlighted in blue) belongs to the vector of blocks* $(1, 1, 4)$*:*

$$(\underbrace{+, +}_{\text{block 1}}, \underbrace{-, -, -}_{\text{block 2}}, \underbrace{+, +, +}_{\text{block 3}}, \underbrace{-, -}_{\text{block 4}}, \underbrace{+, +, +}_{\text{block 5}}).$$

The next lemma describes how the block structure can be used to compute the K-depth as mentioned in Example 5.1.

Lemma 5.5. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top \in \mathbb{R}^N$ *be a residual vector with* $r_n \neq 0$, $n = 1, \ldots, N$ *and* $K \geq 2$. *According to Definition 5.2, the number of blocks is denoted by* B *and the block lengths are denoted by* $q_1, \ldots, q_B$. *We define the set of all* K*-tuples containing block indices with alternating parities by*

$$\mathcal{A}_{K,B} := \{(i_1, \ldots, i_K)^\top \in \{1, \ldots, B\}^K; i_k - i_{k-1} \text{ is odd}, i_k - i_{k-1} > 0, k = 2, \ldots, K\}.$$

Then the K*-depth can be computed by*

$$d_K(\mathbf{r}) = \frac{1}{\binom{N}{K}} \sum_{(i_1,\ldots,i_K) \in \mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k}.$$

The elements of $\mathcal{A}_{K,B}$ represent all K-tuples which contain block indices with alternating parities. This follows from the fact that $i_k - i_{k-1}$ is odd which implies that neighbored indices have different parities. Note further that the parity of the block index indicates the sign of its block. If the first entry of the residual has a positive sign, then an odd block index corresponds to a positive sign and an even block index corresponds to a negative sign and vice versa.

Before proving Lemma 5.5, we will highlight the case $B < K$. Then $\mathcal{A}_{K,B} = \emptyset$ so that the sum in Lemma 5.5 is zero:

Lemma 5.6. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top \in \mathbb{R}^N$ *be an arbitrary residual vector with* $r_n \neq 0$, $n = 1, \ldots, N$ *and let* $B(\mathbf{r})$ *the number of blocks. Then*

$$B(\mathbf{r}) \leq K - 1 \quad \Leftrightarrow \quad d_K(\mathbf{r}) = 0.$$

Proof of Lemma 5.6: If there are $K - 1$ blocks or less, it is impossible to choose K-tuples with $K - 1$ sign changes, so the K-depth will be zero. For the reversed implication, suppose $B(\mathbf{r}) \geq K$. Then, we can construct one K-tuple which elements are from the first K different blocks. In particular, this K-tuple has alternating signs which implies $d_K(\mathbf{r}) > 0$. □

Proof of Lemma 5.5: According to Lemma 5.6, we only need to focus on the case $B \geq K$. For $1 \leq n_1 < \ldots < n_K \leq N$, let $(r_{n_1}, \ldots, r_{n_K})^\top$ be an arbitrary K-tuple of the residual vector. Let

$$\mathcal{B}_{K,B} := \{(n_1, \ldots, n_K)^\top \in \{1, \ldots, N\}^K; 1 \leq n_1 < \ldots < n_K \leq N,$$
$$(r_{n_1}, \ldots, r_{n_K})^\top \text{ belongs to the blocks } (b_{j_1}, \ldots, b_{j_K})^\top \text{ with } (j_1, \ldots, j_K)^\top \in \mathcal{A}_{K,B}\}.$$

The set $\mathcal{B}_{K,B}$ only contains K-tuples from the residual vector with alternating signs.

I.e., for a K-tuple $(r_{n_1},\dots,r_{n_K})^\top$ with $(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B}$, we have

$$\prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k < 0\} = 1.$$

For $(r_{n_1},\dots,r_{n_K})^\top$ with $(n_1,\dots,n_K)^\top \notin \mathcal{B}_{K,B}$ and $1 \le n_1 < \ldots < n_K \le N$

$$\prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k < 0\} = 0,$$

since the parities of a block index indicate the sign of the block. Therefore, we would have the same parities for at least two consecutive block indices which yields the same sign for these two consecutive entries and alternating signs would be impossible. Thus, we only need to consider the indices in $\mathcal{B}_{K,B}$ in the definition of the K-depth:

$$\begin{aligned} d_K(\mathbf{r}) &= \frac{1}{\binom{N}{K}} \sum_{1\le n_1<\ldots<n_K\le N} \left(\prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k < 0\} \right) \\ &= \frac{1}{\binom{N}{K}} \sum_{(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B}} \left(\prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{r_{n_k}(-1)^k < 0\} \right) \\ &= \frac{1}{\binom{N}{K}} \sum_{(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B}} 1. \end{aligned}$$

Further, we consider an arbitrary vector $(i_1,\dots,i_K)^\top \in \mathcal{A}_{K,B}$ and corresponding K-tuples $(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B}$ which lead to residual vectors $(r_{n_1},\dots,r_{n_K})^\top$ that belongs to the block vector $(b_{i_1},\dots,b_{i_K})^\top$. By elementary combinatorics, we have

$$\sum_{\substack{(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B} \text{ with} \\ (r_{n_1},\dots,r_{n_K})^\top \text{ belonging to } (b_{i_1},\dots,b_{i_K})^\top}} 1 = \prod_{k=1}^{K} q_{i_k}.$$

Then, we can split the elements in $\mathcal{B}_{K,B}$ as follows:

$$\begin{aligned} d_K(\mathbf{r}) &= \frac{1}{\binom{N}{K}} \sum_{(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B}} 1 \\ &= \frac{1}{\binom{N}{K}} \sum_{(i_1,\dots,i_K)^\top \in \mathcal{A}_{K,B}} \sum_{\substack{(n_1,\dots,n_K)^\top \in \mathcal{B}_{K,B} \text{ with} \\ (r_{n_1},\dots,r_{n_K})^\top \text{ belonging to } (b_{i_1},\dots,b_{i_K})^\top}} 1 \\ &= \frac{1}{\binom{N}{K}} \sum_{(i_1,\dots,i_K)^\top \in \mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k} \end{aligned}$$

and the assertion follows. □

The algorithm based on Lemma 5.5 will be called *naive block-algorithm* in the following since it will be improved immensely, cf. Chapter 5.2. Before that, we discuss the time complexity of it.

Computational complexity of the naive block-algorithm

For the analysis of the computational complexity of the naive block-algorithm, we use the following lemma:

Lemma 5.7. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top \in \mathbb{R}^N$ *be a residual vector with alternating signs, i.e.,* $\psi(r_n) = -\psi(r_{n+1})$ *for* $n = 1, \ldots, N-1$. *Then, for all* $K \leq N$

$$d_K(\mathbf{r}) = \frac{1}{\binom{N}{K}} \left(\binom{\lfloor (N+K)/2 \rfloor}{K} + \binom{\lceil (N+K-2)/2 \rceil}{K} \right).$$

The proof of Lemma 5.7 is given in Leckey et al. (2020). The main idea for this proof is to distinct the two cases whether the first entry of a K-tuple in $\mathcal{A}_{K,N}$ is an even or odd number. The number of K-tuples in $\mathcal{A}_{K,N}$ with an odd first entry is $\binom{\lfloor (N+K)/2 \rfloor}{K}$ and the number of K-tuples with an even first entry is $\binom{\lceil (N+K-2)/2 \rceil}{K}$. This lemma leads to the cardinality of the set $\mathcal{A}_{K,B}$ from Lemma 5.5.

Corollary 5.8. *For* $K \geq 2$

$$|\mathcal{A}_{K,B}| = \binom{\lfloor (B+K)/2 \rfloor}{K} + \binom{\lceil (B+K-2)/2 \rceil}{K}.$$

Proof of Corollary 5.8: Let $\mathbf{r}$ be a residual vector of length B with alternating signs, i.e., $q_1 = \ldots = q_B = 1$. According to Lemma 5.5, we have for the K-depth

$$\binom{B}{K} d_K(\mathbf{r}) = \sum_{(i_1, \ldots, i_K)^\top \in \mathcal{A}_{K,B}} 1 = |\mathcal{A}_{K,B}|$$

and according to Lemma 5.7, we have

$$\binom{B}{K} d_K(\mathbf{r}) = \binom{\lfloor (B+K)/2 \rfloor}{K} + \binom{\lceil (B+K-2)/2 \rceil}{K}.$$

Since the left side of both equations are equal, this implies the assertion. Note that the set $\mathcal{A}_{K,B}$ does not depend on the particular residual vector and its length but only on the block structure. □

If we assume that the block structure is already computed (this has always time complexity $\Theta(N)$), then the algorithm based on Lemma 5.5 has a computational

complexity of $\Theta(B^K)$ since, according to Corollary 5.8, we have to compute the product of block lengths $|\mathcal{A}_{K,B}| = \Theta(B^K)$ times. In the worst case, this is $\mathcal{O}(N^K)$ due to the fact that $B \leq N$.

5.2 Exact computation of the K-depth in linear time based on blocks

The representation of the K-depth in Lemma 5.5 can be used to construct an exact algorithm in linear time. This can be realized by reorganizing the sums in a nested structure. Then, we compute the inner sum and memorize the results in vectors of cumulated sums. If we compute outer sums, we can use these memorized values to avoid multiple computations of them.

For $K = 3$, this idea is very simple:

Example 5.9. *Let $\mathbf{r} = (r_1, \ldots, r_N)^\top$ be a realization of the residual vector with B blocks and with blocks lengths $q_1, \ldots, q_B$ and $K = 3$. According to Lemma 5.5, the 3-depth can be represented by*

$$\binom{N}{3} d_3(\mathbf{r}) = \sum_{(i_1,i_2,i_3)^\top \in \mathcal{A}_{3,B}} q_{i_1} q_{i_2} q_{i_3} = \sum_{i_1=1}^{B-2} \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-1} \sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B} q_{i_1} q_{i_2} q_{i_3}.$$

We exchange the sums by putting the sum with the index i_2 into the front. Thus, the other two sums only depend on i_2:

$$\sum_{i_1=1}^{B-2} \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-1} \sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B} q_{i_1} q_{i_2} q_{i_3} = \sum_{i_2=2}^{B-1} q_{i_2} \left(\sum_{\substack{i_1=1 \\ i_2-i_1 \text{ odd}}}^{i_2-1} q_{i_1} \right) \left(\sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B} q_{i_3} \right) = \sum_{i_2=2}^{B-1} q_{i_2} \mathcal{F}(i_2) \mathcal{G}(i_2)$$

with $\mathcal{F}(i_2) := \sum_{\substack{i_1=1 \\ i_2-i_1 \text{ odd}}}^{i_2-1} q_{i_1}$ *and* $\mathcal{G}(i_2) := \sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B} q_{i_3}$, $i_2 = 2, \ldots, B-1$.

The values of $(\mathcal{F}(i_2), \mathcal{G}(i_2))_{i_2=2,\ldots,B-1}$ can be computed with time complexity of $\Theta(B)$ by a cumulative sum. Note that these values will be computed separately for the cases i_2 odd or even. This cumulative sum will be stored. Then computing the 3-depth by

$$d_3(\mathbf{r}) = \frac{1}{\binom{N}{3}} \sum_{i_2=2}^{B-1} q_{i_2} \mathcal{F}(i_2) \mathcal{G}(i_2)$$

has time complexity $\Theta(B)$. Note that the computation of $q_1, \ldots, q_B$ has time complexity $\Theta(N)$ and the total computation of the 3-depth is linear in N.

For $K \geq 4$, we have to store nested cumulated sums. Before the general case will be introduced, we will give another example for $K = 4$ for an easier understanding.

Example 5.10. *Let $\mathbf{r} = (r_1, \ldots, r_N)^\top$ be a realization of the residual vector with B blocks and with blocks lengths $q_1, \ldots, q_B$ and $K = 4$. Then, according to Lemma 5.5, we obtain*

$$\begin{aligned}\binom{N}{4} d_4(\mathbf{r}) &= \sum_{i_1=1}^{B-3} q_{i_1} \left(\sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-2} q_{i_2} \sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B-1} q_{i_3} \sum_{\substack{i_4=i_3+1 \\ i_4-i_3 \text{ odd}}}^{B} q_{i_4} \right) \\ &= \sum_{i_1=1}^{B-3} q_{i_1} \mathcal{G}_3(i_1), \end{aligned} \tag{64}$$

where $\mathcal{G}_3(i_1), i_1 = 1, \ldots, B-3$, is defined in three steps recursively:

$$\begin{aligned}\mathcal{G}_3(i_1) &= \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-2} q_{i_2} \mathcal{G}_2(i_2), i_1 = 1, \ldots, B-3, \\ \mathcal{G}_2(i_2) &= \sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B-1} q_{i_3} \mathcal{G}_1(i_3), i_2 = 2, \ldots, B-2, \\ \mathcal{G}_1(i_3) &= \sum_{\substack{i_4=i_3+1 \\ i_4-i_3 \text{ odd}}}^{B} q_{i_4}, i_3 = 3, \ldots, B-1, . \end{aligned}$$

In the first step, $\mathcal{G}_1(i_3), i_3 = 3, \ldots, B-1$ can be computed with time complexity $\Theta(B)$ by a cumulated sum and stored. We can compute $\mathcal{G}_2(i_2), i_2 = 2, \ldots, B-2$ with linear time complexity in B as well by using the memorized values of $\mathcal{G}_1(i_3)$ from the first step. Analogously, we can compute $\mathcal{G}_3(i_1), i_1 = 1, \ldots, B-3$ with a $\Theta(B)$ time complexity. Finally, we also obtain a linear computation of the sum in Formula (64). Note that (i_1, i_2, i_3, i_4) with parities (even, odd, even, odd) and (odd, even, odd, even) have to be considered separately, see the explanations after Theorem 5.11 for more details.

This procedure in Example 5.10 by factorizing q_{i_1} can also be done for $K = 3$. However, the procedure in Example 5.9 can appear to be more natural. Therefore, we also wanted to introduce Example 5.9 done in another manner. The next theorem summarizes the general procedure for all K.

Theorem 5.11. *Let* $\mathbf{r} = (r_1, \ldots, r_N)^\top$ *be a realization of the residual vector with* B *blocks and with blocks lengths* $q_1, \ldots, q_B$*. Then*

$$d_K(\mathbf{r}) = \frac{1}{\binom{N}{K}} \sum_{i_1=1}^{B-K+1} q_{i_1} \mathcal{G}_{K-1}(i_1) \text{ with}$$

$$\mathcal{G}_0 \equiv 1, \mathcal{G}_j(i_{K-j}) := \sum_{\substack{i_{K-j+1}=i_{K-j}+1 \\ i_{K-j+1}-i_{K-j} \text{ odd}}}^{B-j+1} q_{i_{K-j+1}} \mathcal{G}_{j-1}(i_{K-j+1}), j = 1, \ldots, K-1.$$

Proof of Theorem 5.11: This theorem can be proven by a simple induction of the following assertion for all given $i_1 = 1, \ldots, B-K+1$:

$$\mathcal{G}_{K-1}(i_1) = \sum_{\substack{(i_2,\ldots,i_K)^\top \in \mathcal{A}_{K-1,B} \\ i_2 > i_1, i_2 - i_1 \text{ odd}}} \prod_{k=2}^{K} q_{i_k} \text{ for all } K \geq 2. \tag{65}$$

The base step for $K = 2$ follows by defining $\mathcal{A}_{1,B} := \{1, \ldots, B\}$:

$$\mathcal{G}_1(i_1) = \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B} q_{i_2} \mathcal{G}_0(i_2) = \sum_{\substack{i_2 \in \mathcal{A}_{1,B} \\ i_2 > i_1, i_2 - i_1 \text{ odd}}} q_{i_2}.$$

We assume now that Formula (65) holds for a fixed K. Then, we have to prove this formula for $K+1$. By definition, we have:

$$\mathcal{G}_K(i_1) = \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-K} q_{i_2} \mathcal{G}_{K-1}(i_2). \tag{66}$$

Then, the induction step implies

$$\mathcal{G}_{K-1}(i_2) = \sum_{\substack{(i_3,\ldots,i_{K+1})^\top \in \mathcal{A}_{K-1,B} \\ i_3 > i_2, i_3 - i_2 \text{ odd}}} \prod_{k=3}^{K+1} q_{i_k} \tag{67}$$

and by plugging Formula (67) into Formula (66), we obtain

$$\mathcal{G}_K(i_1) = \sum_{\substack{i_2=i_1+1 \\ i_2-i_1 \text{ odd}}}^{B-K} q_{i_2} \left(\sum_{\substack{(i_3,\ldots,i_{K+1})^\top \in \mathcal{A}_{K-1,B} \\ i_3 > i_2, i_3 - i_2 \text{ odd}}} \prod_{k=3}^{K+1} q_{i_k} \right) = \sum_{\substack{(i_2,\ldots,i_{K+1})^\top \in \mathcal{A}_{K,B} \\ i_2 > i_1, i_2 - i_1 \text{ odd}}} \prod_{k=2}^{K+1} q_{i_k}.$$

Using Formula (65), the main assertion of Theorem 5.11 follows as well. □

We will discuss briefly the implementation based on Theorem 5.11. First of all, we split the main sum into the odd and even parts:

$$\sum_{i_1=1}^{B-K+1} q_{i_1}\mathcal{G}_{K-1}(i_1) = \sum_{\substack{i_1=1 \\ i_1 \text{ odd}}}^{B-K+1} q_{i_1}\mathcal{G}_{K-1}(i_1) + \sum_{\substack{i_1=2 \\ i_1 \text{ even}}}^{B-K+1} q_{i_1}\mathcal{G}_{K-1}(i_1).$$

This separation is practicable since we have to consider the terms $\mathcal{G}_{K-1}(i_1)$ distinctly for i_1 odd or even. The values $\mathcal{G}_j(i_{K-j})$ for $j = 1, \ldots, K-1$, $i_{K-j} \in \{K-j, \ldots, B-j\}$ can be memorized in two matrices based on these two cases for i_1.
Moreover, we consider the two cases $K+B$ even and odd. If $K+B$ is even, we compute the following two matrices:

$$\mathcal{Q}_1 = \begin{pmatrix} \mathcal{G}_1(B) & \mathcal{G}_1(B-2) & \cdots & \mathcal{G}_1(K) \\ \mathcal{G}_2(B-1) & \mathcal{G}_2(B-3) & \cdots & \mathcal{G}_2(K-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{G}_{K-1}(B-K+2) & \mathcal{G}_{K-1}(B-K) & \cdots & \mathcal{G}_{K-1}(2) \end{pmatrix} \in \mathbb{N}^{(K-1)\times\left(\frac{B-K}{2}+1\right)},$$

$$\mathcal{Q}_2 = \begin{pmatrix} \mathcal{G}_1(B-1) & \mathcal{G}_1(B-3) & \cdots & \mathcal{G}_1(K+1) \\ \mathcal{G}_2(B-2) & \mathcal{G}_2(B-4) & \cdots & \mathcal{G}_2(K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{G}_{K-1}(B-K+1) & \mathcal{G}_{K-1}(B-K-1) & \cdots & \mathcal{G}_{K-1}(3) \end{pmatrix} \in \mathbb{N}^{(K-1)\times\left(\frac{B-K}{2}\right)}.$$

The first line from each matrix can be computed in linear time in B by a cumulative sum. Each of the following lines has to be computed by the above one in order to have linear time in B. If $K+B$ is odd, we reduce computing the K-depth to the above case by the following lemma (Leckey et al., 2020):

Lemma 5.12. *Let $K \geq 3$, $B \geq K$ and $N \geq B$. We define the function:*

$$d_{K,N,B} : (0,N)^B \to \mathbb{R},\ d_{K,N,B}(q_1, \ldots, q_B) = \frac{1}{\binom{N}{K}} \sum_{(i_1,\ldots,i_K)^\top \in \mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k}.$$

If $K+B$ is odd, then

$$d_{K,N,B}(q_1, \ldots, q_B) = d_{K,N,B-1}(q_1 + q_B, q_2, \ldots, q_{B-1}). \tag{68}$$

Note that $d_{K,N,B}$ has values in the B-dimensional cube $(0,N)^B$ although we will just need it for $(0,N)^B \cap \mathbb{N}^B$ in this chapter. However, Lemma 5.12 will be applied in Chapter 5.4 with the B-dimensional cube as the domain of $d_{K,N,B}$.

Proof of Lemma 5.12: For this proof, all column vectors are written as row vectors. For $x \in \mathbb{R}$ and $\mathbf{w} = (w_1, \ldots, w_J) \in \mathbb{R}^J$, we define

$$(x, \mathbf{w}) := (x, w_1, \ldots, w_J) \text{ and } (\mathbf{w}, x) := (w_1, \ldots, w_J, x).$$

First, we will prove for $K + B$ odd and $\mathbf{i} \in \{2, \ldots, B-1\}^{K-1}$:

(a) $(1, \mathbf{i}) \in \mathcal{A}_{K,B}$ if and only if $(\mathbf{i}, B) \in \mathcal{A}_{K,B}$,

(b) there is no vector $\mathbf{j} \in \{2, \ldots, B-1\}^{K-2}$ with $(1, \mathbf{j}, B) \in \mathcal{A}_{K,B}$.

For (a) note that $(1, \mathbf{i}) \in \mathcal{A}_{K,B}$ requires $\mathbf{i} = (i_1, \ldots, i_{K-1})$ to start with an even index i_1. Now, we have to consider the two cases K odd, B even and K even, B odd such that $K + B$ is odd. For the case K odd and B even, the length $K-1$ of $\mathbf{i}$ is even. Therefore, the last index i_{K-1} of the vector has to be odd if i_1 is even. Since B is even, B and i_{K-1} alternate between even and odd. Hence $(\mathbf{i}, B) \in \mathcal{A}_{K,B}$. Similarly, $(\mathbf{i}, B) \in \mathcal{A}_{K,B}$ requires i_{K-1} to be odd and thus i_1 has to be even. This implies $(1, \mathbf{i}) \in \mathcal{A}_{K,B}$. For the case K even and B odd, the length $K-1$ of $\mathbf{i}$ is odd. If i_1 is even, then i_{K-1} has to be even as well. Since B is odd, B and i_{K-1} alternate between odd and even and thus $(\mathbf{i}, B) \in \mathcal{A}_{K,B}$. The reversed implication for $(\mathbf{i}, B) \in \mathcal{A}_{K,B}$ follows analogously.
For (b) assume for the sake of contradiction that $(1, \mathbf{j}, B) \in \mathcal{A}_{K,B}$ for a vector $\mathbf{j} = (j_1, \ldots, j_{K-2}) \in \{2, \ldots, B-1\}^{K-2}$. Since 1 is odd, j_1 is always even. For K odd and B even, the length $K-2$ of $\mathbf{j}$ is odd. Thus, j_1 and j_{K-2} have the same parity and j_{K-2} must be even as well. However, B has to be odd in order to have $(1, \mathbf{j}, B) \in \mathcal{A}_{K,B}$, which leads to a contradiction to the assumption that B is even. If K is even and B is odd, $(1, \mathbf{j}, B) \in \mathcal{A}_{K,B}$ implies that j_1 is even and j_{K-2} is odd, since $K-2$ is even. This leads analogously to a contradiction since j_{K-2} and B are both odd.
For the next part of the proof, we consider the following sets:

$$\begin{aligned} \mathcal{V}_{K,B} &:= \{\mathbf{i} \in \{2, \ldots, B-1\}^{K-1}; (1, \mathbf{i}) \in \mathcal{A}_{K,B}\}, \\ \mathcal{W}_{K,B} &:= \mathcal{A}_{K,B} \cap \{2, \ldots, B-1\}^K. \end{aligned}$$

We split $\mathcal{A}_{K,B}$ into three parts

(i) $\mathcal{A}^{(1)}_{K,B} := \{1\} \times \mathcal{V}_{K,B}$,

(ii) $\mathcal{A}^{(2)}_{K,B} := \mathcal{V}_{K,B} \times \{B\}$,

(iii) $\mathcal{A}^{(3)}_{K,B} := \mathcal{W}_{K,B}$.

Vectors of the form $(i_1, \ldots, i_K) \in \mathcal{A}_{K,B}$ with $i_1 = 1$ and $i_K = B$ are impossible according to (b). Then (a) implies the following disjoint decomposition:

$$\mathcal{A}_{K,B} = \mathcal{A}_{K,B}^{(1)} \uplus \mathcal{A}_{K,B}^{(2)} \uplus \mathcal{A}_{K,B}^{(3)}.$$

Then

$$\sum_{(i_1,\ldots,i_K)\in\mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k} = (q_1 + q_B) \sum_{(i_1,\ldots,i_{K-1})\in\mathcal{V}_{K,B}} \prod_{k=1}^{K-1} q_{i_k} + \sum_{(i_1,\ldots,i_K)\in\mathcal{W}_{K,B}} \prod_{k=1}^{K} q_{i_k}.$$

Furthermore, note that

$$\mathcal{A}_{K,B-1} = \mathcal{A}_{K,B}^{(1)} \uplus \mathcal{A}_{K,B}^{(3)}.$$

In particular, if $\tilde{q}_1 = q_1 + q_B$ and $\tilde{q}_j = q_j$ for $j = 2, \ldots, B-1$, then

$$\sum_{(i_1,\ldots,i_K)\in\mathcal{A}_{K,B-1}} \prod_{k=1}^{K} \tilde{q}_{i_k} = \tilde{q}_1 \sum_{(i_1,\ldots,i_{K-1})\in\mathcal{V}_{K,B}} \prod_{k=1}^{K-1} \tilde{q}_{i_k} + \sum_{(i_1,\ldots,i_K)\in\mathcal{W}_{K,B}} \prod_{k=1}^{K} \tilde{q}_{i_k}.$$

Hence $d_{K,N,B}(q_1, \ldots, q_B) = d_{K,N,B-1}(\tilde{q}_1, \ldots, \tilde{q}_{B-1})$, which is the assertion. □

According to Lemma 5.12, we can merge the first and last block and consider the block number $B-1$ for the above algorithm if $K+B$ is odd since $K+B-1$ is even then. Moreover, we should treat the special cases for $K \leq B+1$ separately. For $K < B$, we can conclude that the sum is zero by Lemma 5.6, p. 79. For $K = B$, we only have the compute the product:

$$d_K(\mathbf{r}) = \frac{1}{\binom{N}{K}} \prod_{k=1}^{B} q_k.$$

Note that in this case, the first matrix $\mathcal{Q}_1$ only contains one column and the second matrix $\mathcal{Q}_2$ cannot be considered, because we would have zero columns. The case $K = B+1$ can be reduced to the case $K = B$ as well since $K+B$ is always odd for $K = B+1$.

Although, we only have integers in the matrices $\mathcal{Q}_1$ and $\mathcal{Q}_2$, it is not recommendable to implement this algorithm by filling the matrix with variables declared as integers. The computed numbers in the matrices will have large values which can cause an overflow for increasing N or K. E.g., the usual machine grid is $[-2^{31}+1, 2^{31}-1] \cap \mathbb{Z}$ in `R`, see the `R`-documentation of `integer`. Therefore, we implemented this algorithm by filling the matrix with variables declared as doubles. `R` provides values from

about 2^{-1023} to 2^{1023} and -2^{1023} to -2^{-1023}, see the R-documentation of `double`. This accuracy of the floating point arithmetic is sufficient for the most applications. Nevertheless, note that for even larger N or K in big data situations, an infinite value can occur as an output. In this case, the algorithm could be extended with some case distinction for large N and K where we would divide the block lengths by 10^{-a} for some $a \in \mathbb{N}$ in order to not obtain too high values. At the end, we would multiply the result with a suitable compensation.

5.3 Runtime comparison between the algorithms from Chapter 4 and Chapter 5.2

In this chapter, we compare the different algorithms from Chapter 4 and 5.2 based on their runtimes. Chapter 4.1 offers an approximate algorithm for the K-depth with linear time complexity $\Theta(N)$ for all K. In Chapter 4.2, an exact computation of the K-depth with linear time complexity in N is derived for $K \in \{3, 4, 5\}$. In Chapter 5.2, we have found another linear *and* exact algorithm for all K. It is interesting to conclude for which K which algorithm is more recommendable. Therefore, we will analyze the runtimes by repeated computations. In contrast to the algorithms from Chapter 4, the time complexity of the algorithm from Chapter 5.2 depends on the number of blocks. Due to this dependency, we will consider a worst case scenario and an average case scenario for the algorithm from Chapter 5.2.

1. $B = N$ describes the *worst case* for the block implementation. The residual vector has then alternating signs. Situation with many blocks occurs e.g. in the case of negative correlations.

2. $B = \frac{N+1}{2}$ describes the expected behavior *under the true parameter* as an *average case*. In the runtimes studie, N will be even and $\left\lfloor \frac{N+1}{2} \right\rfloor$ will be considered therefore. We will consider residual vectors with the following sign structure for the sake of simplicity:

$$\mathbf{r} = (1, 1, -1, -1, 1, 1, -1, -1, \ldots)^\top.$$

Note that the algorithm from Chapter 5.2 is faster the smaller B. The best case scenario occurs when $B < K$ since only the signs of the residual vector and its block structure have to be computed then, since it is then clear that the K-depth is zero according to Lemma 5.6. The expected number of blocks are computed in the following lemma in order to explain the average case.

Lemma 5.13. *For a random vector* $\mathbf{R} = (R_1, \ldots, R_N)^\top$ *of independent random variables with* $\mathbb{P}(R_n > 0) = p_n$ *and* $\mathbb{P}(R_n < 0) = 1 - p_n$

$$\mathbb{E}(B(\mathbf{R})) = 1 + \sum_{n=2}^{N} (p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1}).$$

For the special case of $\mathbb{P}(R_n > 0) = \frac{1}{2} = \mathbb{P}(R_n < 0)$

$$\mathbb{E}(B(\mathbf{R})) = \frac{N+1}{2}.$$

Proof of Lemma 5.13: By the law of total probability and the independence of $R_1, \ldots, R_N$, we obtain

$$\begin{aligned}
&\mathbb{P}(\psi(R_n) \neq \psi(R_{n-1})) \\
&= \mathbb{P}(\psi(R_n) \neq \psi(R_{n-1}) \mid \psi(R_n) > 0, \psi(R_{n-1}) > 0) \cdot \mathbb{P}(\psi(R_n) > 0, \psi(R_{n-1}) > 0) \\
&\quad + \mathbb{P}(\psi(R_n) \neq \psi(R_{n-1}) \mid \psi(R_n) > 0, \psi(R_{n-1}) < 0) \cdot \mathbb{P}(\psi(R_n) > 0, \psi(R_{n-1}) < 0) \\
&\quad + \mathbb{P}(\psi(R_n) \neq \psi(R_{n-1}) \mid \psi(R_n) < 0, \psi(R_{n-1}) > 0) \cdot \mathbb{P}(\psi(R_n) < 0, \psi(R_{n-1}) > 0) \\
&\quad + \mathbb{P}(\psi(R_n) \neq \psi(R_{n-1}) \mid \psi(R_n) < 0, \psi(R_{n-1}) < 0) \cdot \mathbb{P}(\psi(R_n) < 0, \psi(R_{n-1}) < 0) \\
&= 0 + 1 \cdot \mathbb{P}(\psi(R_n) > 0, \psi(R_{n-1}) < 0) + 1 \cdot \mathbb{P}(\psi(R_n) < 0, \psi(R_{n-1}) > 0) + 0 \\
&= \mathbb{P}(\psi(R_n) > 0, \psi(R_{n-1}) < 0) + \mathbb{P}(\psi(R_n) < 0, \psi(R_{n-1}) > 0) \\
&= p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1}.
\end{aligned}$$

Then, the expected value of the number of blocks is

$$\begin{aligned}
\mathbb{E}(B(\mathbf{R})) &= \mathbb{E}\left(1 + \sum_{n=2}^{N} \mathbb{1}\{\psi(R_n) \neq \psi(R_{n-1})\}\right) \\
&= 1 + \sum_{n=2}^{N} \mathbb{P}(\psi(R_n) \neq \psi(R_{n-1})) = 1 + \sum_{n=2}^{N} (p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1}).
\end{aligned}$$

If $p_n = \frac{1}{2}$ for all $n = 1, \ldots, N$, then

$$\mathbb{E}(B(\mathbf{R})) = 1 + \sum_{n=2}^{N} (p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1}) = 1 + \frac{N-1}{2} = \frac{N+1}{2}$$

and the assertion follows. □

We want to underline the interpretation of the expected value. The summands in Lemma 5.13 can be rewritten by $p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1} = p_n - 2p_np_{n-1} + p_{n-1}$. Figure 16 illustrates the behavior of the block lengths by considering neighbored

probabilities. The higher the distance between p_{n-1} and p_n, the more probable a

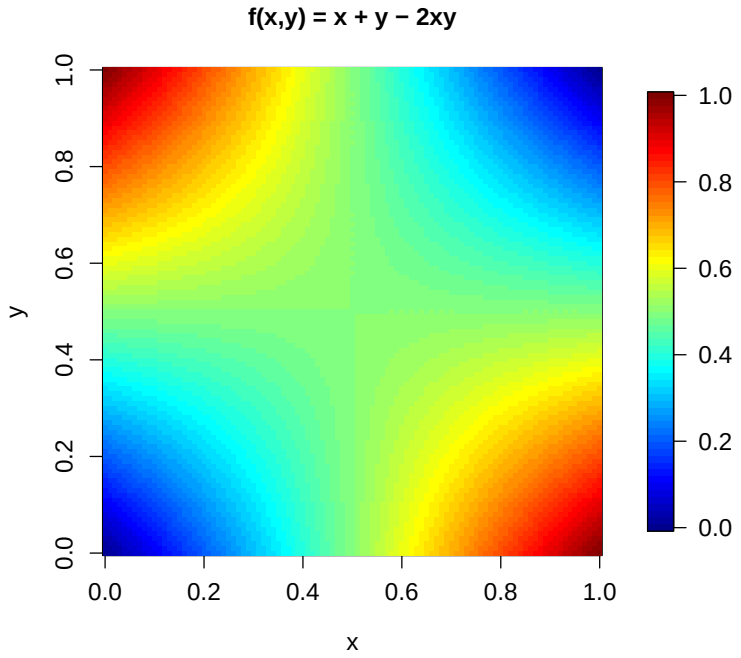

Figure 16: A visualization of the functional behavior of a summand of the expected block length for two neighbored probabilities

sign change occurs and thus a new block.

The blue edges in Figure 16 correspond to situations with similar consecutive probabilities and the red edges for alternating probabilities, i.e., probabilities with opposite values. Thus, the number of expected blocks can be computed for positively or negatively correlated residuals under some assumptions on the dependency structure of $\mathbf{R}$ in Lemma 5.13 (e.g., an autoregressive time series).

In Figure 17, the runtimes of the approximate algorithm from Chapter 4.1 and the algorithm from Chapter 5.2 are compared. The algorithm from Chapter 5.2 is even in the worst case situation (i.e., alternating signs) better than the algorithms in Chapter 4.1. Since these are only the worst and average cases, we can conclude that the block algorithm is clearly better. Moreover, the improvements between the average and worst case are not high. The boxplots are omitted here compared to Chapter 4.3 since the distinction between the lines is better visible without boxplots in the figure.

We further see that for increasing K, the computational costs do not increase that much if the block implementation is used. Therefore, we compare the runtimes for $K \in \{3, 5, 7\}$ in Figure 18. This result leads to the conclusion that increasing K does not have much influence on the runtime. Compared to the algorithms from Chapter 4, this algorithm is clearly better from every perspective. Thus, the block

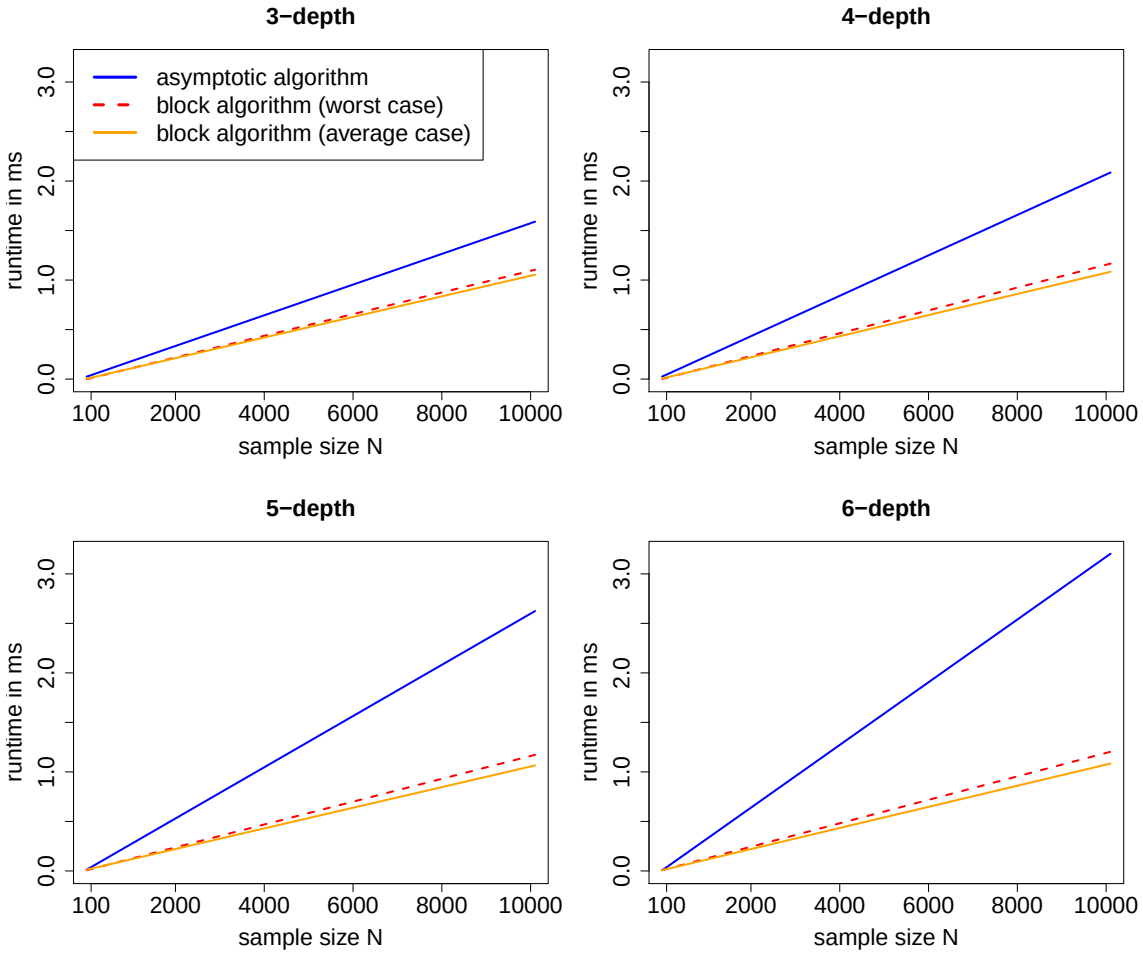

Figure 17: The simulated runtimes of the approximate algorithm in Chapter 4.1 and the exact algorithm in Chapter 5.2 for $K \in \{3, 4, 5, 6\}$

algorithm from Chapter 5.2 is recommended to use generally.

As an explanation for the better performance of the algorithm from Chapter 5.2, we will subsequently analyze the number of arithmetic operations in each implementation. Recall that the formula from Lemma 4.1, p. 56, needs to compute the following expression in advance:

$$S_{n,\alpha}^{N} = \sum_{k=1}^{n} \left(\frac{k}{N}\right)^{\alpha} \psi(r_n) \text{ for } n \in \{1, \ldots, N\} \text{ and } \alpha \in \{0, \ldots, K-2\}. \tag{69}$$

In the block algorithm, we compute $\mathcal{Q}_1$ and $\mathcal{Q}_2$ in advance. We consider the case $B = N$ here to compare the algorithm from Chapter 4.2 with the block algorithm in the worst case. The additional power functions in Formula (69) increase the computational costs compared to the algorithm from Chapter 5 where only the signs

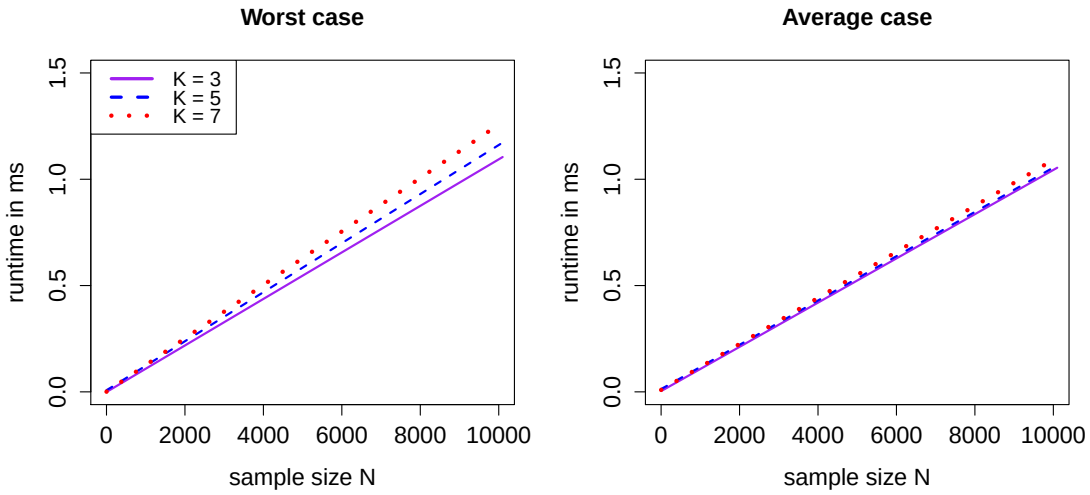

Figure 18: The simulated runtimes of the algorithm in Chapter 5.2 for $K \in \{3, 5, 7\}$ for the worst case and the average case

of the residuals are summed up. The number of arithmetic operations to compute $S_{n,\alpha}$ for $\alpha \in \{0, \ldots, K-1\}$ and for $\mathcal{Q}_1, \mathcal{Q}_2$ (with $B = N$) are for both approximately $N \cdot (K-1)$ iterations nevertheless.
Furthermore, the resulting value of the K-depth can be computed by

$$\begin{aligned} & N\left(d_K(\mathbf{r}) - \frac{1}{2^{K-1}}\right) \\ = & -\frac{N^{K-1}K(K-1)}{2\langle N\rangle_K} \sum_{j=0}^{K-2} \binom{K-2}{j} \sum_{n=2}^{N} \left(\frac{1}{2} - \frac{n}{N}\right)^{K-2-j} \psi(r_n) S_{n-1,j}^N. \end{aligned}$$

Approximately $N \cdot (K-1)$ additional iterations are needed to compute this term. In contrast, the block algorithm only needs approximately N iterations by computing

$$d_K(\mathbf{r}) = \frac{1}{\binom{N}{K}} \sum_{i_1=1}^{N-K+1} q_{i_1} \mathcal{G}_{K-1}(i_1)$$

from Theorem 5.11, p. 84. We can see that the preparation step only depends on K in the block algorithm, i.e., the asymptotic algorithm has more arithmetic operations.

5.4 Further impact of the block implementation

The block algorithm does not only provide an efficient computation of the K-depth but also a better understanding of theoretical properties. According to Lemma 5.6, p. 79, we obtain conditions on the block structures when the K-depth is zero. The-

refore, the following simple heuristic seems to be plausible: The lower the number of blocks is, the lower the value of the K-depth since a smaller number of alternating signs occur. On the other hand, the more blocks exist, the more elements the set $\mathcal{A}_{K,B}$ has and we expect high values for the K-depth.
In Leckey et al. (2020), various properties of the K-depth are given in dependence of the block structure. In this subsection, we will highlight one conjecture from this paper that is based on the block representation, because it implies some confirmations of the maximality and minimality properties of the K-depth. This understanding can have impact for further theoretical research, e.g., for consistency analysis of statistical tests based on the K-depth.

Conjecture 5.14. *Let $K \geq 3$, $B \geq K$ and $N \geq B$. We define*

$$d_{K,N,B} : (0,N)^B \to \mathbb{R},\ d_{K,N,B}(q_1,\ldots,q_B) = \frac{1}{\binom{N}{K}} \sum_{(i_1,\ldots,i_K)^\top \in \mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k} \text{ and}$$

$$\mathcal{L}_{K,N,B} = \operatorname{argmax} \left\{ d_{K,N,B}(q_1,\ldots,q_B); (q_1,\ldots,q_B)^\top \in (0,N)^B; \sum_{b=1}^{B} q_b = N \right\}.$$

Then the following holds:

(a) If $K+B$ is even then:

$$\mathcal{L}_{K,N,B} = \left\{ \left(\frac{N}{B}, \ldots, \frac{N}{B} \right)^\top \right\} \text{ with } d_{K,N,B}\left(\frac{N}{B}, \ldots, \frac{N}{B} \right) = \frac{\langle \frac{B+K-2}{2} \rangle_{K-1}}{B^{K-1}} \frac{N^K}{\langle N \rangle_K} \quad (70)$$

where $\langle x \rangle_k := \frac{x!}{(x-k)!}$ for $x, k \in \mathbb{N}$. If $\frac{N}{B} \in \mathbb{N}$, then this is the maximal K-depth for residuals of length N with B blocks.

(b) If $K+B$ is odd then:

$$\mathcal{L}_{K,N,B} = \left\{ \left(\frac{\beta N}{B-1}, \frac{N}{B-1}, \ldots, \frac{N}{B-1}, \frac{(1-\beta)N}{B-1} \right)^\top ; \beta \in (0,1) \right\} \text{ with}$$

$$d_{K,N,B}\left(\frac{\beta N}{B-1}, \frac{N}{B-1}, \ldots, \frac{N}{B-1}, \frac{(1-\beta)N}{B-1} \right) = \frac{\langle \frac{B+K-3}{2} \rangle_{K-1}}{(B-1)^{K-1}} \frac{N^K}{\langle N \rangle_K}. \quad (71)$$

If $\frac{N}{B-1} \in \mathbb{N}$ and $\frac{\beta N}{B-1} \in \mathbb{N}$ for $\beta \in (0,1)$, then this is the maximal K-depth for residuals of length N with B blocks.

Conjecture 5.14 yields the block configuration for maximal $d_{K,N,B}$ under fixed K, N and B. The values of the function $d_{K,N,B}$ for its assumed maximal points in Formula (70) and (71) can be derived by Corollary 5.8, p. 81, cf. Leckey et al. (2020). For

some N and B, we cannot construct blocks $q_1, \ldots, q_B \in \mathbb{N}$ such that we obtain the maximum of $d_{K,N,B}$. In these cases, the conjecture only delivers an *upper bound*. The case $K = 2$ is excluded in this conjecture since the 2-depth is maximal if the sum of the signs from the residuals is minimal (Leckey et al., 2020). This is the case if the number of positive and negative signs are equal (or nearly equal for odd N). The distinction between the cases $K + B$ odd or even may be surprising, but it gets intuitive if Lemma 5.12, p. 85 is recalled. Thus, maximizing under the situation in (b), i.e. $K + B$ is odd, can be translated to the situation (a) for $K + B - 1$. Therefore, if Conjecture 5.14 (a) is correct, then Lemma 5.12 implies the statement in Conjecture 5.14 (b) as well.
Finding a proof for this conjecture might be easy at first glance by a Lagrange optimization approach. The members of our research group could so far prove that the points in the set $\mathcal{L}_{K,N,B}$ from the conjecture are critical points. For $K \leq 4$, we could also prove that the points in $\mathcal{L}_{K,N,B}$ are the only critical points by solving the resulting system of nonlinear equations from the Lagrange approach analytically. However, we have not been successful to prove the uniqueness for $K \geq 5$ yet.
The next lemma shows the monotony of the maximal values of $d_{K,N,B}$ in N.

Lemma 5.15. *Let $K \geq 3$. We define the function*

$$f_K : \mathbb{N} \cap [K, \infty) \to \mathbb{R}, f_K(B) := \begin{cases} \frac{\langle \frac{B+K-2}{2} \rangle_{K-1}}{B^{K-1}} & \text{for } K+B \text{ even,} \\ \frac{\langle \frac{B+K-3}{2} \rangle_{K-1}}{(B-1)^{K-1}} & \text{for } K+B \text{ odd.} \end{cases}$$

Then f_K is monotonously increasing.

Proof of Lemma 5.15: First, we consider the case $K + B$ even and show that

$$f_K(B) \leq f_K(B+2).$$

This inequality is equivalent to

$$\frac{1}{B^{K-1}} \prod_{i=0}^{K-2} \left(\frac{B+K-2-2i}{2} \right) \leq \frac{1}{(B+2)^{K-1}} \prod_{i=0}^{K-2} \left(\frac{B+K-2i}{2} \right).$$

We can divide the inequality by $\prod_{i=0}^{K-3} \left(\frac{B+K-2-2i}{2} \right)$ and obtain:

$$\frac{B-K+2}{B^{K-1}} \leq \frac{B+K}{(B+2)^{K-1}}.$$

After applying the logarithm, we obtain the following equivalent formula:

$$\log(B+K) + (K-1)\log(B) - \log(B-K+2) - (K-1)\log(B+2) \geq 0. \quad (72)$$

We prove this inequality by the following argumentation. Let

$$h_K(B) := \log(B+K) + (K-1)\log(B) - \log(B-K+2) - (K-1)\log(B+2).$$

For $K = B$, $K + B = 2K$ is still even. It is straight forward to obtain by using elementary logarithm identities:

$$h_K(K) = \log\left(\frac{K^K}{(K+2)^{K-1}}\right).$$

For $K \geq 3$, it holds $\frac{K^K}{(K+2)^{K-1}} > 1$ and thus $h_K(K) > 0$. Taking the derivative of h_K yields

$$h'_K(B) = \frac{1}{B+K} + \frac{K-1}{B} - \frac{1}{B-K+2} - \frac{K-1}{B+2} \overset{!}{<} 0.$$

It is straight forward to show that $h'_K(B) \overset{!}{<} 0$ for $K \geq 3$ is equivalent to

$$2(K-1)(-K^2+2K) \overset{!}{<} 0 \quad (73)$$

by finding the common denominator of all fractions. Formula (73) is correct for $K \geq 3$ since $(K-1) > 0$ and $(-K^2+2K) < 0$. By $\lim_{B\to\infty} h_K(B) = 0$, $h_K(K) > 0$ and $h_K(B)$ is monotonically decreasing for increasing $B \geq K$, Formula (72) follows. Furthermore, it is straight forward to check that

$$f_K(B) = f_K(B+1) \text{ for } K+B \text{ even.}$$

This implies that f_K is monotonously increasing in B for $B \geq K$. □

Multiplying f_K with $\frac{N}{\langle N\rangle_K}$ leads to the supposed maximal values of $d_{K,N,B}$. Since f_K is increasing in B, the maximal values are increasing in B as well. Thus, the K-depth is maximal under alternating signs for $N+K$ even. For $N+K$ odd, this argumentation does not yield the conclusion of the maximality under alternating signs since the upper bounds are not sharp. Due to Lemma 5.12, alternating signs are not the only situation where the K-depth is maximal for $N+K$ odd especially.

6 Tests based on the K-sign depth

In this chapter, two types of tests based on the K-depth will be introduced. We can test the fit of a model parameter or the independence of residuals which is basically testing one of the two conditions from Assumption 2.1. For the first test, we assume that (A1) is correct and we test (A2) or vice versa for the second test. These two types of tests are presented in Chapter 6.1.
In Chapter 6.2, we discuss how to compute the asymptotic quantiles for $K \geq 3$. For small sample sizes, we can consider the exact distribution of the K-depth and compute its quantiles exactly. The exact distribution can be computed since the K-depth is *distribution free* under the true parameter. For medium-sized sample sizes, we can simulate the finite sample quantiles by a large number of repeated evaluations of the K-depth. In Chapter 6.3, we study the type-I errors of tests based on the asymptotic quantiles. The type-I errors are estimated in simulation studies by the rate of rejections under residuals satisfying Assumption 2.1. In Chapter 6.4, we consider testing the assumption (A2) under several polynomial models and compute the power under several parameters by simulation studies. The hypotheses are referred as *point hypotheses* since the null hypothesis only contains a single parameter:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{74}$$

This simulation study leads to conclusions how to choose the hyper-parameter K for a given model class. Moreover, we will understand the behavior of the power functions under several sign change structures for given K. In Chapter 6.5, we extend the class of tested hypotheses by *set hypotheses*:

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0, H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0.$$

In contrast to Formula (74), the null hypothesis is not restricted to a single parameter. For this method, we will introduce an approach to test H_0 by solving an optimization problem with the highest K-depth under $\boldsymbol{\Theta}_0$ as a test statistic.
In Chapter 6.6, we consider tests for stochastic independence or also known as tests for randomness for residuals $R_1, \ldots, R_N$ with a particular order (e.g., the time):

$$H_0 : R_1, \ldots, R_N \text{ are independent}, H_1 : R_1, \ldots, R_N \text{ are not independent.}$$

Tests based on the K-depth and some modifications will be presented.

6.1 Types of tests based on the K-sign depth

Let $q^K_{\alpha,N}$ be the α-quantile of the *exact* distribution of the K-depth and let q^K_α be the *asymptotic* α-quantile. The K-depth is *distribution free* for all types of errors satisfying Assumption 2.1. This implies that the finite sample quantiles are available for every sample size N and should be used for small N in particular. Two main types of tests will be presented in this chapter. Recall that T_K denotes the rescaled K-depth given in Formula (6), p. 12.

6.1.1 Testing the fit of the model parameters

In the first case, we assume that the residuals satisfy (A1), i.e., the errors are independent. Then, we can test the fit of a suggested parameter $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

Theorem 6.1. *Let $Y_1, \ldots, Y_N$ be random variables that can be described by the model equation (1), p. 4, from Chapter 2.1:*

$$Y_n = g(\mathbf{X}_n, \boldsymbol{\theta}^*) + E_n \text{ for } n = 1, \ldots, N$$

with true parameter $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Let $E_1, \ldots, E_N$ satisfy Assumption 2.1.

(a) Consider the point-hypotheses

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

for a given $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Then we obtain an α-level test for the following decision rule based on the residual vector $\mathbf{R}(\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta}_0$:

$$\text{Reject } H_0 \text{ if } T_K(\mathbf{R}(\boldsymbol{\theta}_0)) < q^K_{\alpha,N}.$$

(b) Consider the set-hypotheses

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0, H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$$

such that $\boldsymbol{\Theta}_0 \uplus \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$ (disjoint union). Then we obtain an α-level test for the following decision rule based on the residual vector $\mathbf{R}(\boldsymbol{\theta})$ for arbitrary $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$:

$$\text{Reject } H_0 \text{ if } \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} T_K(\mathbf{R}(\boldsymbol{\theta})) < q^K_{\alpha,N}.$$

If we exchange $q^K_{\alpha,N}$ by q^K_α, we have asymptotic α-level tests.

Since the test statistic is discrete, we have to consider a strict inequality with the α-quantile in the decision rule. Note that tests with an analogous structure based on other depth functions have already been considered, e.g., for tests based on the Likelihood depth in Müller (2005).

Proof of Theorem 6.1: (a) holds by definition of the quantile. For proving (b), let $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$ be the true parameter. Then we have for the type-I-error:

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} T_K(\mathbf{R}(\boldsymbol{\theta})) < q_{\alpha,N}^K\right) \leq \mathbb{P}\left(T_K(\mathbf{R}(\boldsymbol{\theta}^*)) < q_{\alpha,N}^K\right) \leq \alpha$$

since $q_{\alpha,N}^K$ is the finite sample quantile. Exchanging $q_{\alpha,N}^K$ by q_α^K yields analogously:

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} T_K(\mathbf{R}(\boldsymbol{\theta})) < q_{\alpha}^K\right) \leq \mathbb{P}\left(T_K(\mathbf{R}(\boldsymbol{\theta}^*)) < q_{\alpha}^K\right) \xrightarrow[N\to\infty]{} \alpha$$

since q_α^K is the asymptotic quantile. □

Test (a) keeps the required type-I-error of α due to the finite sample quantiles. Since the statistic has a discrete distribution, the type-I-error is not exactly α. The convergence of the type-I-errors to α for the asymptotic test can be from above and below, see Chapter 6.3. For test (b), the behavior of the type-I-error does not only depend on N and K, but rather on the model and $\boldsymbol{\Theta}_0$. Moreover, the inequality in the proof of Theorem 6.1 may be rough in general. More details concerning the computation of the supremum can be found in Chapter 6.5.

6.1.2 Test for independence

The literature offers a large pool of tests for independence. One popular test for independence is the Durbin-Watson test based on an estimation of the Bravais-Pearson correlation coefficient under normality (Verbeek, 2012, p. 102). The asymptotic Ljung-Box test is based on estimated autocorrelations with a hyper-parameter for the number of considered lags (Ljung and Box, 1978). The Wald-Wolfowitz runs-test is based on dichotomous data. The signs of the residuals can also be interpreted as dichotomous data and used for the runs-test. Runs of the same signs is an identical term for 'blocks' from Chapter 5. This test checks if the number of blocks is too small or high. The critical values of this test are based on the conditional distribution of the number of blocks given the number of positive or negative signs which leads asymptotically to a normal distribution (Gibbons and Chakraborti, 2003, p. 78-85). Note that this is very close to the concept of the K-depth. The Von-Neumann-Ratio-Rank test modifies the Von-Neumann test statistic by ranks as a

robustification of the original test (Bartels, 1982). The distance correlation (Székely et al., 2007) detects general classes of dependencies which are not only correlated. A test procedure for independence can also be applied for the K-depth as the next theorem presents.

Theorem 6.2. *Let* $\mathbf{Z} = (Z_1, \ldots, Z_N)^\top$ *be a random vector in* $\mathbb{R}$ *with*

$$\mathbb{P}(Z_n > \mu) = \mathbb{P}(Z_n < \mu) = \frac{1}{2} \text{ for } n = 1, \ldots, N.$$

Consider the hypotheses

$$H_0 : Z_1, \ldots, Z_N \text{ are independent, } H_1 : Z_1, \ldots, Z_N \text{ are not independent.}$$

Then we obtain an α*-level test for the following decision rule based on the residuals* $\mathbf{R} = (Z_1 - \mu, \ldots, Z_N - \mu)^\top$*:*

$$\text{Reject } H_0 \text{ if } T_K(\mathbf{R}) < q^K_{\alpha_1,N} \text{ or } T_K(\mathbf{R}) > q^K_{1-\alpha_2,N}$$

with $\alpha_1, \alpha_2 \in [0, \alpha]$ *and* $\alpha_1 + \alpha_2 = \alpha$*. If we exchange* $q^K_{\alpha_i,N}$ *by* $q^K_{\alpha_i}$ *for* $i = 1, 2$*, we have an asymptotic* α*-level test.*

The main idea of this test is that under H_0, Assumption 2.1 is satisfied. Positively or negatively correlated residuals lead to a lower or higher number of sign changes, respectively, with untypical values of the K-depth under H_0. In practice, we have to estimate μ, e.g., by the sample median, in order to apply the test in Theorem 6.2. Note that other structures as trends or seasonal components from a time series should be removed before using this test (Neusser, 2016, p. 332).
The test in Theorem 6.2 is two-sided in contrast to the test in Theorem 6.1. We choose usually $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. A too small test statistic is an indication for positive correlation since consecutive signs are more frequently the same and the number of blocks smaller then. In contrast, too large values of the K-depth indicate negative correlation since a suspiciously high number of sign changes occur then.

6.2 Quantiles and their computation

The asymptotic quantiles of the 2-depth can be derived easily since we can state its asymptotic distribution explicitly. Theorem 3.29, p. 50, yields for a random vector $\mathbf{E} = (E_1, \ldots, E_N)^\top$ satisfying Assumption 2.1:

$$N\left(d_2(\mathbf{E}) - \frac{1}{2}\right) = \frac{1}{2}\left(1 - (S_1^N)^2\right)\frac{N^2}{\langle N \rangle_2} \xrightarrow[N\to\infty]{\mathcal{D}} \frac{1}{2}\left(1 - X^2\right)$$

with $X \sim \mathcal{N}(0,1)$ due to the Central Limit Theorem and $\frac{N^2}{\langle N \rangle_2} \xrightarrow[N\to\infty]{} 1$. We obtain the asymptotic quantiles of the 2-depth from this transformed χ^2_1-distribution directly. For $K \geq 3$, the asymptotic distribution does not have an easy closed form, as in Theorem 3.29, p. 50, already mentioned. The algorithms in Chapter 4 and 5 do not provide formulas that lead to a direct computation of realizations of the asymptotic distribution either.

The quantiles of the K-depth can be simulated by computing the multiple integrals of simulated paths of Brownian motions by Riemann approximations. Computing a J-dimensional integral by the Riemann sums with N steps has usually a time complexity of $\Theta(N^J)$. If a path of the Brownian motion is divided into N parts in the Riemann approximation, then computing a realization of the asymptotic representation of the K-depth has a time complexity of $\Theta(N^{K-2})$ based on naive Riemann sums (cf. the $(K-2)$-dimensional integral in Theorem 3.29). We will show that the integrals from the asymptotic representation can be computed in linear time $\Theta(N)$ by combinatorial arguments to sum up the Riemann sums efficiently.

In practice, it is often more handy to simulate a *high number* of realizations of the K-depth from large residual vectors satisfying Assumption 2.1 since the K-depth can be computed efficiently with the algorithms from Chapter 4 and 5. Then the empirical quantiles can be computed and approximate the asymptotic quantiles. However, the following results may be useful for some generalized versions of the K-depth, cf. Chapter 7.1.

Computation of the integral terms based on combinatorics

In the following, we will write $\mathbf{t} = (t_1, \ldots, t_J)$ for the integration variable and

$$W_J(\mathbf{t}) := (t_1 \wedge \ldots \wedge t_J) + \frac{1}{2},$$
$$V_J(\mathbf{t}) := (t_1 \vee \ldots \vee t_J) - \frac{1}{2}.$$

Further, we want to consider projections of W_J and V_J such that

$$W_j^i(\mathbf{t}) := (t_i \wedge \ldots \wedge t_j) + \frac{1}{2},$$
$$V_j^i(\mathbf{t}) := (t_i \vee \ldots \vee t_j) - \frac{1}{2},$$

for $1 \leq i < j \leq J$. If $i = 1$, we will omit notating this index, i.e.,

$$W_j(\mathbf{t}) := W_j^1(\mathbf{t}) \text{ and } V_j(\mathbf{t}) := V_j^1(\mathbf{t}).$$

The following calculation will be done for paths from a càdlàg-process $(S_t^N)_{t\in[0,1]}$ with a finite number of equidistant jumps at $t = \frac{1}{N}, \ldots, \frac{N-1}{N}, 1$. Without loss of generality, the following calculation can be done for paths from a Brownian motion $(B_t)_{t\in[0,1]}$ since we can consider the Brownian motion on N fixed time points and approximate it by a step function as follows:

$$B_{\frac{n}{N}} = S_{\frac{n}{N}}^N \text{ for } n \in \{0, 1, \ldots, N\}.$$

By increasing N, the continuous paths of the Brownian motion can be approximated arbitrarily precisely by step functions. Further, recall the definition of the sets $\mathcal{C}_{J,N}$ and $\mathcal{C}_J$, see Lemma 3.27, p. 47, and that the asymptotic integral representation of the K-depth from Theorem 3.22, p. 39, contains terms of the form:

$$\int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \mathbb{1}_{\mathcal{C}_J}(\mathbf{t}) \left(S_{W_J(\mathbf{t})\wedge 1}^N - S_{V_J(\mathbf{t})\vee 0}^N\right)^2 d\mathbf{t} \text{ for } J = 1, \ldots, K-2.$$

Note that for finite N, we actually have to consider $\mathbb{1}_{\mathcal{C}_{J,N}}$ instead of $\mathbb{1}_{\mathcal{C}_J}$. Since $\mathbb{1}_{\mathcal{C}_{J,N}}$ is more complicated to deal with and we want to consider the asymptotic integrals, we simplify $\mathbb{1}_{\mathcal{C}_{J,N}} \approx \mathbb{1}_{\mathcal{C}_J}$. In the next step, we split this J-dimensional integral into three parts by splitting the domain $\left(-\frac{1}{2},\frac{3}{2}\right)^J$ up such that:

$$\begin{aligned}
&\int_{\left(-\frac{1}{2},\frac{3}{2}\right)^J} \mathbb{1}_{\mathcal{C}_J}(\mathbf{t}) \left(S_{W_J(\mathbf{t})\wedge 1}^N - S_{V_J(\mathbf{t})\vee 0}^N\right)^2 d\mathbf{t} \\
&= \int_{\left(-\frac{1}{2},\frac{1}{2}\right)^J} \left(S_{W_J(\mathbf{t})}^N\right)^2 d\mathbf{t} + \int_{\left[\frac{1}{2},\frac{3}{2}\right)^J} \left(S_1^N - S_{V_J(\mathbf{t})}^N\right)^2 d\mathbf{t} \\
&+ \sum_{L=1}^{J-1} \binom{J}{L} \int_{\left(-\frac{1}{2},\frac{1}{2}\right)^{J-L}\times\left[\frac{1}{2},\frac{3}{2}\right)^L} \prod_{\substack{i=1,\ldots,J-L \\ j=J-L+1,\ldots,J}} \mathbb{1}\{|t_j - t_i| < 1\} \left(S_{W_J(\mathbf{t})\wedge 1}^N - S_{V_J(\mathbf{t})\vee 0}^N\right)^2 d\mathbf{t}.
\end{aligned}$$

Without loss of generality, the indices $i = 1, \ldots, J-L$ and $j = J-L+1, \ldots, J$ should correspond to $t_i \in \left(-\frac{1}{2},\frac{1}{2}\right)$ or $t_j \in \left[\frac{1}{2},\frac{3}{2}\right)$, respectively. Therefore, we can simplify for any $L = 1, \ldots, J-1$:

$$\begin{aligned}
&\int_{\left(-\frac{1}{2},\frac{1}{2}\right)^{J-L}\times\left[\frac{1}{2},\frac{3}{2}\right)^L} \prod_{\substack{i=1,\ldots,J-L \\ j=J-L+1,\ldots,J}} \mathbb{1}\{|t_j - t_i| < 1\} \left(S_{W_J(\mathbf{t})\wedge 1}^N - S_{V_J(\mathbf{t})\vee 0}^N\right)^2 d\mathbf{t} \\
&= \int_{\left(-\frac{1}{2},\frac{1}{2}\right)^{J-L}\times\left[\frac{1}{2},\frac{3}{2}\right)^L} \prod_{\substack{i=1,\ldots,J-L \\ j=J-L+1,\ldots,J}} \mathbb{1}\{t_j - t_i < 1\} \left(S_{W_{J-L}(\mathbf{t})}^N - S_{V_J^{J-L+1}(\mathbf{t})}^N\right)^2 d\mathbf{t}.
\end{aligned}$$

Then, each integral can be rewritten by

$$\begin{aligned}
&\int_{\left(-\frac{1}{2},\frac{1}{2}\right)^J} \left(S^N_{W_J(\mathbf{t})}\right)^2 d\mathbf{t} + \int_{\left[\frac{1}{2},\frac{3}{2}\right)^J} \left(S^N_1 - S^N_{V_J(\mathbf{t})}\right)^2 d\mathbf{t} \\
&+ \sum_{L=1}^{J-1} \binom{J}{L} \int_{\left(-\frac{1}{2},\frac{1}{2}\right)^{J-L}\times\left[\frac{1}{2},\frac{3}{2}\right)^L} \prod_{\substack{i=1,\dots,J-L \\ j=J-L+1,\dots,J}} \mathbb{1}\{|t_j - t_i| < 1\} \left(S^N_{W_J(\mathbf{t})\wedge 1} - S^N_{V_J(\mathbf{t})\vee 0}\right)^2 d\mathbf{t} \\
&= \int_{(0,1)^J} \left(S^N_{W_J(\mathbf{t})-\frac{1}{2}}\right)^2 d\mathbf{t} + \int_{(0,1)^J} \left(S^N_1 - S^N_{V_J(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t} \\
&+ \sum_{L=1}^{J-1} \binom{J}{L} \int_{(0,1)^J} \prod_{\substack{i=1,\dots,J-L \\ j=J-L+1,\dots,J}} \mathbb{1}\{t_j < t_i\} \left(S^N_{W_{J-L}(\mathbf{t})-\frac{1}{2}} - S^N_{V_J^{J-L+1}(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t}
\end{aligned}$$

The next lemma shows that each part can be computed in linear time.

Lemma 6.3. *Let $(S^N_t)_{t\in[0,1]}$ be a càdlàg-process with deterministic jump points in $t = \frac{1}{N}, \dots, \frac{N-1}{N}, 1$. Let $J \in \mathbb{N}$ and $L \in \{1, \dots, J-1\}$. Then*

$$\begin{aligned}
&(a) \int_{(0,1)^J} \left(S^N_{W_J(\mathbf{t})-\frac{1}{2}}\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{m=1}^{N} \left((N-m+1)^J - (N-m)^J\right) \left(S^N_{m/N}\right)^2, \\
&(b) \int_{(0,1)^J} \left(S^N_1 - S^N_{V_J(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{m=1}^{N} (m^J - (m-1)^J) \left(S^N_1 - S^N_{m/N}\right)^2 \\
&(c) \int_{(0,1)^J} \prod_{\substack{i=1,\dots,J-L \\ j=J-L+1,\dots,J}} \mathbb{1}\{t_j < t_i\} \left(S^N_{W_{J-L}(\mathbf{t})-\frac{1}{2}} - S^N_{V_J^{J-L+1}(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t} \\
&= \frac{1}{N^J} \Bigg(\sum_{m_2=1}^{N} m_2^L \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right) \left(S_{m_2/N}\right)^2 \\
&\quad + \sum_{m_1=1}^{N} (N-m_1+1)^{J-L} (m_1^L - (m_1-1)^L) \left(S^N_{m_1/N}\right)^2 \\
&\quad - 2 \sum_{m_2=1}^{N} \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right) S^N_{m_2/N} \tilde{S}^L(m_2) \Bigg)
\end{aligned} \tag{75}$$

where we define $\tilde{S}^L(m) := \sum_{\ell=1}^{m} (m^L - (m-1)^L) S^N_{m/N}$ for $m = 1, \dots, N$.

This lemma shows that all integrals can be computed in linear time in N. The vector $\left(\tilde{S}^L(1), \dots, \tilde{S}^L(N)\right)^\top$ can be computed in advance in linear time and stored. Such as in the algorithms in Lemma 4.1, p. 56, and Theorem 5.11, p. 84, we use the memorized vector so that we can compute the sum in Formula (75) with linear time complexity in N as well. The equations are also correct if we have Brownian motions as integrands on the left side and the limit for $N \to \infty$ on the right side.

Proof of Lemma 6.3: First, we can rewrite all integrals by sums over J indices:

$$(a) \int_{(0,1)^J} \left(S^N_{W_J(\mathbf{t})-\frac{1}{2}}\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S^N_{\min\{n_1,\dots,n_J\}/N}\right)^2$$

$$(b) \int_{(0,1)^J} \left(S^N_1 - S^N_{V_J(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S^N_1 - S^N_{\max\{n_1,\dots,n_J\}/N}\right)^2$$

$$(c) \int_{(0,1)^J} \prod_{\substack{i=1,\dots,J-L \\ j=J-L+1,\dots,J}} \mathbb{1}\{t_j < t_i\} \left(S^N_{W_{J-L}(\mathbf{t})-\frac{1}{2}} - S^N_{V_J^{J-L+1}(\mathbf{t})+\frac{1}{2}}\right)^2 d\mathbf{t}$$

$$= \frac{1}{N^J} \sum_{\substack{n_1,\dots,n_J=1 \\ n_1,\dots,n_{J-L} \geq n_{J-L+1},\dots,n_J}}^{N} \left(S^N_{\min\{n_1,\dots,n_{J-L}\}/N} - S^N_{\max\{n_{J-L+1},\dots,n_J\}/N}\right)^2$$

We will consider (a) - (c) separately. In the following, L will denote the number of t_i in the vector $\mathbf{t}$ so that $t_i \in \left[\frac{1}{2}, \frac{3}{2}\right)$. Then, the three cases are $L = 0$, $L = J$ and $L \in \{1, \dots, J-1\}$ are of interest for the assertions (a) - (c), respectively.

(a) **Case $L = 0$:**

We consider the first part:

$$\int_{\left(-\frac{1}{2},\frac{1}{2}\right)^J} \left(S^N_{W_J(\mathbf{t})}\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S^N_{\min\{n_1,\dots,n_J\}/N}\right)^2.$$

In order to simplify this sum, we consider the set $\mathcal{V}_J(m)$ for $m \in \{1, \dots, N\}$:

$$\mathcal{V}_J(m) := \Big\{(n_1, \dots, n_J) \in \{1, \dots, N\}^J; \min\{n_1, \dots, n_J\} = m\Big\}.$$

The cardinality of $\mathcal{V}_J(m)$ can be obtained by counting all J-tuples where exactly j entries are equal to m and the other $J - j$ entries are greater than m:

$$|\mathcal{V}_J(m)| = \sum_{j=1}^{J} \binom{J}{j} (N-m)^{J-j} = (N-m+1)^J - (N-m)^J.$$

In the last equation, the binomial theorem is applied. Then, we obtain

$$\frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S^N_{\min\{n_1,\dots,n_J\}/N}\right)^2 = \frac{1}{N^J} \sum_{m=1}^{N} |\mathcal{V}_J(m)| \left(S^N_{m/N}\right)^2$$

$$= \frac{1}{N^J} \sum_{m=1}^{N} \left((N-m+1)^J - (N-m)^J\right) \left(S^N_{m/N}\right)^2.$$

(b) **Case $L = J$:**

We consider the second part:

$$\int_{\left[\frac{1}{2},\frac{3}{2}\right)^J} \left(S_1^N - S_{V_J(\mathbf{t})}^N\right)^2 d\mathbf{t} = \frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S_1^N - S_{\max\{n_1,\dots,n_J\}/N}^N\right)^2.$$

Analogously, we simplify the sum by considering the cardinality of the set $\mathcal{W}_J(m)$ for $m \in \{1,\dots,N\}$:

$$\mathcal{W}_J(m) := \Big\{(n_1,\dots,n_J) \in \{1,\dots,N\}^J;\, \max\{n_1,\dots,n_J\} = m\Big\}.$$

The number of elements in $\mathcal{W}_J(m)$ can be counted by considering all J-tuples where exactly j entries are equal to m and the other $J-j$ entries are smaller than m:

$$|\mathcal{W}_J(m)| = \sum_{j=1}^{J} \binom{J}{j} (m-1)^{J-j} = m^J - (m-1)^J.$$

We obtain the result:

$$\begin{aligned}\frac{1}{N^J} \sum_{n_1,\dots,n_J=1}^{N} \left(S_1^N - S_{\max\{n_1,\dots,n_J\}/N}^N\right)^2 &= \frac{1}{N^J} \sum_{m=1}^{N} |\mathcal{W}_J(m)| \left(S_1^N - S_{m/N}^N\right)^2 \\ &= \frac{1}{N^J} \sum_{m=1}^{N} (m^J - (m-1)^J) \left(S_1^N - S_{m/N}^N\right)^2.\end{aligned}$$

(c) **Case $L \in \{1,\dots,J-1\}$:**

We consider the last and more complicated part:

$$\begin{aligned}&\int_{\left(-\frac{1}{2},\frac{1}{2}\right)^{J-L} \times \left[\frac{1}{2},\frac{3}{2}\right)^L} \prod_{\substack{i=1,\dots,J-L \\ j=J-L+1,\dots,J}} \mathbb{1}\{|t_j - t_i| < 1\} \left(S_{W_J(\mathbf{t})\wedge 1}^N - S_{V_J(\mathbf{t})\vee 0}^N\right)^2 d\mathbf{t} \\ =&\frac{1}{N^J} \sum_{\substack{n_1,\dots,n_J=1 \\ n_1,\dots,n_{J-L} \geq n_{J-L+1},\dots,n_J}}^{N} \left(S_{\min\{n_1,\dots,n_{J-L}\}/N}^N - S_{\max\{n_{J-L+1},\dots,n_J\}/N}^N\right)^2. \qquad (76)\end{aligned}$$

Now, we fix pairs (m_1, m_2) with $1 \leq m_1 \leq m_2 \leq N$ such that

$$\min\{n_1,\dots,n_{J-L}\} = m_2 \text{ and } \max\{n_{J-L+1},\dots,n_J\} = m_1.$$

In order to count all of possible J-tuples with these values for fixed (m_1, m_2), we

can consider the cartesian product of the sets

$$\mathcal{V}_{J-L}(m_2) := \left\{(n_1,\dots,n_{J-L}) \in \{1,\dots,N\}^{J-L}; \min\{n_1,\dots,n_{J-L}\} = m_2\right\}$$
$$\mathcal{W}_J^{J-L+1}(m_1) := \left\{(n_{J-L+1},\dots,n_J) \in \{1,\dots,N\}^{L}; \max\{n_1,\dots,n_J\} = m_1\right\}$$

which are projections of $\mathcal{V}_J(m_2)$ and $\mathcal{W}_J(m_1)$ from the previous cases in $L = 0$ and $L = J$ to the first $J-L$ or last L coordinates, respectively. In particular, $\mathcal{V}_{J-L}(m_2)$ contains $(J-L)$-tuples and $\mathcal{W}_J^{J-L+1}(m_1)$ contains L-tuples such that their cartesian product contains J-tuples. For the cardinality of the cartesian product:

$$\begin{aligned}&|\mathcal{V}_{J-L}(m_2) \times \mathcal{W}_J^{J-L+1}(m_1)| = |\mathcal{V}_{J-L}(m_2)| \cdot |\mathcal{W}_J^{J-L+1}(m_1)| \\ =&((N-m_2+1)^J - (N-m_2)^J)(m_1^L - (m_1-1)^L).\end{aligned}$$

Then, we can simplify the sum in Formula (76) by:

$$\begin{aligned}&\frac{1}{N^J} \sum_{\substack{n_1,\dots,n_J=1 \\ n_1,\dots,n_{J-L} \geq n_{J-L+1},\dots,n_J}}^{N} \left(S^N_{\min\{n_1,\dots,n_{J-L}\}/N} - S^N_{\max\{n_{J-L+1},\dots,n_J\}/N}\right)^2 \\ =&\frac{1}{N^J} \sum_{m_2=1}^{N} \sum_{m_1=1}^{m_2} |\mathcal{V}_{J-L}(m_2)| \cdot |\mathcal{W}_J^{J-L+1}(m_1)| \cdot \left(S^N_{m_2/N} - S^N_{m_1/N}\right)^2.\end{aligned}$$

We split $\left(S^N_{\frac{m_2}{N}} - S^N_{\frac{m_1}{N}}\right)^2$ into the three parts

$$\left(S^N_{m_2/N} - S^N_{m_1/N}\right)^2 = \left(S^N_{m_2/N}\right)^2 + \left(S^N_{m_1/N}\right)^2 - 2S^N_{m_2/N}S^N_{m_1/N} \tag{77}$$

and consider these parts separately:

$\underline{\left(S^N_{m_2/N}\right)^2}$:

$$\begin{aligned}&\sum_{m_2=1}^{N} \sum_{m_1=1}^{m_2} |\mathcal{V}_{J-L}(m_2)| \cdot |\mathcal{W}_J^{J-L+1}(m_1)| \cdot \left(S^N_{m_2/N}\right)^2 \\ =&\sum_{m_2=1}^{N} \sum_{m_1=1}^{m_2} \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right)\left(m_1^L - (m_1-1)^L\right)\left(S_{m_2/N}\right)^2 \\ =&\sum_{m_2=1}^{N} m_2^L \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right)\left(S_{m_2/N}\right)^2.\end{aligned}$$

$\underline{\left(S^N_{m_1/N}\right)^2}$:

$$\begin{aligned}
&\sum_{m_2=1}^{N}\sum_{m_1=1}^{m_2} |\mathcal{V}_{J-L}(m_2)| \cdot |\mathcal{W}_J^{J-L+1}(m_1)| \cdot \left(S^N_{m_1/N}\right)^2 \\
=&\sum_{m_1=1}^{N}\sum_{m_2=m_1}^{N} \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right)\left(m_1^L - (m_1-1)^L\right)\left(S^N_{m_1/N}\right)^2 \\
=&\sum_{m_1=1}^{N} (N-m_1+1)^{J-L}(m_1^L - (m_1-1)^L)\left(S^N_{m_1/N}\right)^2.
\end{aligned}$$

$\underline{-2S^N_{m_2/N}S^N_{m_1/N}}$:

$$\begin{aligned}
&-2\sum_{m_2=1}^{N}\sum_{m_1=1}^{m_2} \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right)\left(m_1^L - (m_1-1)^L\right) S^N_{m_2/N}S^N_{m_1/N} \\
=&-2\sum_{m_2=1}^{N} \left((N-m_2+1)^{J-L} - (N-m_2)^{J-L}\right) S^N_{m_2/N}\tilde{S}^L(m_2).
\end{aligned}$$

Inserting all three parts into Formula (77) yields (c). □

For the 4-depth, this representation can be further simplified such that this storage method is not necessary (Malcherczyk, 2018a, p. 65-71). In particular, we consider the third part of Formula (77) from the previous proof for $K = 4$, $J = 2$ and $L = 1$:

$$\begin{aligned}
&\sum_{m_2=1}^{N}\sum_{m_1=1}^{m_2} \left((N-m_2+1) - (N-m_2)\right)\left(m_1 - (m_1-1)\right) S^N_{m_2/N}S^N_{m_1/N} \\
=&\sum_{m_2=1}^{N}\sum_{m_1=1}^{m_2} S^N_{m_2/N}S^N_{m_1/N} = \frac{1}{2}\sum_{m_2=1}^{N}\sum_{m_1=1}^{N} S^N_{m_2/N}S^N_{m_1/N} + \frac{1}{2}\sum_{m=1}^{N}\left(S^N_{m/N}\right)^2 \\
=&\frac{1}{2}\left(\sum_{m=1}^{N} S^N_{m/N}\right)^2 + \frac{1}{2}\sum_{m=1}^{N}\left(S^N_{m/N}\right)^2
\end{aligned}$$

which can also be computed in linear time in N and only requires storing $S^N_{m/N}$ for $m \in \{1, \ldots, N\}$.

Computation of the finite sample quantiles

The finite sample quantiles $q^K_{\alpha,N}$ can be computed by observing all 2^N combinations for residuals with elements in $\{1, -1\}$ since the signs are only relevant for the resulting value of the K-depth. However, this method is not recommended for large

sample sizes, since the number of possible elements 2^N increases exponentially. In the package `GSignTest` Horn (2021a), the finite sample quantiles $q^K_{\alpha,N}$ are computed for $N \in \{K, \ldots, 25\}$ and $K \in \{3, 4, 5\}$. For larger N, the finite sample quantiles are approximated by 1,000,000 of repeatedly computed realizations of the rescaled K-depth T_K under the Assumption 2.1. For this thesis, the finite sample are computed by 100,000 repetitions of the K-depth under Assumption 2.1. We always use finite sample quantiles if not mentioned otherwise.
The finite sample quantiles have the drawback that they can only be computed exactly for small sample sizes and that we have to compute them for each N compared to the asymptotic quantiles. On the other hand, the finite sample quantiles always provides the α-level while the asymptotic quantiles only approximately provide this property (see also the proof of Theorem 6.1).

6.3 Type-I-error and required sample sizes

If we apply the K-depth test with asymptotic quantiles, we should check how large the sample size N has to be for a particularly given type-I-error. The type-I-error is the probability that we reject the null hypothesis H_0 although it is actually true. In Chapter 6.3.1, the type-I-errors will be computed by simulations for various K. We consider the asymptotic point-hypotheses test in Theorem 6.1 (a), p. 97, for commonly used levels $\alpha \in \{0.1, 0.05, 0.01\}$. In Chapter 6.3.2, we will also discuss the problem that the null hypothesis cannot be rejected for any residuals if the sample size N is too small due to the discrete distribution of the K-depth.

6.3.1 Type-I-error simulations

At first, we give an overview of the following simulation study for computing the type-I-errors. For $N \in \{10, 15, \ldots, 200\}$, $K \in \{3, 4, \ldots, 11\}$ and $\alpha \in \{0.1, 0.05, 0.01\}$, we check 1,000,000 times if H_0 is rejected under errors $E_1, \ldots, E_N$ satisfying Assumption 2.1 applying the test in Theorem 6.1 with the asymptotic quantiles. The case $K = 2$ is not considered here since the 2-depth is equivalent to counting the number of positive signs (Leckey et al., 2020). Due to the law of large numbers, the relative number of rejections approximates the type-I-error. These type-I-errors are presented in Figure 19, 20 and 21 for respective $\alpha \in \{0.1, 0.05, 0.01\}$. The red dashed line represents the benchmark of the different levels $\alpha \in \{0.1, 0.05, 0.01\}$.
In Figure 19, the case $\alpha = 0.1$ is considered. The type-I-errors converge from above to the desired level so that the tests are liberal for too small N. Note also that for too small N, H_0 can never be rejected. Therefore, the curves for $K \geq 7$ start at zero and

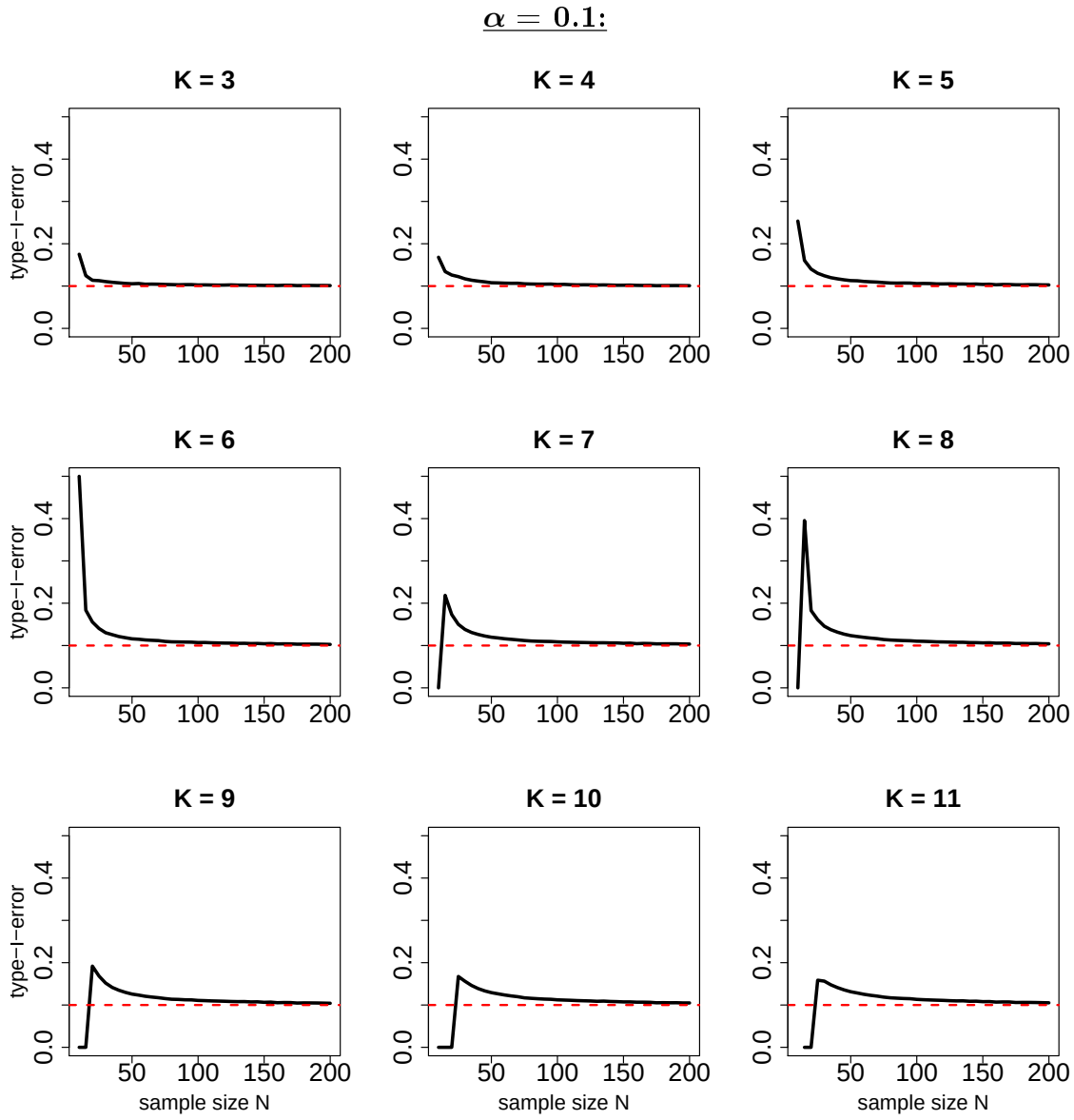

Figure 19: Simulated type-I-errors of the asymptotic K-depth test for $\alpha = 0.1$

jump to a value over the dashed line. It is noticeable that for $K \in \{6, 8\}$, the type-I-errors are very high for small sample sizes compared to the other K. For $K \in \{5, 7\}$, we can also recognize higher results but not that extreme. For increasing K, the type-I-errors are not that high for small N, but rejections are only for higher N possible. For approximately $N \geq 100$, we have acceptable type-I-errors for general K. For $K \in \{3, 4\}$, we could have also for smaller N, e.g., $N \geq 50$, acceptable type-I-errors. Otherwise, the finite sample quantiles should be used instead.
The case $\alpha = 0.05$ is presented in Figure 20. This case yields similar results to $\alpha = 0.1$, but the type-I-errors are relatively smaller compared to the other case. We

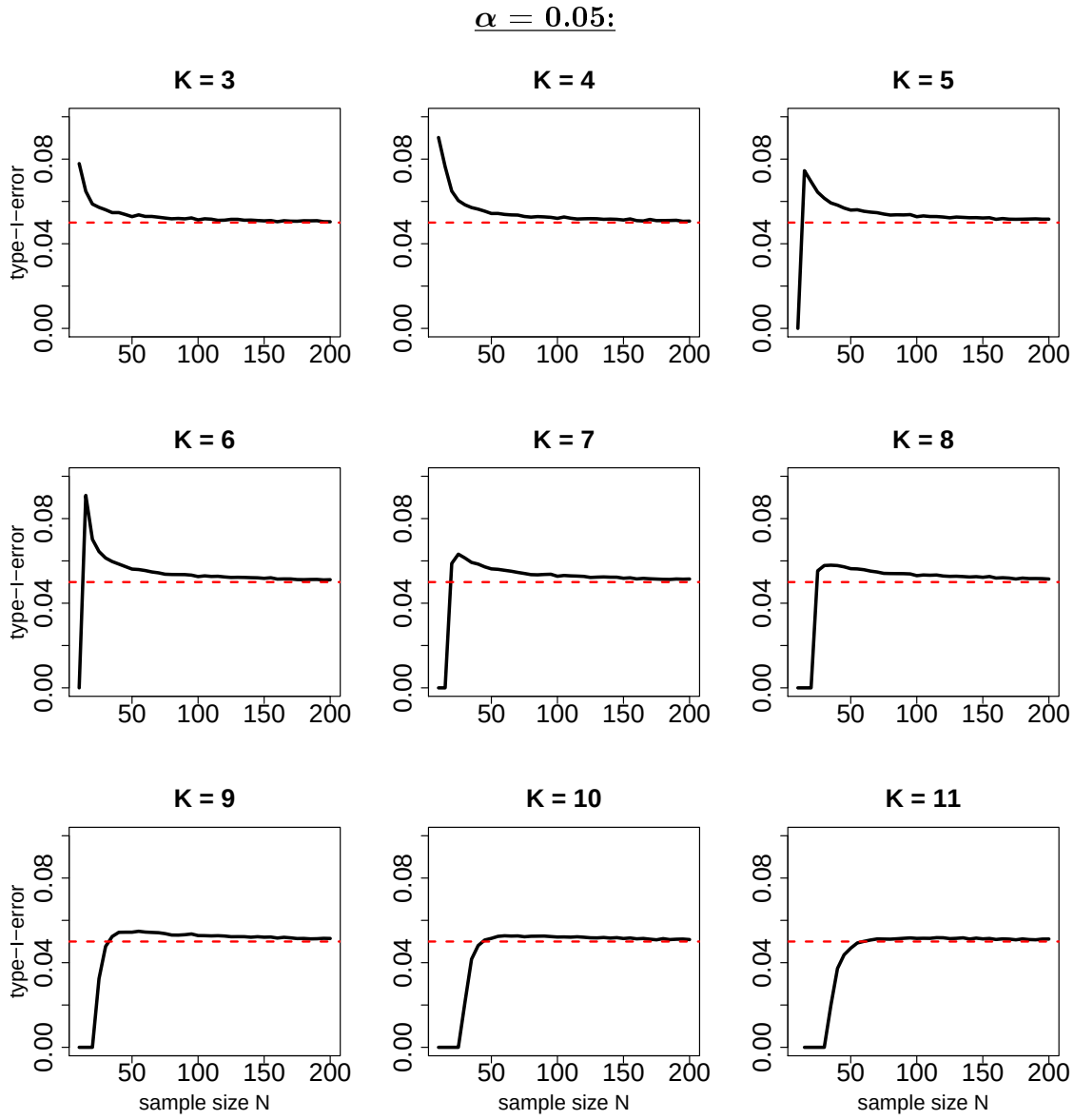

Figure 20: Simulated type-I-errors of the asymptotic K-depth test for $\alpha = 0.05$

can further notice that the necessary sample size N for possible rejections of H_0 increases for decreasing α.

The situation for $\alpha = 0.01$ in Figure 21 shows convergence from below for $K \geq 4$, in contrast to the previous scenarios. Especially, the tests are conservative. We suggest $N \geq 100$ for higher K and also $N \geq 50$ for smaller K as a proper choice. However, a clear choice of N for all K is not possible and has to be concluded in advance.

The changing behavior of the convergence from above and below of the type-I-errors can be explained by a comparison of the exact and asymptotic α-quantiles for several α, e.g., by comparing the distribution functions. If a finite sample quantile

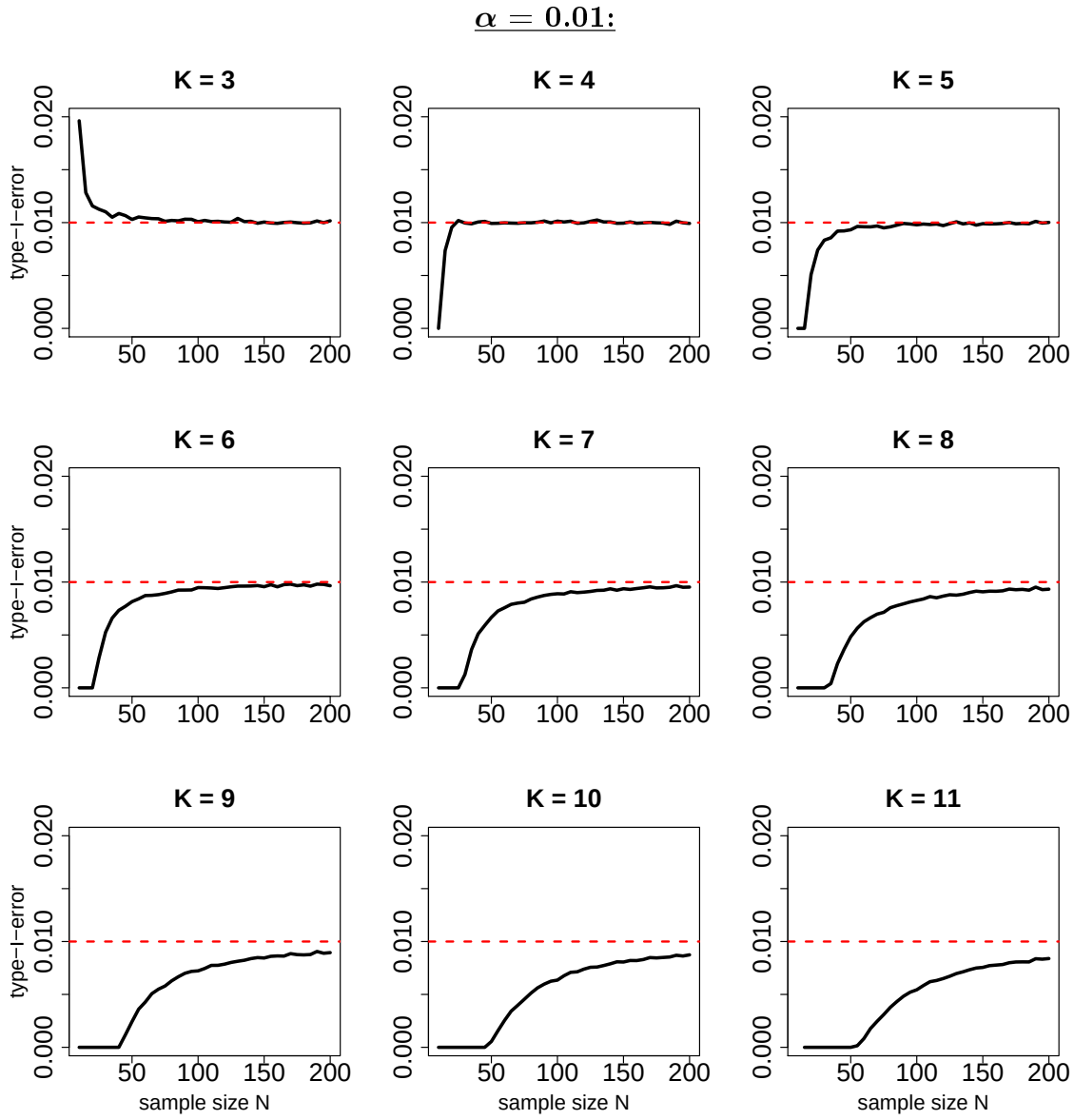

Figure 21: Simulated type-I-errors of the asymptotic K-depth test for $\alpha = 0.01$

converges from below to the asymptotic quantile, then the type-I-errors will converge from above so that the test will be liberal for small N and vice versa.
If the type-I-errors are too high, we should consider finite sample quantiles which will lead to tests with type-I-errors equal or less than a given α-level. However, testing with the finite sample quantiles can also fail if the sample sizes are too small since H_0 can never be rejected then. This problem will be discussed in the next subsection.

6.3.2 Required sample sizes

We will now focus on the problem that for too small sample sizes, the K-depth test never rejects H_0. The next lemma gives a condition when the K-depth test is able to reject the null hypotheses for given α-level and K for the test in Theorem 6.1 (a).

Lemma 6.4. *Let $\alpha \in (0,1)$ and $K \in \mathbb{N} \setminus \{1\}$ and $N \in \mathbb{N}$. Only if the condition*

$$-\frac{N}{2^{K-1}} < q^K_{\alpha,N}, \tag{78}$$

is satisfied, the K-depth test in Theorem 6.1 (a), p. 97, can reject the null hypothesis under some residuals. Formula (78) is satisfied if

$$p^N_K \leq \alpha \text{ for } p^N_K = \mathbb{P}(X_N \leq K-2) \text{ with } X_N \sim Bin(N-1, \tfrac{1}{2}).$$

Proof of Lemma 6.4: The left side of Formula (78) is the minimum of the rescaled K-depth T_K for sample size N. If the α-quantile coincides with its minimum, the condition for rejecting the null hypothesis can never be fulfilled.
According to Lemma 5.6, p. 79, the K-depth is minimal for residuals with $K-1$ or less blocks. Therefore, the probability that the K-depth is minimal for residuals under Assumption 2.1 is given by

$$p^N_K = \mathbb{P}(X_N \leq K-2) \text{ with } X_N \sim Bin(N-1, \tfrac{1}{2})$$

since $X_N + 1$ describes the number of blocks and $p^N_K = \mathbb{P}(X_N + 1 \leq K-1)$ is the probability that $K-1$ or less blocks occur. For given K and α, the smallest sample size N which satisfies Formula (78) can be obtained by computing the smallest N such that $p^N_K \leq \alpha$. □

The problem that an α-quantile coincides with the minimum of the distribution occurs for most tests based on discrete statistics, because the statistics have only a few number of possible values. E.g., the Wilcoxon rank-sum test cannot reject the null hypothesis for too small sample sizes either (Corder and Foreman, 2009, p. 228). When N increases and α is not too small, there will be more possible values for the statistics such that the α-quantiles starts to differ from the minimal value.
Table 4 shows the results of the required sample sizes for $\alpha \in \{0.1, 0.05, 0.01\}$. E.g. for $K = 7$ and $\alpha = 0.05$, a sample size of $N \geq 19$ satisfies the necessary condition. Note that the cut off points in Table 4 do not coincide exactly with cut off points in Figure 19, 20 and 21, because the exact distribution is considered here instead of the asymptotic distribution. If we want to apply the K-depth for large K, we have

α \ K	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.1	8	10	13	15	18	20	22	25	27	29	31	34	36	38	40
0.05	9	12	14	17	19	22	24	27	29	32	34	36	39	41	43
0.01	12	15	18	20	23	26	28	31	34	36	39	41	44	46	48

Table 4: Required sample sizes for given K and α

to consider the needed sample size in the planning of an experiment in best case. According to Figure 22, we conclude that the necessary sample sizes can be linearly

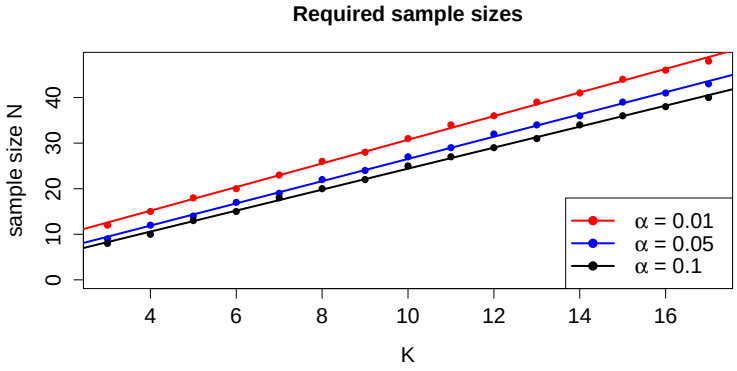

Figure 22: Required sample sizes for several α and K

extrapolated for higher K by the following equations obtained by regression lines:

$$\text{for } \alpha = 0.01 : N_{0.01}(K) = 4.8 + 2.6 \cdot K,$$
$$\text{for } \alpha = 0.05 : N_{0.05}(K) = 2.1 + 2.4 \cdot K,$$
$$\text{for } \alpha = 0.1 : N_{0.1}(K) = 1.3 + 2.3 \cdot K.$$

Nevertheless, these regression lines are only approximate rules and should always be confirmed by the procedure in Lemma 6.4.

Required sample sizes for the two-sided test

The previous procedure only considers the one-sided test from Theorem 6.1, but we also want to consider the two-sided test in Theorem 6.2. For $\alpha \in (0, 1)$, we usually have the quantiles $q^K_{\frac{\alpha}{2},N}$ and $q^K_{1-\frac{\alpha}{2},N}$ as the lower and upper critical value. The same procedure as before can be done for finding the required sample sizes that exceeding is possible. However, the required sample size for deceeding the lower value or exceeding the upper value can be different. Therefore, we will analyze both situations. For finding the necessary sample size to exceed the upper critical value, we have to check if the upper critical value is equal to the maximal K-depth. We

suppose that the K-depth is maximal under alternating signs, but note that we have not already proven this assertion, cf. Conjecture 5.14 and Lemma 5.15. Therefore, we only provide the following conjecture for finding the required sample sizes.

Conjecture 6.5. *Let $\alpha \in (0,1)$ and $K \in \mathbb{N} \setminus \{1\}$ and $N \in \mathbb{N}$. Let $\mathbf{r} \in \mathbb{R}^N$ be a residual vector with alternating signs. Only if the condition*

$$q^K_{1-\frac{\alpha}{2},N} < T_K(\mathbf{r})$$

is satisfied, the test statistic from the K-depth test in Theorem 6.2 can exceed the upper critical value.

In Table 5, the required sample sizes from the procedure in Lemma 6.4 are listed such that deceeding the lower critical value can be possible. Table 6 and 7 present the required sample sizes for the possibility to exceed the upper critical value. The simulations show here that a distinction between even and odd N is necessary. Therefore, the first even sample sizes for the possibility to exceed the upper critical value are given in Table 6 and the first odd sample sizes are given in Table 7. According to Conjecture 5.14, this distinction may not be surprising since the solution sets for the maximal K-depth are also different for even or odd number of blocks. It is noticeable that the required sample sizes in Table 6 and

$\frac{\alpha}{2}$ \ K	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.05	9	12	14	17	19	22	24	27	29	32	34	36	39	41	43
0.025	10	13	16	18	21	23	26	29	31	33	36	38	41	43	45
0.005	13	16	19	22	25	27	30	33	35	38	40	43	45	48	50

Table 5: Required sample sizes for the two-sided K-depth test (possibility to deceed the lower critical value $q^K_{\frac{\alpha}{2},N}$) for given K and $\frac{\alpha}{2}$

$1-\frac{\alpha}{2}$ \ K	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.95	12	6	10	6	10	8	10	10	12	12	14	14	16	16	18
0.975	14	8	10	8	10	8	10	10	12	12	14	14	16	16	18
0.995	18	10	14	10	14	10	14	10	14	12	14	14	16	16	18

Table 6: Required sample sizes (even N) for the two-sided K-depth test (possibility to exceed the upper critical value $q^K_{1-\frac{\alpha}{2},N}$) for given K and $1-\frac{\alpha}{2}$

7 are smaller than in Table 5. For $K+N$ even, we can conclude the rule that the upper critical value $q^K_{1-\frac{\alpha}{2}}$ can be exceeded if the condition $\frac{1}{2^{N-1}} < \frac{\alpha}{2}$ holds. Note that $\mathbb{P}((X_1,\ldots,X_N) \text{ has alternating signs}) = \frac{1}{2^{N-1}}$ for i.i.d. random variables

$1-\frac{\alpha}{2}$ \ K	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.95	7	9	7	9	7	9	9	11	11	13	13	15	15	17	17
0.975	7	11	7	11	7	11	9	11	11	13	13	15	15	17	17
0.995	9	13	9	13	9	13	9	13	11	13	13	15	15	17	17

Table 7: Required sample sizes (odd N) for the two-sided K-depth test (possibility to exceed the upper critical value $q^K_{1-\frac{\alpha}{2},N}$) for given K and $1-\frac{\alpha}{2}$

$X_1, \ldots, X_N \sim Bin(1, \frac{1}{2})$ which is the supposed probability to have maximal K-depth under H_0 according to Conjecture 5.14 (a) and Lemma 5.15. Thus, fulfilling this condition leads to upper $(1-\frac{\alpha}{2})$-quantiles which are smaller than the maximal K-depth. For $K+N$ odd, similar explanations may be concluded based on Conjecture 5.14 (b). Since this case is more complicated, we leave out a detailed explanation. The regressions lines for the required sample sizes in Table 5 are given by:

$$\text{for } \frac{\alpha}{2} = 0.005 : N_{0.005}(K) = 6 + 2.6 \cdot K,$$
$$\text{for } \frac{\alpha}{2} = 0.025 : N_{0.025}(K) = 3.2 + 2.5 \cdot K.$$

In Figure 23, the corresponding required sample sizes for exceeding the upper critical value are given. The distinction between the even and odd sample sizes is visualized

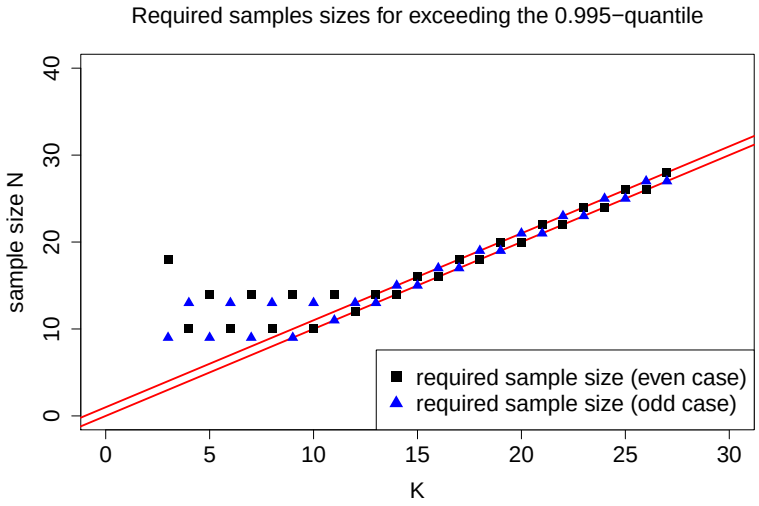

Figure 23: Required sample sizes for several α and K

by plotting squares and triangles. The plot shows the required sample sizes for $K \leq 27$. For smaller K, the required sample sizes are alternating between the even and odd case. Up to $K \geq 12$, the required samples are well located on the lines $N(K) = K$ (lower line) or $N(K) = 1 + K$ (upper line).

6.4 Choice of the hyper-parameter K

In this chapter, we will test the fit of parameters in several polynomial models and compute the type-II-errors by simulations. The general model equation is

$$Y_n = g(X_n, \boldsymbol{\theta}^*) + E_n, n = 1, \dots N$$

with the true parameter $\boldsymbol{\theta}^*$, cf. Chapter 2.1. Point-hypotheses of the form:

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}_0\}, H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$$

will be tested in this chapter. One central aim is to answer these questions:

- *How should we choose the hyper-parameter K for a particular model class?*
- *Can we conclude rules for the power functions in dependence of K?*
- *Are there mathematical explanations for the answers of these questions?*

The answers of the first two questions will depend on the sign structure of the difference function between the correct model and a suggested model denoted by $\boldsymbol{\theta}$:

$$D(x, \boldsymbol{\theta}^*, \boldsymbol{\theta}) = g(x, \boldsymbol{\theta}^*) - g(x, \boldsymbol{\theta}) \text{ for arbitrary explanatory variables } x \in \mathbb{R}.$$

The sign structure of the residuals under alternatives depends strongly on the function D since it contains information about areas of the true model that cannot be explained by the alternative parameter. In other words, the function D corresponds to the residuals omitting the errors. Thus, if D has a high number of sign changes, then the residual vector will also have a great chance for a high number of sign changes. In this case, the alternative under the parameter $\boldsymbol{\theta}$ will be harder to reject since the K-depth will be higher according to the heuristics from Conjecture 5.14, p. 93. Furthermore, Lemma 5.6, p. 79, leads to the idea that the K-depth is small if we expect $K-2$ or less sign changes in D. To summarize these ideas, we conclude:

The K-depth test may have a high power for an alternative indicated by $\boldsymbol{\theta}$ if the difference function $D(\bullet, \boldsymbol{\theta}^*, \boldsymbol{\theta})$ has $K-2$ or less sign changes.

The simulation study in Chapter 6.4.1 will consider models with alternatives that have a fixed number of sign changes denoted by Q. In order to answer the upper questions, we will vary the number of sign changes in different models and conclude how K should be chosen. Considering polynomial regression models has the advantage that we can construct easily alternatives with fixed number of sign changes.

More complex continuous model functions can often be approximated locally by polynomial models. Therefore, our results can be transferred for other models in certain ways. In Chapter 6.4.2, the results and the impact for the practice will be discussed. The choice for the hyper-parameter K will also be concluded.

6.4.1 Simulation study with fixed numbers of sign changes

In the following simulation study, we fix subclasses of polynomial regression functions in order to understand the influence of the number of sign changes to the power of the tests better. We consider polynomial models such that the difference functions have Q sign changes with $Q \in \{1, 2, 3, 4, 5\}$. The alternatives are produced by scaled versions of the true polynomial models or by alternatives with a reversed sign and shifted then.

The explanatory variables are equidistant in the fixed interval $[-3, 3]$ for samples sizes $N \in \{25, 50, 100, 200\}$. The situation of expanding intervals of the explanatory variables for increasing N will not be considered since polynomial functions are strictly monotonous for sufficiently high input values. This will lead to uninteresting, constant sign structures for sufficiently high explanatory variables and the K-depth will be rejected easily for several alternatives.

Moreover, we have independent Cauchy distributed errors $E_1, \ldots, E_N \sim Cau(0, 1)$ to express the outlier robustness of the methods. Note that under other symmetric distributions, such as the normal distribution, the results are similar. The simulations of the power are 1,000 times repeated for each alternative based on the test in Theorem 6.1 (a) using the finite sample quantiles. The parameter areas vary in several cases but are always discretized in steps of length 0.1. We consider $K \in \{3, \ldots, 10\}$, but for $N = 25$, we omit the case $K = 10$ due to the small sample size and the impossible rejection, cf. Table 4, p. 112, in Chapter 6.3.2. The 2-depth test is omitted due to its bad performance. Simulations based on the 2-depth test can be found in Leckey et al. (2020).

Furthermore, Table 8, p. 117, lists all considered models. The models with a variation of scaling are indexed by „a“ and the models with variation of the intercept are indexed by „b“. The true parameters are denoted by $\boldsymbol{\theta}^*$ and $\boldsymbol{\Theta}_1$ describes the set of the alternatives. We choose here polynomials of degree Q since they can at most have Q intersections with another polynomial of the same degree. The number of intersections corresponds to the number of sign changes in the difference function. An illustration of the model situations are given combined with the results of the power functions for a better overview.

model	Q	model function	$\boldsymbol{\theta}^*$	$\boldsymbol{\Theta}_1$
I	1	$g(x,\theta) = \theta x$	1	$[-4,4] \setminus \{1\}$
IIa	2	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x^2$	$(-1,1)^\top$	$\{(-\theta,\theta)^\top; \theta \in [0,4], \theta \neq 1\}$
IIb	2	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x^2$	$(-1,1)^\top$	$[1,7] \times \{-1\}$
IIIa	3	$g(x,\boldsymbol{\theta}) = \theta_0 x + \theta_1 x^3$	$(-2,1)^\top$	$\{(-2\theta,\theta)^\top; \theta \in [0,4], \theta \neq 1\}$
IIIb	3	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^3$	$(0,-2,1)^\top$	$[-3,3] \times \{(2,-1)^\top\}$
IVa	4	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x^2 + \theta_2 x^4$	$\left(\frac{1}{2},-2,1\right)^\top$	$\{\left(\frac{\theta}{2},-2\theta,\theta\right)^\top; \theta \in [0,4], \theta \neq 1\}$
IVb	4	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x^2 + \theta_2 x^4$	$(\frac{1}{2},-2,1)^\top$	$[-3,3] \times \{(2,-1)^\top\}$
Va	5	$g(x,\boldsymbol{\theta}) = \theta_0 x + \theta_1 x^3 + \theta_2 x^5$	$(2,-3,\frac{3}{4})^\top$	$\{(2\theta,-3\theta,\frac{3}{4}\theta)^\top; \theta \in [0,6], \theta \neq 1\}$
Vb	5	$g(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^3 + \theta_3 x^5$	$(0,2,-3,\frac{3}{4})^\top$	$[-3,3] \times \{(-2,3,\frac{3}{4})^\top\}$

Table 8: Overview of the models having alternatives with Q sign changes in the difference functions

Results of the simulation study

1-sign change model (I)

In order to have only one sign change, several linear functions are considered for model (I), cf. Figure 24. The model equation is given by $g(x, \theta) = \theta x$, $\theta \in [-4, 4]$.

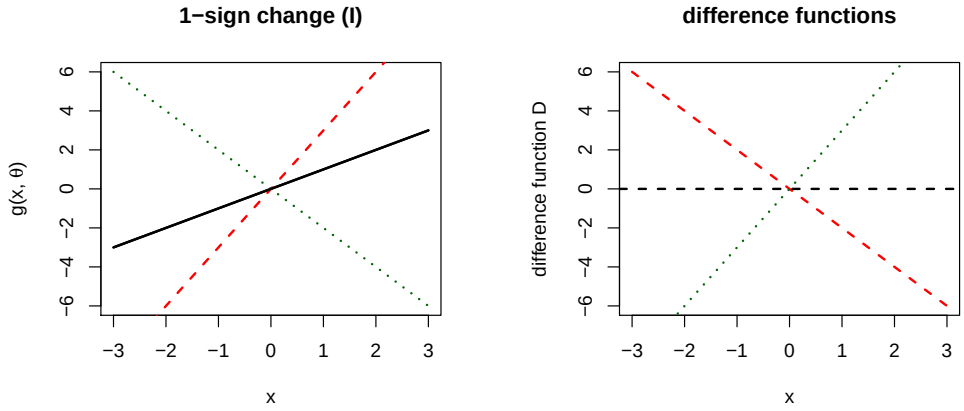

Figure 24: Illustration of the model class I with true parameter $\theta^* = 1$ (black), alternatives $\theta = 3$ (red) and $\theta = -2$ (green)

Considering one parameter and determining the intercept to zero has the advantage that the simulations run faster and the visualization of the one-dimensional parameter space for the power functions are more convenient. On the right side of this figure, we see the difference functions for the two examples $\theta \in \{3, -2\}$ from the alternative space $\mathbf{\Theta}_1 = [-4, 4] \setminus \{1\}$. In Table 9, the mean of the computed power

N	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$	$K = 8$	$K = 9$	$K = 10$
25	0.86	0.73	0.84	0.75	0.82	0.74	0.79	0.00
50	0.93	0.87	0.92	0.88	0.91	0.88	0.91	0.88
100	0.96	0.93	0.95	0.94	0.95	0.94	0.95	0.93
200	0.98	0.96	0.97	0.96	0.97	0.96	0.97	0.96

Table 9: Mean of the simulated power for the 1-sign change alternatives (I)

of all alternatives in $\mathbf{\Theta}_1$ are given. Figure 25 contains the computed power for each alternative parameter. We can conclude that the K-depth has slightly higher power for odd K. These differences are especially noticeable for $N = 25$. Moreover, increasing the sample size N can compensate the power loss of even K. For $N \geq 50$, the differences between even and odd K exist but they become less relevant. In particular, the case $K = 3$ delivers the best result, cf. Table 9. Further, the mean power decreases for increasing odd K and does not vary for different even K.

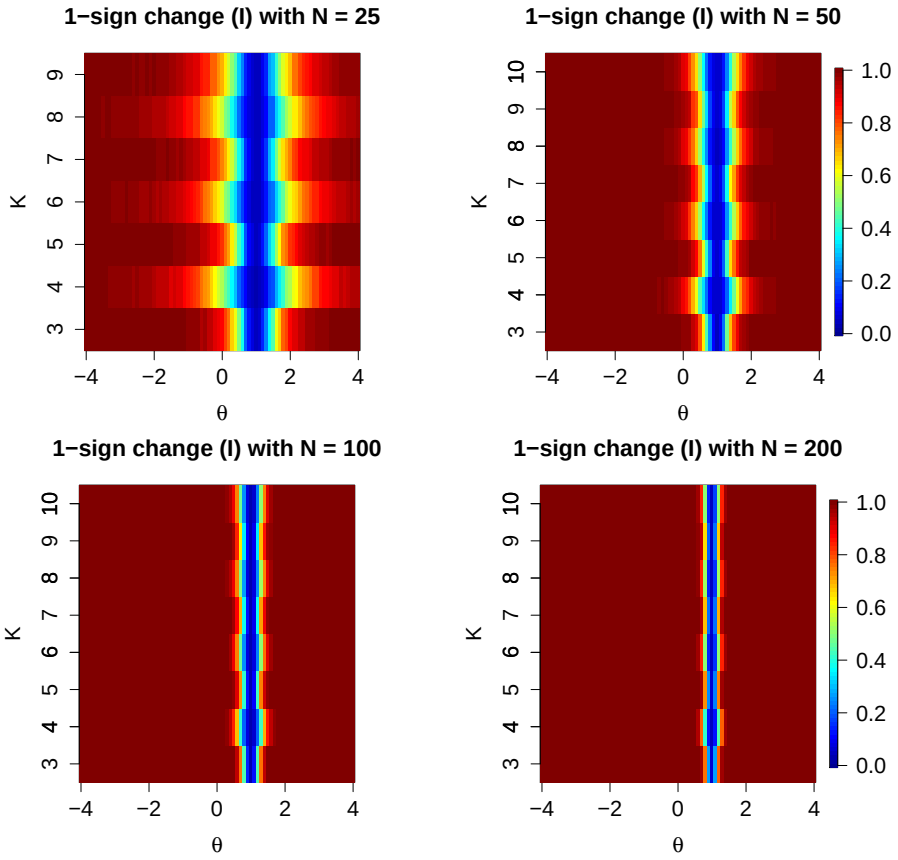

Figure 25: Power functions for the 1-sign change alternatives (I)

2-sign changes model (IIa)

In Figure 26, the black line shows the true model $g(x, \boldsymbol{\theta}^*) = x^2 - 1$. The parameters

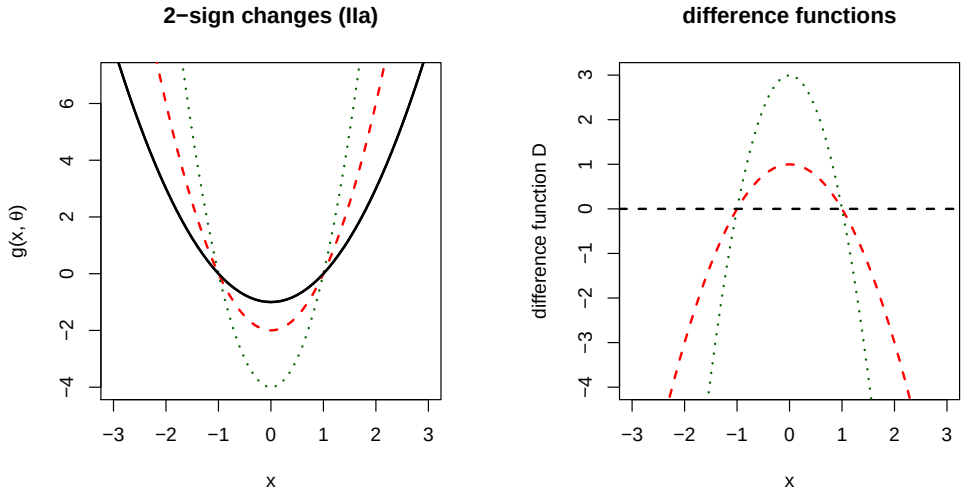

Figure 26: Illustration of the model class IIa with true parameter $\boldsymbol{\theta}^* = (-1, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (-2, 2)^\top$ (red) and $\boldsymbol{\theta} = (-4, 4)^\top$ (green)

in $\boldsymbol{\Theta}_1 = \{(-\theta, \theta)^\top; \theta \in [0,4], \theta \neq 1\}$ describe scaled versions of this quadratic function by $g(x, \boldsymbol{\theta}) = \theta \cdot g(x, \boldsymbol{\theta}^*)$ for $\theta \in [0,4] \setminus \{1\}$. This scaling leads to compressed or stretched parabolas. In all cases, the difference function has exactly two sign changes. The power functions in Figure 27 show that the K-depths perform the best for even K. Especially for $N = 25$, this effect is noticeable. The case $K = 4$ is slightly the best, cf. Table 10. For odd $K \geq 5$, the power increases strongly for

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.03	0.79	0.53	0.78	0.59	0.75	0.59	0.00
50	0.07	0.91	0.78	0.90	0.82	0.90	0.83	0.89
100	0.46	0.96	0.90	0.96	0.91	0.95	0.92	0.95
200	0.82	0.98	0.95	0.98	0.96	0.98	0.96	0.98

Table 10: Mean of the simulated power for 2-sign change alternatives (IIa)

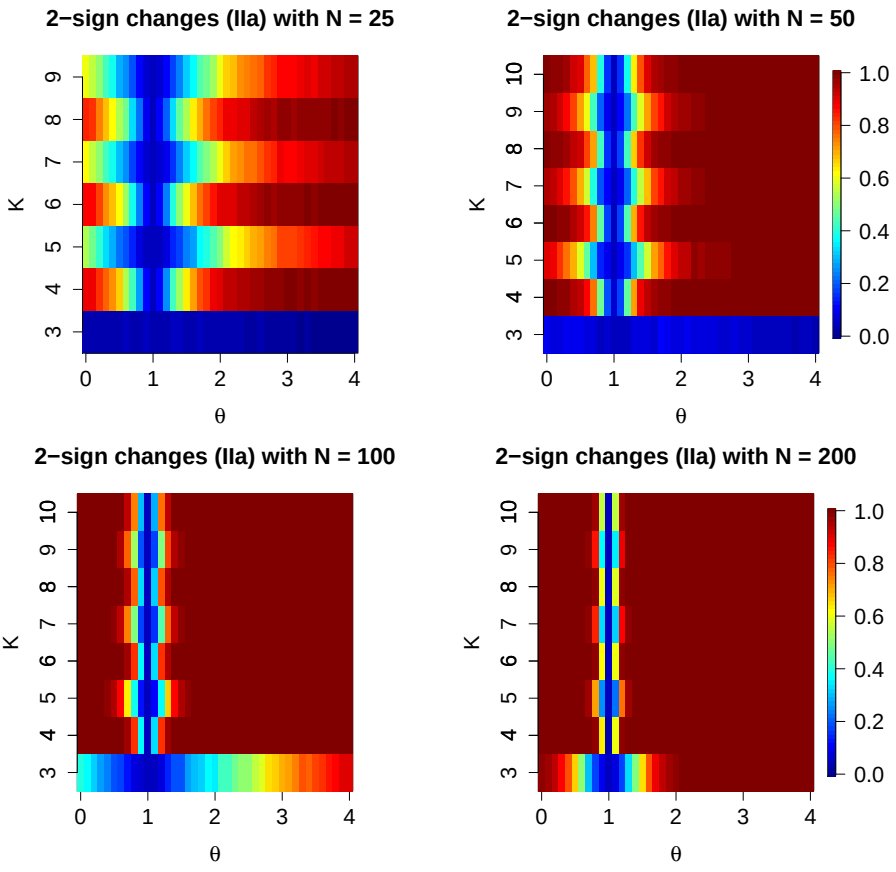

Figure 27: Power functions for the 2-sign changes alternatives (IIa)

increasing N and the differences of the power are not that high compared to even K. The case $K = 3$ performs still badly. Even for $N = 200$, the power function is noticeably worse than for the other cases. The bad performance for $K = 3$ is not

surprising based on the conjecture that the K-depth does not perform well if the number of sign changes in D is higher than $K - 2$.

2-sign changes model (IIb)

Figure 28 shows the alternatives with two sign changes by shifts of the true model with negative sign. In particular, the true model is here also $g(x, \boldsymbol{\theta}^*) = x^2 - 1$ and

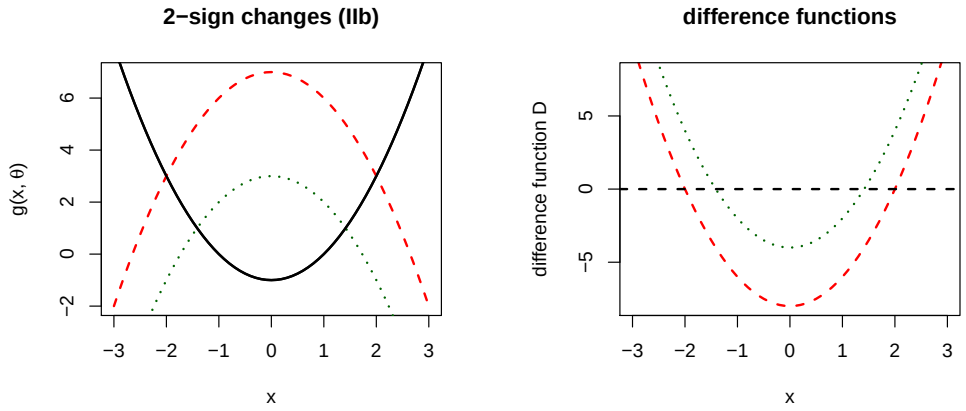

Figure 28: Illustration of the model class IIb with true parameter $\boldsymbol{\theta}^* = (-1, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (7, -1)^\top$ (red) and $\boldsymbol{\theta} = (3, -1)^\top$ (green)

the alternative have the form $g(x, \boldsymbol{\theta}) = -x^2 + 1 + \theta$ for $\theta \in [1, 7]$. Note that we do not obtain the true parameter (situation of H_0) for any θ in Figure 29 so that the power functions should have a power of 1 for all θ. We see similar results as

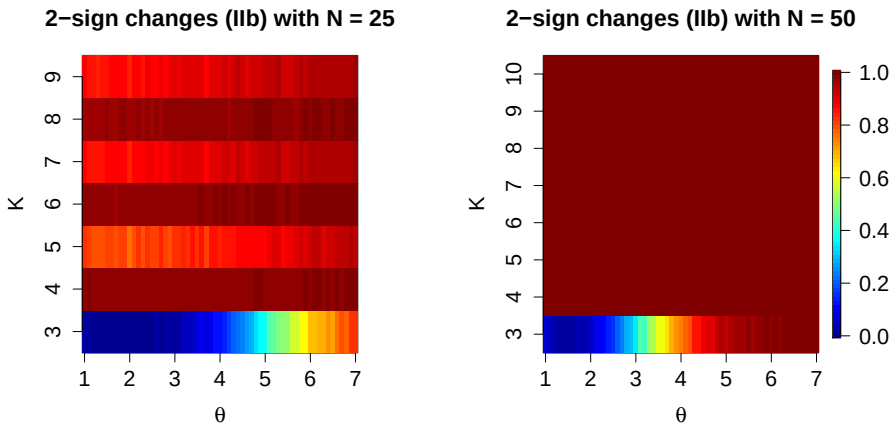

Figure 29: Power functions for the 2-sign change alternatives (IIb)

in Figure 27, especially that odd K perform worse than even K. Nevertheless, the power functions are clearly higher here for smaller N which can be caused by the fact that the functions are easier to distinguish based on their sign structure. Therefore,

the figures for $N \in \{100, 200\}$ are not shown since the power function is almost at one everywhere, cf. Table 11.

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.27	0.98	0.85	0.98	0.90	0.97	0.90	0.00
50	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 11: Mean of the simulated power for 2-sign change alternatives (IIb)

3-sign changes model (IIIa)

Figure 30 shows the 3-sign changes alternatives which are rescaled versions of the true model $g(x, \boldsymbol{\theta}^*) = x^3 - 2x$. In particular, the alternatives are analogously to model (IIa) generated by multiplying this model equation with $\theta \in [0, 4] \setminus \{1\}$. In

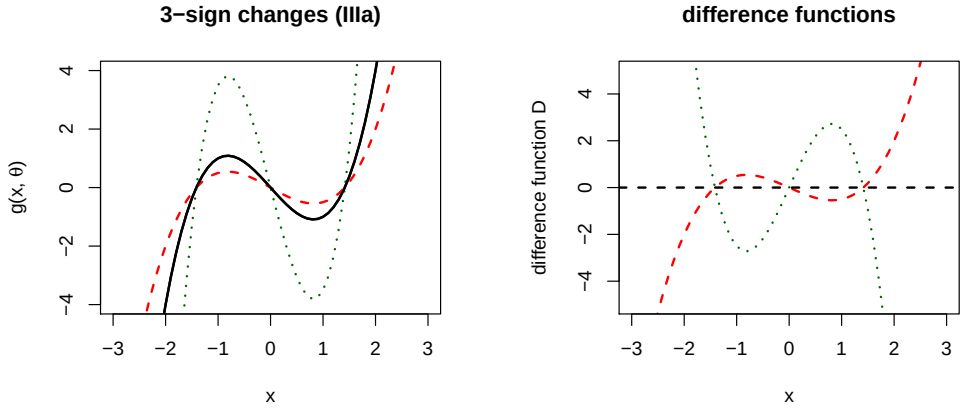

Figure 30: Illustration of the model class IIIa with true parameter $\boldsymbol{\theta}^* = (-2, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (-1, 0.5)^\top$ (red) and $\boldsymbol{\theta} = (-7, 3.5)^\top$ (green)

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.63	0.05	0.83	0.50	0.82	0.60	0.79	0.00
50	0.92	0.27	0.95	0.82	0.95	0.86	0.94	0.88
100	0.98	0.83	0.99	0.94	0.99	0.95	0.99	0.96
200	1.00	0.95	1.00	0.98	1.00	0.99	1.00	0.99

Table 12: Mean of the simulated power for 3-sign change alternatives (IIIa)

Table 12 and Figure 31, we obtain reversed results concerning the performance of the parity of K. Odd $K \geq 5$ are here better than even K. We can also see that the results are better for $K = 3$ than for $K = 4$ but also slightly worse than for higher odd K. This can be explained by the mentioned conjecture that the 3-depth cannot

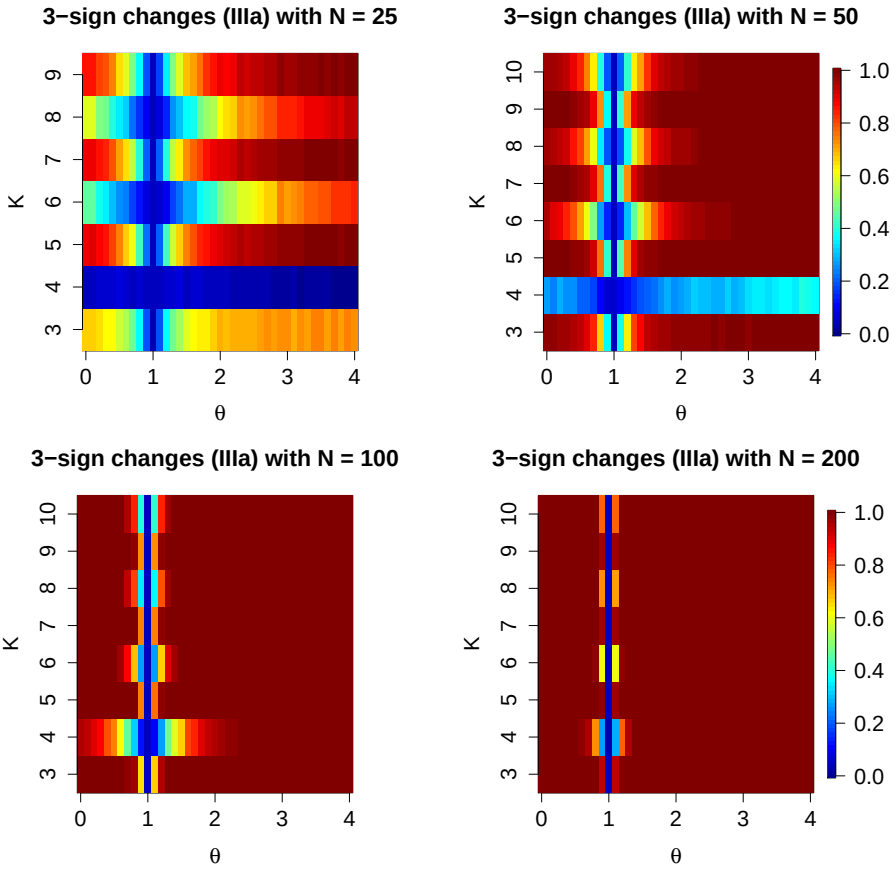

Figure 31: Power functions for the 3-sign changes alternatives (IIIa)

reject model situations with more than two sign changes under small N. However, this deficit of the 3-depth is compensated by increasing N. For even $K \geq 6$, the results are also noticeably better up $N \geq 50$ and also the 4-depth gets better for $N \geq 100$.

3-sign changes model (IIIb)

Figure 32 shows the true model equation $g(x, \boldsymbol{\theta}^*) = x^3 - 2x$ in black again. The

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.87	0.35	0.98	0.82	0.98	0.88	0.97	0.00
50	1.00	0.77	1.00	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 13: Mean of the simulated power for 3-sign change alternatives (IIIb)

alternatives are given by $g(x, \boldsymbol{\theta}) = -g(x, \boldsymbol{\theta}^*) + \theta$ for $\theta \in [-3, 3]$. Figure 33 presents a similar results to the other 3-sign changes case (IIIa). The power functions for the K-depth with odd K are higher than for even K. The 3-depth performs worse than

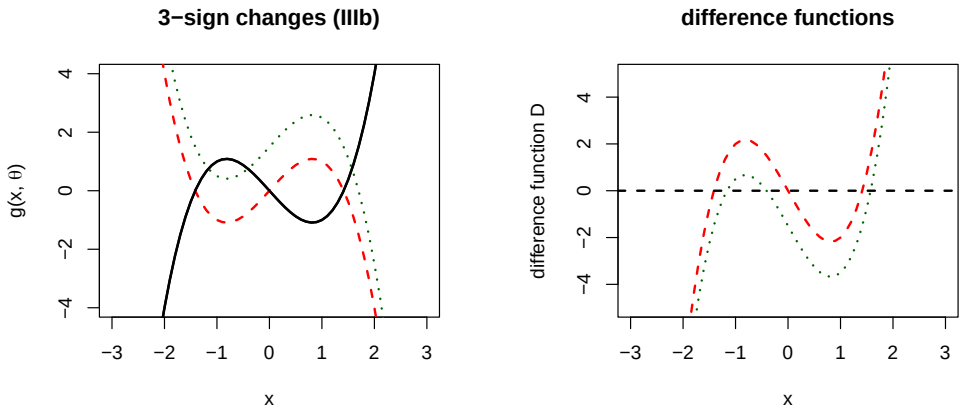

Figure 32: Illustration of the model class IIIb with true parameter $\boldsymbol{\theta}^* = (0, -2, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (0, 2, -1)^\top$ (red) and $\boldsymbol{\theta} = (1.5, 2, -1)^\top$ (green)

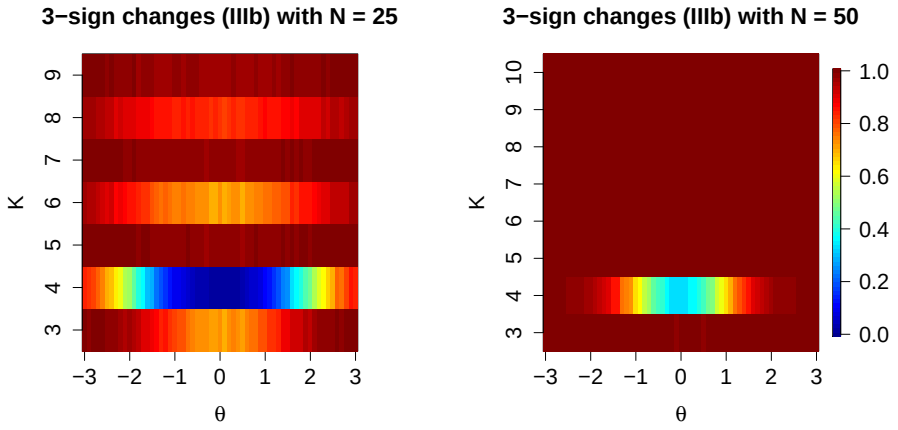

Figure 33: Power functions for the 3-sign change alternatives (IIIb)

the K-depth for odd $K \geq 5$. The case $K = 4$ is here noticeably worse than the other even $K \geq 6$. For $N \geq 100$, all power function are almost equal to one, cf. Table 13.

4-sign changes model (IVa)

Figure 34 shows 4-sign changes alternatives generated by rescaling the true model equation $g(x, \boldsymbol{\theta}^*) = \frac{1}{2} - 2x^2 + x^4$ with $\theta \cdot g(x, \boldsymbol{\theta}^*)$ for $\theta \in [0, 4] \setminus \{1\}$.
Similar results to the alternatives with 2-sign changes are given in Figure 35 for the power functions. However, it is noticeable that the 5-depth performs worse than for the 2-sign change case for $N = 25$. For $N = 50$, the power for all $K \geq 4$ is very high. For $K = 3$, the sample sizes $N \in \{50, 100\}$ do not deliver good results.
The worse performance of the 5-depth can be explained by the fact that the difference functions have four intersections. Therefore, we expect five blocks in the residuals which explains the better performance of the K-depths for odd $K \geq 7$ as

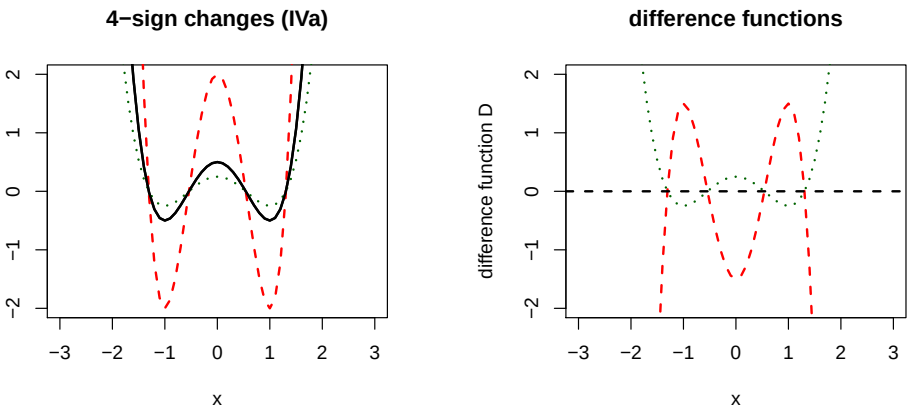

Figure 34: Illustration of the model class IVa with true parameter $\boldsymbol{\theta}^* = (0.5, -2, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (2, -8, 4)^\top$ (red) and $\boldsymbol{\theta} = (0.25, -1, 0.5)^\top$ (green)

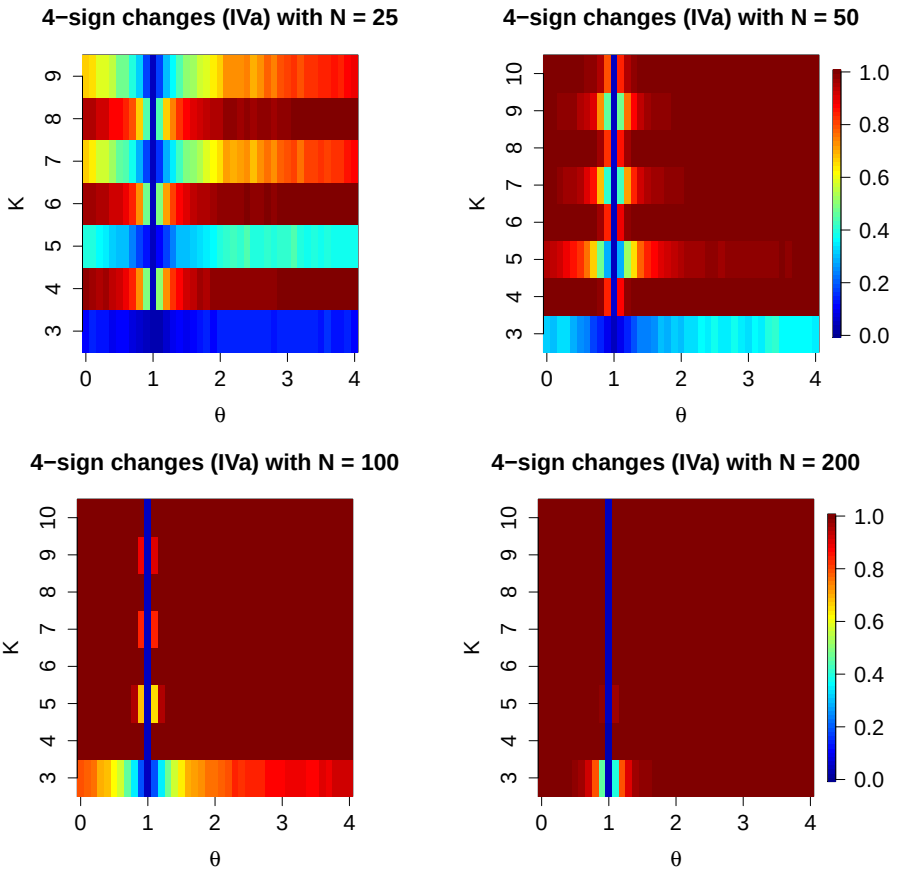

Figure 35: Power functions for the 4-sign changes alternatives (IVa)

well. Nevertheless, the 4-depth has a surprisingly high power although the hyper-parameter $K = 4$ is too small compared to the number of sign changes. E.g., the 3-depth performed much worse in the (IIIa)-model. However, similar effects, we see for the 3-depth in the (IIIa)-model, could be obtained for the 4-depth in the (IVa)-

model by extending the distances between the intersections. Table 14 presents the average power values.

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.13	0.92	0.35	0.92	0.62	0.90	0.64	0.00
50	0.30	0.99	0.87	0.99	0.94	0.99	0.95	0.99
100	0.73	1.00	0.98	1.00	0.99	1.00	0.99	1.00
200	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 14: Mean of the simulated power for 4-sign change alternatives (IVa)

4-sign changes model (IVb)

Figure 36 presents alternatives based on the same true model (IVa) with reversed signs given by $g(x, \boldsymbol{\theta}^*) = \frac{1}{2} - 2x^2 + x^4$ and alternatives generated by $-g(x, \boldsymbol{\theta}^*) + \theta$ for $\theta \in [-3, 3]$. In Figure 37 and Table 15, we can see that for $N = 25$ and odd K, the power function are worse than for even K.

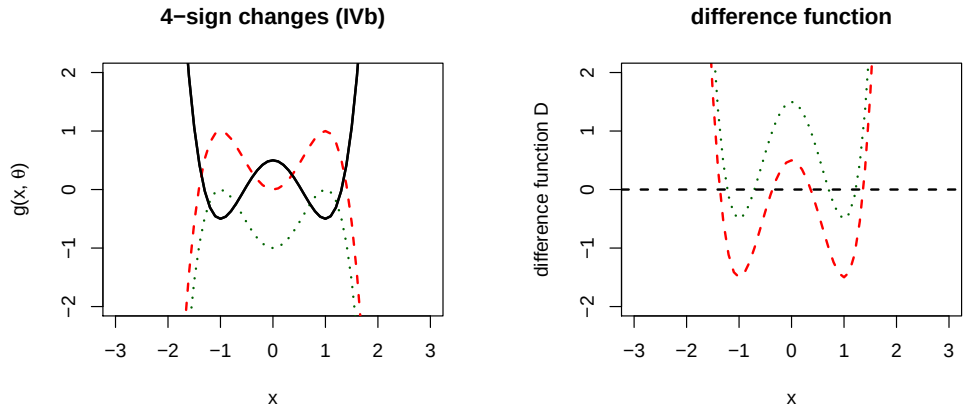

Figure 36: Illustration of the model class IVb with true parameter $\boldsymbol{\theta}^* = (0.5, -2, 1)^\top$ (black), alternatives $\boldsymbol{\theta} = (0, 2, -1)^\top$ (red) and $\boldsymbol{\theta} = (-1, 2, -1)^\top$ (green)

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.32	0.99	0.69	1.00	0.88	0.99	0.89	0.00
50	0.53	1.00	0.99	1.00	1.00	1.00	1.00	1.00
100	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 15: Mean of the simulated power for 4-sign change alternatives (IVb)

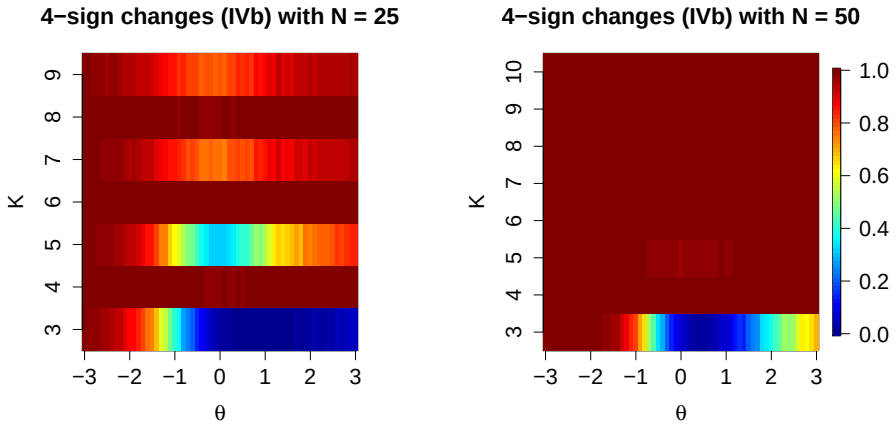

Figure 37: Power functions for the 4-sign change alternatives (IVb)

5-sign changes model (Va)

Analogously, the 5-sign change alternatives by scalings and their power functions are presented in Figure 38 and Figure 39, respectively. Table 16 gives an overview of

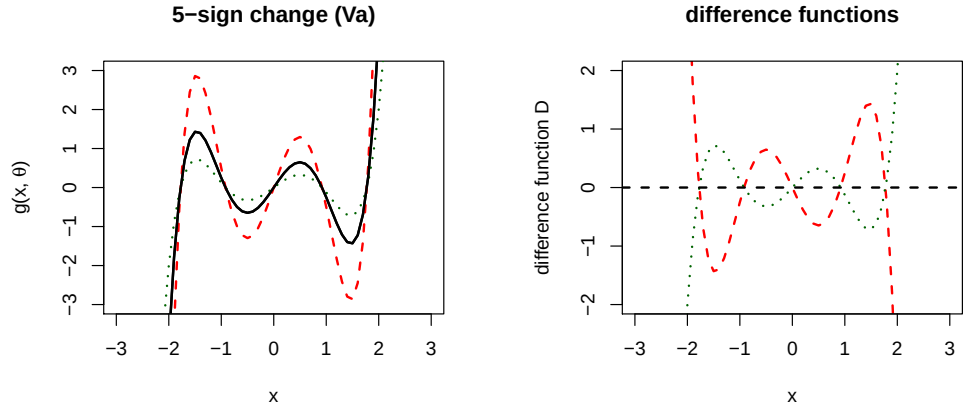

Figure 38: Illustration of the model class Va with true parameter $\boldsymbol{\theta}^* = (2, -3, 0.75)^\top$ (black), alternatives $\boldsymbol{\theta} = (4, -6, 1.5)^\top$ (red) and $\boldsymbol{\theta} = (1, -1.5, 0.375)^\top$ (green)

the average power values. We have again the result that for odd K with $K \geq Q+2$,

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.34	0.02	0.69	0.07	0.86	0.37	0.85	0.00
50	0.94	0.06	0.99	0.67	0.99	0.90	0.99	0.93
100	1.00	0.44	1.00	0.97	1.00	0.99	1.00	0.99
200	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00

Table 16: Mean of the simulated power for 5-sign change alternatives (Va)

the power functions are high for smaller N and even K perform worse.

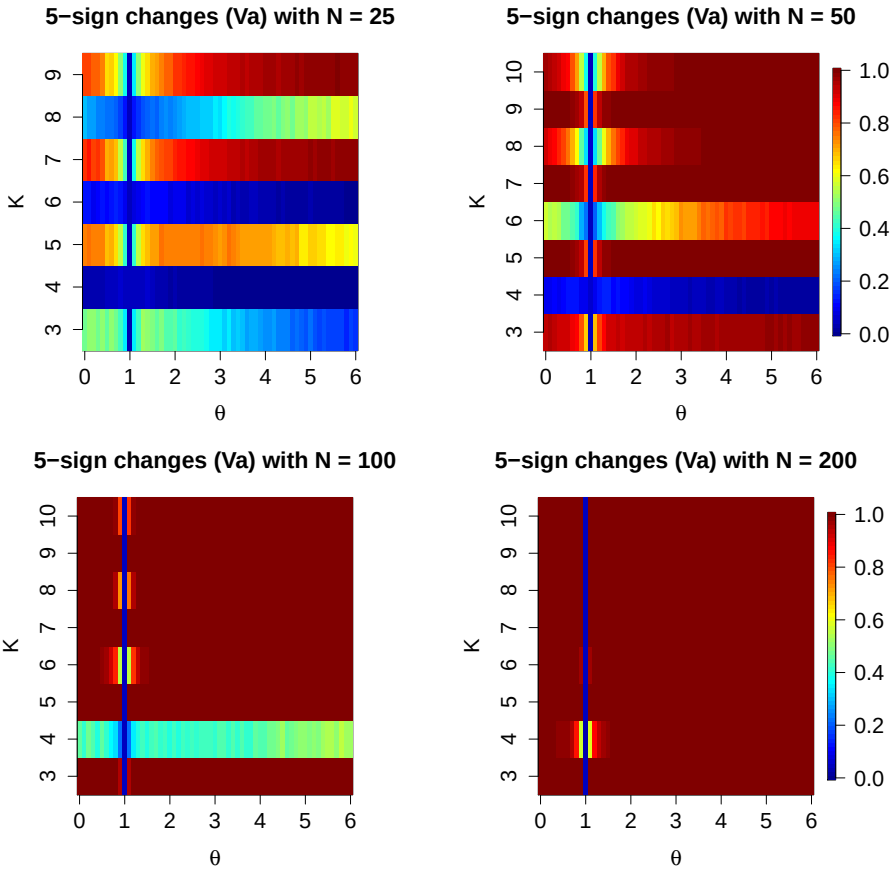

Figure 39: Power functions for the 5-sign change alternatives (Va)

5-sign changes model (Vb)

Further, the 5-sign change alternatives by shifts and their power functions are presented in Figure 40 and Figure 41, respectively.

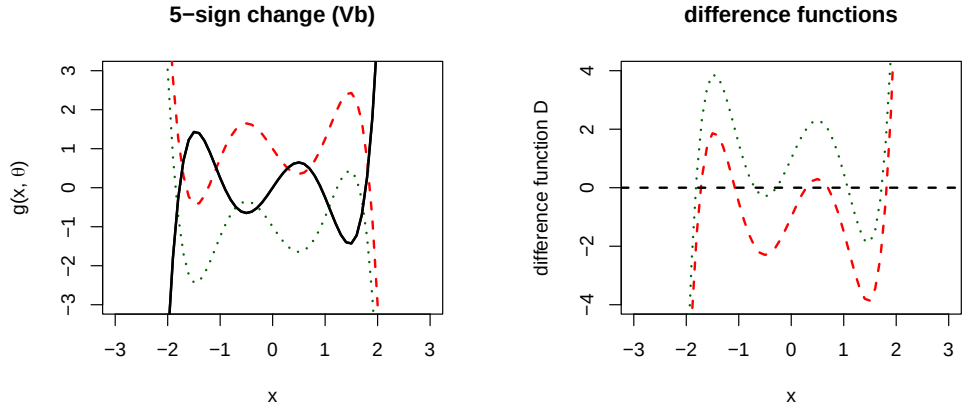

Figure 40: Illustration of the model class Vb with true parameter $\boldsymbol{\theta}^* = (0, 2, -3, 0.75)^\top$ (black), alternatives $\boldsymbol{\theta} = (1, -2, 3, -0.75)^\top$ (red) and $\boldsymbol{\theta} = (-1, -2, 3, -0.75)^\top$ (green)

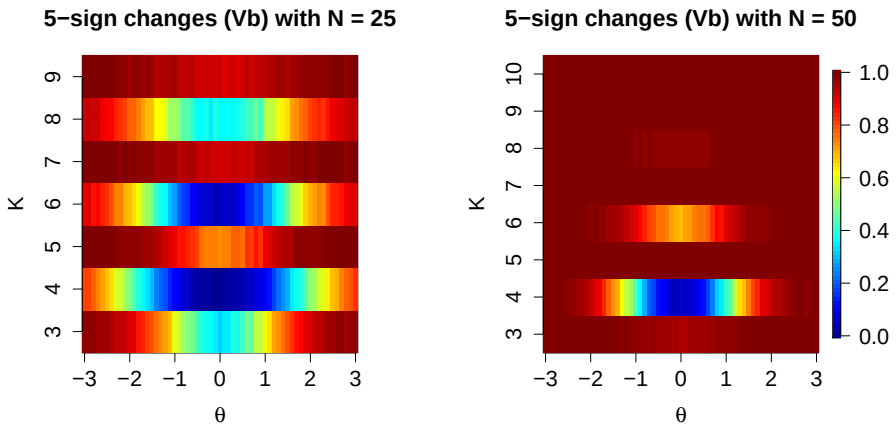

Figure 41: Power functions for the 5-sign change alternatives (Vb)

Table 17 gives an overview of the average power values.

N	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
25	0.73	0.35	0.91	0.47	0.96	0.65	0.96	0.00
50	0.99	0.63	1.00	0.92	1.00	1.00	1.00	1.00
100	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 17: Mean of the simulated power for 5-sign change alternatives (Vb)

6.4.2 Conclusions for the choice of K

Let Q be the number of sign changes in the difference function of a given alternative parameter. Then the simulation study leads to the following conclusions:

> The K-sign tests have the highest power for $K = Q + 2$. For $K = Q + 2 + 2q$ with $q \in \mathbb{N}$, the power will be slightly lower for increasing q. If we consider $K = Q + 2 + 2q - 1$, then the power is worse compared to K with reversed parities. However, if $K < Q + 2$, then the power will be low, especially for K and Q with different parities. This low power can be compensated by increasing the sample size N. The results of the simulation study lead to the suggestion of consistency properties for all $K \geq 3$ in these scenarios. The convergence rate seems to be higher for $K \geq Q + 2$, i.e., we do not need such high sample sizes to gain higher powers.

Note that the case $K = 2$ yields a bad power and inconsistency properties in many situations since the 2-depth is asymptotically equivalent to counting the signs of the residuals (Leckey et al., 2020). The only situation the 2-depth performs well is the

case of constant models with model function $g(\theta) = \theta$ which fits also to the above conjectures since $Q = 0$ then (Lehmann and Romano, 2005, p. 701).
We want to give an illustration how to explain these results. Therefore, we consider the quadratic function $g(x) = x^2 - 1$ as the true model and $h(x) = -x^2 + 1$ as an alternative, cf. Figure 42. On the vertical coordinate axis of this figure, we see the sign

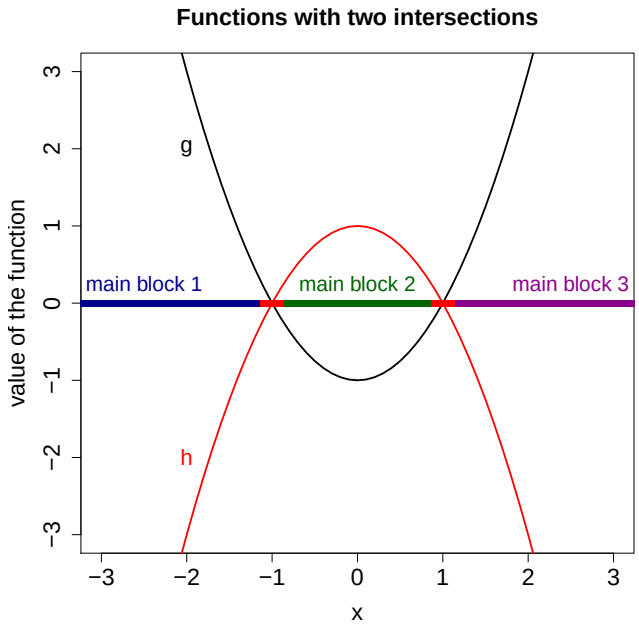

Figure 42: Example of two functions with two intersections

structure of the difference function. The difference function $g - h$ has positive signs for main block 1 and 3 and negative signs for main block 2. Due to the additionally given errors, we will expect a qualitative structure of the signs of the residuals as given in Figure 43. We have three **main blocks**, i.e., these blocks are the longest

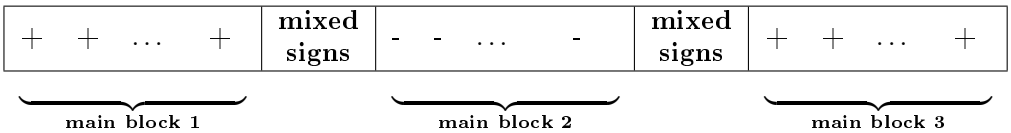

Figure 43: Overview of the scenario with the three main blocks

and have a high impact on the value of the depth. Between these main blocks, we have **mixed signs** since near to the intersections of the functions, the probability of a positive and negative sign is approximately equal and block lengths are small. If we consider $K = 3$, then according to the block representation:

$$d_3(r_1, \ldots, r_N) = \frac{1}{\binom{N}{3}} \sum_{(i_1,i_2,i_3) \in \mathcal{A}_{3,B}} \prod_{k=1}^{3} q_{i_k}$$

all main blocks are once multiplied as one summand in the above sum. This will also be the case for other odd K which can have even more K-tuples containing the main blocks. The multiplication of the three main blocks leads to a strong increase of the K-depth. Especially for small N, the multiplication of the three main blocks in one summand dominates the resulting depth. For increasing odd K, there may occur the situation that the number of blocks in the mostly alternating parts are such small that combinations with all main blocks are not possible and the K-depth decreases. For even K, it is not possible to construct a K-tuple containing all main blocks since the block index of the first and last need to have different parities. However, the first and last main block index share the same parities and therefore, we can only combine two main blocks for each summand at most.
This interpretation can be done for other number of sign changes analogously. Especially for odd number of sign changes, we will obtain reversed conclusions for the parities. The bad results for $K \leq Q + 1$ can be explained as well since there are more than one combination of large blocks multiplied with each other.
Note that these explanations are only heuristics and based on some examples. However, these heuristics can be useful for future research to understand the performance of tests theoretically and for the choice of the hyper-parameter K in general. Furthermore, the proportions of the main blocks are constant as $N \to \infty$ for such scenarios. This property may lead to an explanation for the consistency. Proofs for the consistency and the other conclusions are beyond this thesis and are tasks for further research.

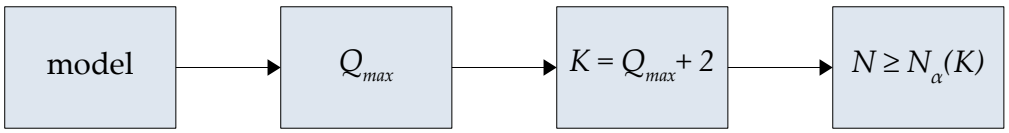

Figure 44: Proceeding for using the K-depth in practice

Figure 44 presents how the K-depth can be applied for statistical analysis. The model, the associated parameters and the hypotheses have to be chosen at first. Then, the maximal number of intersections Q_{max} between the model function under the null hypothesis H_0 and an alternative can be derived. The previous simulation study concludes the choice $K = Q_{max} + 2$. For given K and α-level, Chapter 6.3 gives the smallest sample size $N_\alpha(K)$ which can be used for a convenient test.
The simulations show that the power functions of the K-depth should be considered separately for odd K and even K. However, models with different numbers of sign changes are not included. For further research it should be investigated how the

different behavior of the K-depth for even and odd K appears in this case. In scenarios with various numbers of sign changes, multiple testing with even and odd K could be an idea to increase the power globally. On the other hand, the power would decrease due to a Bonferroni correction of the level. The choice of K should also be investigated for the multiple regeression (Horn, 2021b).

6.5 Testing the fit of set hypotheses

Instead of null hypotheses with $|\mathbf{\Theta}_0| = 1$ as in Chapter 6.4, we also want to consider set hypotheses. From a practical perspective, set hypotheses are commonly used. E.g., we can test if one parameter is zero when the other parameters can be arbitrary so that $\mathbf{\Theta}_0$ is a $(p-1)$-dimensionale subset of $\mathbf{\Theta} = \mathbb{R}^p$. Testing these kinds of hypotheses is useful to test if an explanatory variable can be declared as significant for the model. Furthermore, comparing some parameters from each other by null hypotheses of the form $H_0 : \theta_1 = \theta_2$ or $H_0 : \theta_1 = \ldots = \theta_p$ is typical in the ANOVA (Paolella, 2018, p. 87). As presented in Theorem 6.1 (b), these type of null hypotheses can be tested by replacing the K-depth under one fixed parameter in $\mathbf{\Theta}_0$ with the supremum of the K-depth under all parameters in $\mathbf{\Theta}_0$. In order to compute the supremum of all K-depths with parameters in $\mathbf{\Theta}_0$, we have to solve an optimization problem. Due to the discontinuity of the objective functions, we need derivative-free optimization algorithms (also known as gradient-free algorithms), i.e., they do not require the derivatives of the objective functions (Conn et al., 2009). In Chapter 6.5.1, such optimization algorithms will be shortly listed. In Chapter 6.5.2, a situation for a two-sample relevance test based on the K-depth is presented and compared with a t-test for this context.

6.5.1 Optimization procedures

For an introductory example, we will consider the model

$$Y_n = \theta^* + E_n \text{ for } n = 1, \ldots, 20$$

with $\theta^* \in \mathbb{R}$ as the true parameter and i.i.d. $E_1, \ldots, E_{20} \sim \mathcal{N}(0, 1)$. For $\theta \in \mathbb{R}$ and realizations $y_1, \ldots, y_{20}$, the residuals $r_1(\theta), \ldots, r_{20}(\theta)$ are defined as usual by

$$r_n(\theta) := y_n - \theta \text{ for } n = 1, \ldots, 20$$

and their K-depth can be computed. Figure 45 presents the 3-depth for several parameters in $\{-3, -2.99, -2.98, \ldots, 3.99, 4\}$ for a single simulation based on the

model $\theta^* = 1$. These kinds of discontinuous functions have to be optimized in the following. For multidimensional parameter spaces, similar functions can be obtained. In the following lemma, we formalize and prove the discontinuity.

Lemma 6.6. *Let $e_1, \ldots, e_N \in \mathbb{R}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^d$ be arbitrary and $\boldsymbol{\Theta}$ a connected subset in $\mathbb{R}^p$.*[3] *Moreover, let*

$$y_n = g(\mathbf{x}_n, \boldsymbol{\theta}^*) + e_n, n = 1, \ldots, N$$

for some model function $g : \mathbb{R}^d \times \boldsymbol{\Theta} \to \mathbb{R}$ and $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. We define the function

$$L : \boldsymbol{\Theta} \to \mathbb{R},\ L(\boldsymbol{\theta}) = d_K(r_1(\boldsymbol{\theta}), \ldots, r_N(\boldsymbol{\theta}))$$

for $r_n(\boldsymbol{\theta}) = y_n - g(\mathbf{x}_n, \boldsymbol{\theta})$, $n = 1, \ldots, N$. Then L is discontinuous or constant.

Proof of Lemma 6.6: For arbitrary N, the K-depth $d_K(r_1(\boldsymbol{\theta}), \ldots, r_N(\boldsymbol{\theta}))$ can only provide a finite number of different values. If L takes at least two different values, then the image of L is disconnected since the K-depth only delivers a finite number of values and thus, the points are isolated. Therefore, L must be discontinuous since this would lead to a contradiction to the intermediate value theorem for general topological spaces otherwise (Yan, 2016, p. 138). This theorem implies that for continuous $f : X \to \mathbb{R}$ with X connected, $f(X)$ is connected as well (Yan, 2016, p. 138). If L takes only one value, then L is constant. □

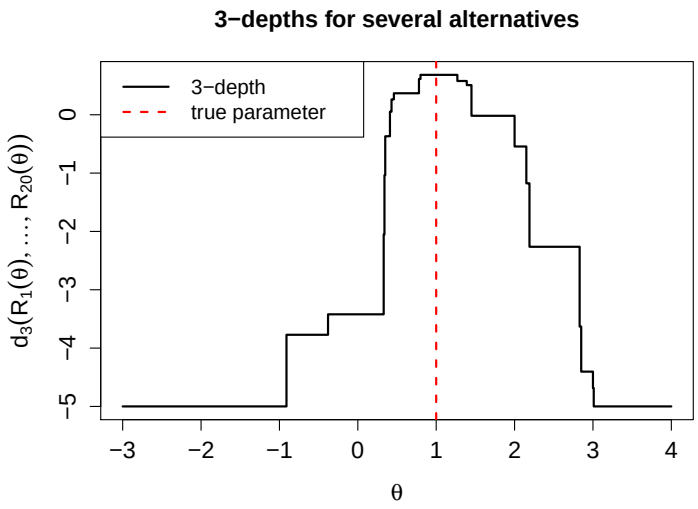

Figure 45: 3-depths for several alternative paramters for illustration of the objective function

[3]The connectedness of $\boldsymbol{\Theta}$ avoids isolated points. The concept of continuity in $\mathbb{R}^p$ according to the canonical topology would not make sense under isolated points.

According to Lemma 6.6, typical optimization procedures as the Newton's method cannot be used since we especially would need differentiable functions (Aragón et al., 2018, p. 209). Note that there exists research that generalizes the theory of differentation for discontinuous functions, see Moreau and Aeyels (2000). Nevertheless, the usual approaches for solving optimization problems for discontinuous functions are derivative-free algorithms. Before we will introduce some prominent optimization procedures, the following estimator should be mentioned in this context as well.

Remark 6.7. *In Figure 45, we see that the highest 3-depths leads also to the true parameter. The idea of using the K-depth as an estimator by finding the maximal depth is intuitive. This approach is known as the* ***maximum-depth estimator****:*

$$\hat{\boldsymbol{\theta}}_{d_K} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, d_K(R_1(\boldsymbol{\theta}), \ldots, R_N(\boldsymbol{\theta})).$$

This estimator is applied for $K = 3$ in Falkenau (2016) for example. However, these estimations are generally not unique and $\hat{\boldsymbol{\theta}}_{d_K}$ is a set therefore. Historically, this approach has already been considered for estimating the center of a multivariate data set as multivariate generalizations of the median. E.g., the halfspace median of Tukey (1975) is the set of $\mathbf{x} \in \mathbb{R}^q$ with the highest halfspace depth or Van Aelst et al. (2002) considers the regression depth median (also known as deepest regression method) as the parameters with the highest regression depth, cf. Chapter 2.3.

List of derivative-free optimization procedures

The number of derivative-free optimization procedures is large and it is beyond this thesis to discuss every method in detail. Therefore, we only will give a brief list of some methods the literature offers.

- **Grid search:** This is a simple algorithm which considers a discrete grid of all parameters in $\boldsymbol{\Theta}_0 \subseteq \mathbb{R}^p$ or an approximation if $\boldsymbol{\Theta}_0$ is unbounded for example. The major drawback of this method are the high computational costs for increasing dimensions of $\boldsymbol{\Theta}_0$ (Audet and Hare, 2017, p. 35).
- **Nelder-Mead method (Downhill simplex method):** This algorithm uses simplices which rotate and shrink to the direction with the largest descent (Conn et al., 2009, p. 141). Note that there exists several examples where this algorithm does not converge (McKinnon, 1998) or can be very slow (Baudin, 2009). Although the lack of convergence theory for this algorithm is often claimed as a major drawback, there exists research where conditions for the convergence is studied (Lagarias et al., 1998).

- **Simulated annealing:** In contrast to a deterministic grid search, simulated annealing is a stochastic algorithm considering several points of $\boldsymbol{\Theta}_0$. It is motivated on the idea of slowly cooling material. After an initial evaluation, this algorithm evaluates nearby solutions by a random choice based on weighted probabilities. Bad directions with smaller probabilities are also possible in order to avoid being stuck in local optima. At the beginning, the jumps between several solutions can be very high. Towards the end, the jumps get smaller until the algorithm 'cooled down' and ends. Simulated annealing is often used for finding the global optimum and the convergence of the algorithm is studied for several models (Dekkers and Aarts, 1991). However, this algorithm involves many tuning parameters influencing the quality of the solution. Therefore, Siarry (2016) gives practical suggestions for the parameter settings.

- **Particle swarm:** This stochastic algorithm was introduced by Kennedy and Eberhart (1995). At the beginning of the algorithm, randomly initialized evaluations are considered as particles of a population where each particle gets a random momentary direction. After each further iteration, a particle changes its position by a combination of some directions, e.g.:
  - the direction from the previous iteration,
  - the best direction a particle had in its past,
  - the direction of the swarm.

  The idea is that the swarm will converge to an optimal point which is inspired by the phenomenon of collective intelligence in biology. Qian and Li (2018) present this class of algorithms in its standard and improved forms and analyze their convergence.

- **Genetic algorithms:** This class of algorithms is inspired by the process of natural selection. We consider a random initial population and some fitness function (this can also be the objective function) evaluating each individual. Then biological processes such as recombination, mutation and selection are applied based on the Darwinian evolution theory (Chambers, 2001). In particular, individuals from the population with a high fitness are more likely to be selected for recombination to produce new individuals for the next generation. Moreover, individuals change randomly causing genetic diversity. Similar to the Nelder-Mead method, a guarantee of convergence is not given (Audet and Hare, 2017, p. 55).

There are other derivative-free algorithms such as the great deluge algorithm (Dueck, 1993) or Bayesian optimization (Frazier, 2018). For a further discussion of derivative-free optimization algorithms, we refer to Audet and Hare (2017). The correct choice of an optimization algorithm depends on the properties of the objective function. In the next chapter, we will show for an example how a specific and simple optimization algorithm can be constructed for a class of objective functions. In other situations, the previously mentioned algorithms could be applied instead.

6.5.2 Example: Two-sample test for relevant differences

We consider two samples given by

$$Y_n = \theta_1 + E_n \text{ for } n = 1, \dots, M$$
$$Y_n = \theta_2 + E_n \text{ for } n = M+1, \dots, N$$

so that $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$ is the unknown parameter vector and $E_1, \dots, E_N$ satisfy Assumption 2.1. We can sum up this model equation to one equation by

$$Y_n = \theta_1 \mathbb{1}\{1 \leq n \leq M\} + \theta_2 \mathbb{1}\{M+1 \leq n \leq N\} + E_n \text{ for } n = 1, \dots, N.$$

We are interested in testing the following hypotheses

$$H_0 : |\theta_1 - \theta_2| \leq \delta, H_1 : |\theta_1 - \theta_2| > \delta \tag{79}$$

for a given relevance parameter $\delta \geq 0$. For $\delta = 0$, we have a standard two-sample test with $H_0 : \theta_1 = \theta_2$. E.g., this type of test can be used to detect a change point between $Y_1, \dots, Y_M$ and $Y_{M+1}, \dots, Y_N$ with jump height δ or more (Dette and Kutta, 2021). The set of parameters under which H_0 in Formula (79) is correct has the form:

$$\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta} \in \mathbb{R}^2; |\theta_1 - \theta_2| \leq \delta\}. \tag{80}$$

For realizations of the residual vector under any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, we can consider the test procedure

$$\text{reject } H_0 \text{ if } \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} d_K(r_1(\boldsymbol{\theta}), \dots, r_N(\boldsymbol{\theta})) < q^K_{\alpha,N}$$

where $q^K_{\alpha,N}$ denotes the exact α-quantile of the K-depth. In the following, we will always assume that $M = \frac{N}{2}$. Figure 46 shows the behavior of the objective function $L(\boldsymbol{\theta}) = d_K(r_1(\boldsymbol{\theta}) \dots, r_N(\boldsymbol{\theta}))$ in an 101×101 grid when the location parameters θ_1

and θ_2 change. The area surrounded by the red lines corresponds to $\mathbf{\Theta}_1 = \mathbb{R}^2 \backslash \mathbf{\Theta}_0$ also

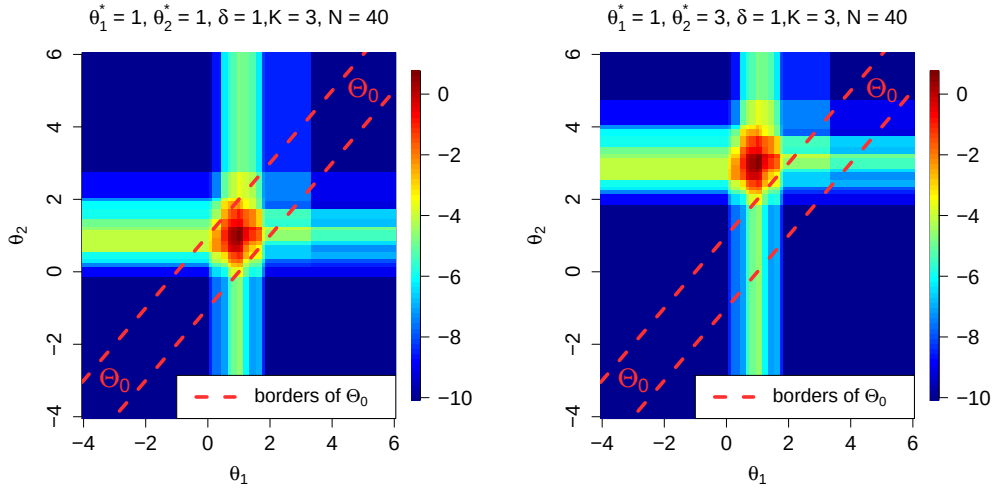

Figure 46: The objective function $L(\boldsymbol{\theta})$ for two cases for the location parameters

known as the relevance area. In contrast, $\mathbf{\Theta}_0$ is referred to the non-relevance area. Both functions were simulated under the same random seed for better comparability. The values of the function L show parallel contour lines and the depth is decreasing for shifts along the θ_1-axis or θ_2-axis from the maximal value. We can take advantage of this property for the next optimization procedure based on a simple heuristic ideas.

Remark 6.8. *The optimization under $\mathbf{\Theta}_0$ in Formula (80) will be done in two steps.*

1. *The maximum K-depth in the whole parameter space $\mathbf{\Theta}$ is **approximately** computed by the median of each sample:*

$$\hat{\boldsymbol{\theta}}_{max} = (med(y_1, \ldots, y_M), med(y_{M+1}, \ldots, y_N))^\top.$$

   In Figure 47, this point is presented by a triangle. However, this value can be outside $\mathbf{\Theta}_0$. If this happens, we consider Step 2.

2. *One coordinate of $\hat{\boldsymbol{\theta}}_{max}$ is fixed and the other coordinate is shifted such that the absolute distance between the two coordinates is lower than δ. This will be done for both coordinates. Formally, we consider the union of the two sets:*

$$\{(\theta_1, \theta_2)^\top \in \mathbb{R}^2; \theta_2 = med(y_{M+1}, \ldots, y_N), |\theta_1 - \theta_2| \leq \delta\}$$
$$\{(\theta_1, \theta_2)^\top \in \mathbb{R}^2; \theta_1 = med(y_1, \ldots, y_M), |\theta_1 - \theta_2| \leq \delta\}$$

   and consider a grid search. In Figure 47, these points from the grid search are highlighted.

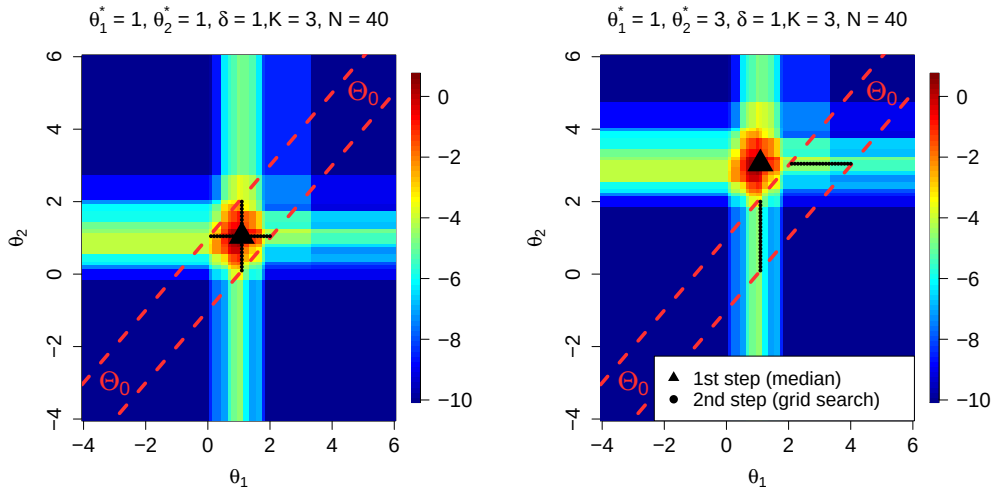

Figure 47: Visulization of an optimization approach from the objective functions from Figure 46

Note that this optimization approach will not work in general. E.g., the objective function L can also have rotated structures in other models and therefore, the shape of L has to be studied in advance.

This remark illustrates how using geometric properties of the objective functions can replace similar information we would obtain from derivatives. For other models and objective functions, similar approaches should be developed.

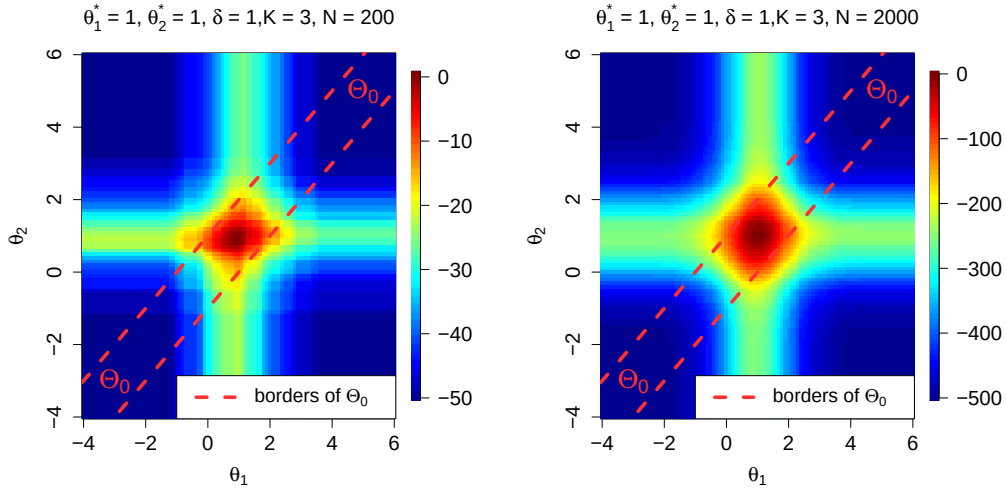

Figure 48: The objective function $L(\boldsymbol{\theta})$ for two different sample sizes

In Figure 48, the objective functions for higher sample sizes are given. In Figure 49, the cases for $K = 4$ and $K = 7$ are presented. Note that the legends change in these figures. On the one hand, higher sample sizes yield an objective function with

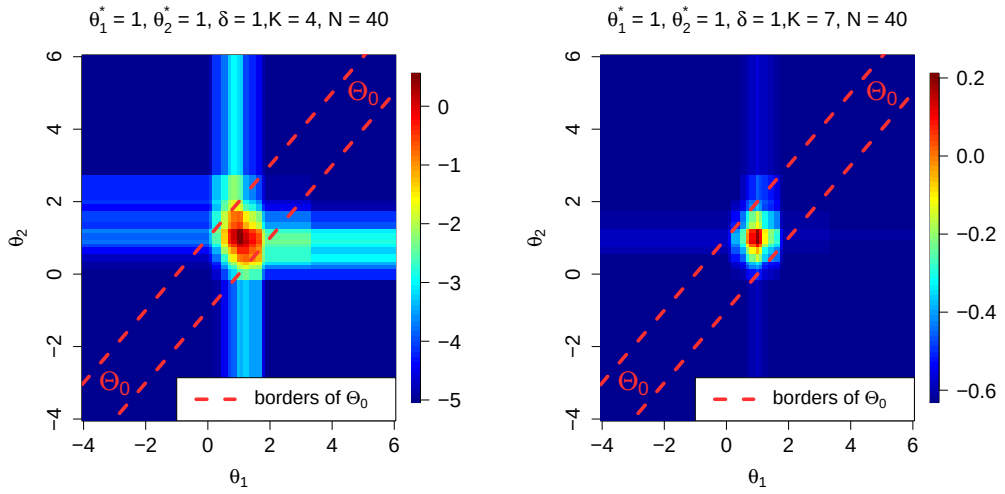

Figure 49: The objective function $L(\boldsymbol{\theta})$ for different K

more values which leads to a smoother shape. On the other hand, increasing K leads to functions more concentrated to the center similar to the estimated density functions of the K-depth in Chapter 3.4. Other scenarios as Cauchy distributed errors or different sizes of groups yield similar figures. A more careful simulation study show that the decision for rejection under H_0 for the optimal depth from the procedure in Remark 6.8 is mostly the same as for the maximal depth based on a grid search in $\boldsymbol{\Theta}_0$. Therefore, we assume that the presented optimization procedure in Remark 6.8 is valid to use in practice for this model. A proof for the convergence of this algorithm to the true maximum depth in $\boldsymbol{\Theta}_0$ is beyond this thesis.
The tests based on the K-depth are compared with a relevance test based on the t-test. This two-sample relevance t-test is applied by testing the following hypotheses:

$$H_0^{(1)} : \theta_1 - \theta_2 \leq \delta, H_1^{(1)} : \theta_1 - \theta_2 > \delta,$$
$$H_0^{(2)} : \theta_1 - \theta_2 \geq \delta, H_1^{(2)} : \theta_1 - \theta_2 < \delta.$$

Each pair of hypotheses is tested by a two-sample t-test. Based on the Bonferroni correction, both tests are adjusted to $\frac{\alpha}{2}$-level tests for testing the null hypothesis H_0 in Formula (79) with level α. H_0 is rejected if either $H_0^{(1)}$ or $H_0^{(2)}$ is rejected. Another two-sample t-test for relevant differences is presented in the Bachelor thesis Malcherczyk (2018b). This test yields similar results compared to the other t-test but the variance is assumed to be known. This is a too restrictive assumption for many applications and problematic if the second moment of the errors do not exist either. Therefore, we do not consider the second t-test from this thesis.

Results of the simulation study

Figure 50 and 51 present the resulting power functions of a simulation study. The parameter $\theta_2 = 1$ is fixed while $\theta_1 \in \{-4, -3.9, \ldots, 4.9, 5\}$ is varied and the null hypothesis $H_0 : |\theta_1 - \theta_2| \leq 1$ is tested 100 times per parameter.

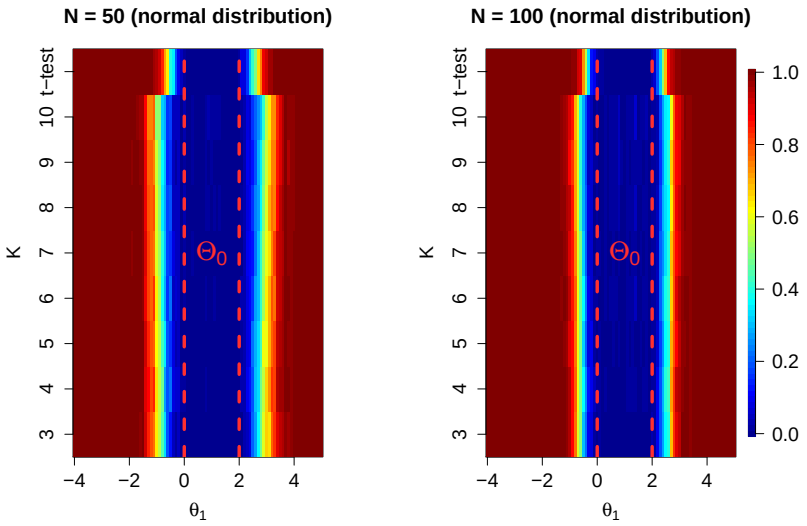

Figure 50: The power function of the relevance tests under $\mathcal{N}(0, 1)$-distribution

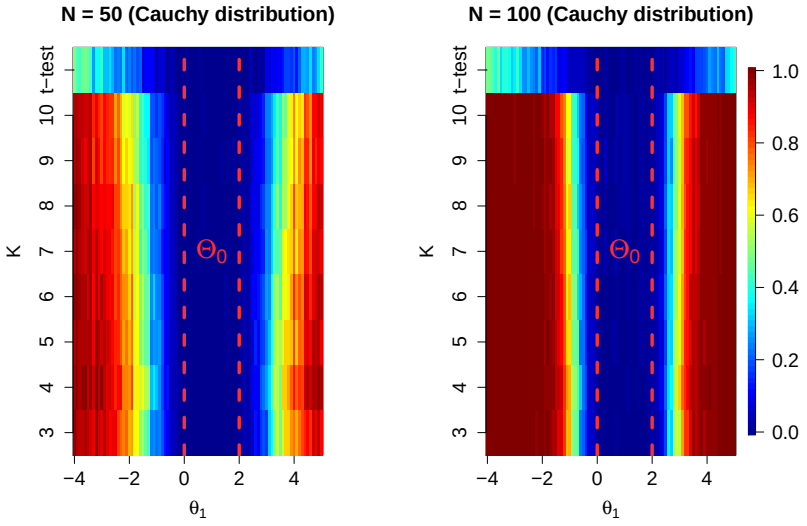

Figure 51: The power function of the relevance tests under $Cau(0, 1)$-distribution

In Figure 50 and Figure 51, the sample sizes $N = 50$ and $N = 100$ with two equal groups and $\mathcal{N}(0, 1)$- or $Cau(0, 1)$-distributed errors, respectively, are considered. For normally distributed errors, the t-test yields better results. The tests based on the

K-depth have only slight differences. A more careful consideration on the power function shows that the power is higher for even K than in all other presented scenarios. Especially, the power function is uniformly the highest for $K = 4$. Furthermore, the K-depth shows consistency properties, i.e., the power increases to one for increasing sample sizes. Similar to the explanations in Chapter 6.4.2, one sign change in the difference function can be expected under several alternatives which explains the better performance of even K.
For Cauchy distributed errors, the t-test fails and seems not to have consistency properties. This is not very surprising since the moments of the Cauchy distribution do not exist. However, the t-statistic is based on the mean and empirical variance in order to estimate the expected value and the variance, respectively, which is not suitable for Cauchy distributed random variables. The K-depth tests yield similar results to the normally distributed case but the improvement from the case $N = 100$ to $N = 50$ is more noticeable. The case $K = 4$ also performs best here.

Outlook for relevance tests based on the K-depth

We will focus on a particular class of parameter constellations in $\boldsymbol{\Theta}_1$ which may be a difficult situation for the K-depth for rejecting the null hypothesis. There exists always an alternative parameter such that one coordinate can be part of the true parameter. Therefore, one group has nearly alternating signs. If the location parameters of both groups have a high distance from each other, the signs of the other group are nearly constant (i.e., the most of them are positive or negative). Such a scenario is given in Figure 52. Note that the associated random walk which

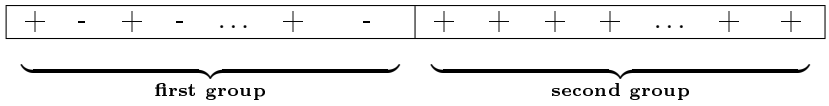

Figure 52: A difficult scenario for the relevance test based on the K-depth

sums up the signs of the residuals, cf. Theorem 3.22, p. 39, tends to ∞ for increasing N under this situation if the relative size of both groups keeps constant. Therefore, the relevance test based on the K-depth can be improved more since considering the depth of the combined samples can yield a power loss. Associated parameters to Figure 52 have not a depth equal to zero although one component of the parameter vector fits poorly. Therefore, the minimum of the depths for each sample seems to be more natural to consider since this will lead to a more sensitive statistic. The next theorem describes how a test can be constructed based on the minimal depth. For the situation of more than two groups, the minimum is also more natural if the

order of the groups is not clearly given. Note that the distribution changes under the minimum so that the quantiles have to be corrected accordingly. As an outlook, we formulate the following test:

Theorem 6.9. *Let N be an even integer and $M = \frac{N}{2}$. Let $Y_1, \dots, Y_M, Y_{M+1}, \dots, Y_N$ be random variables such that*

$$Y_n = \begin{cases} \mu_1^* + E_n, & \text{for } n = 1, \dots, M \\ \mu_2^* + E_n, & \text{for } n = M+1, \dots, N \end{cases}$$

with $(\mu_1^*, \mu_2^*)^\top \in \mathbb{R}^2$. Let $E_1, \dots, E_N$ satisfy Assumption 2.1. Consider the hypotheses

$$H_0 : |\mu_1 - \mu_2| \leq \delta, H_1 : |\mu_1 - \mu_2| > \delta$$

for $\delta \geq 0$. For $(\mu_1, \mu_2)^\top \in \mathbb{R}^2$, we define the residual vectors by

$$\mathbf{R}^{(1)}(\mu_1) := (Y_1 - \mu_1, \dots, Y_M - \mu_1)^\top, \mathbf{R}^{(2)}(\mu_2) := (Y_{M+1} - \mu_2, \dots, Y_N - \mu_2)^\top.$$

Then we obtain an α-level test for the following decision rule:

$$\text{Reject } H_0 \text{ if } \sup_{|\mu_1 - \mu_2| \leq \delta} \min\{T_K(\mathbf{R}^{(1)}(\mu_1)), T_K(\mathbf{R}^{(2)}(\mu_2))\} < q^K_{\tilde{\alpha},N}$$

for $\tilde{\alpha} = 1 - \sqrt{1-\alpha}$ and $q^K_{\alpha,N}$ as the finite sample quantile of the K-depth.

Proof of Theorem 6.9: Let $(\mu_1^*, \mu_2^*) \in \mathbb{R}^2$ be under H_0, i.e., $|\mu_1^* - \mu_2^*| \leq \delta$. By using the independence of $\mathbf{R}^{(1)}(\mu_1^*)$ and $\mathbf{R}^{(2)}(\mu_2^*)$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{|\mu_1 - \mu_2| \leq \delta} \min\{T_K(\mathbf{R}^{(1)}(\mu_1)), T_K(\mathbf{R}^{(2)}(\mu_2))\} < q^K_{\tilde{\alpha},N}\right) \\
&\leq \mathbb{P}\left(\min\{T_K(\mathbf{R}^{(1)}(\mu_1^*)), T_K(\mathbf{R}^{(2)}(\mu_2^*))\} < q^K_{\tilde{\alpha},N}\right) \\
&= 1 - (1 - \mathbb{P}\left(T_K(\mathbf{R}^{(1)}(\mu_1^*)) < q^K_{\tilde{\alpha},N}\right)) \cdot (1 - \mathbb{P}\left(T_K(\mathbf{R}^{(2)}(\mu_2^*)) < q^K_{\tilde{\alpha},N}\right)) \\
&\leq 1 - (1 - \tilde{\alpha})(1 - \tilde{\alpha}) = 1 - (1 - \tilde{\alpha})^2 = \alpha
\end{aligned}$$

which implies that we have an α-level test. □

If $M \neq \frac{N}{2}$ then the quantile $q_{\tilde{\alpha}}$ in Theorem 6.9 cannot be used. Instead of this, the quantiles of the statistic have to be simulated under the true parameter. Theorem 6.9 can be extended for more than two samples to an ANOVA-type test based on the minimum of the K-depths. Nevertheless, computing the supremum is still challenging in higher dimensional spaces and more research has to be done in future.

6.6 Independence test approaches based on the K-sign depth

In this chapter, we will give a short overview of several notions of sign depths which can be used for testing independence. In the Master thesis Dohme (2021) and in Dohme et al. (2021), the power of the test in Theorem 6.2 for independence is simulated for various situations and compared with other popular tests from this context. Figure 53 shows an example of this result for the following AR(1)-model

$$Y_0 = 0 \text{ and } Y_n = \theta \cdot Y_{n-1} + E_n, n \in \{1, \dots, N\}$$

with i.i.d. $E_n \sim \mathcal{N}(0, 1)$. The parameter $\theta \in (-1, 1)$ denotes the correlation between

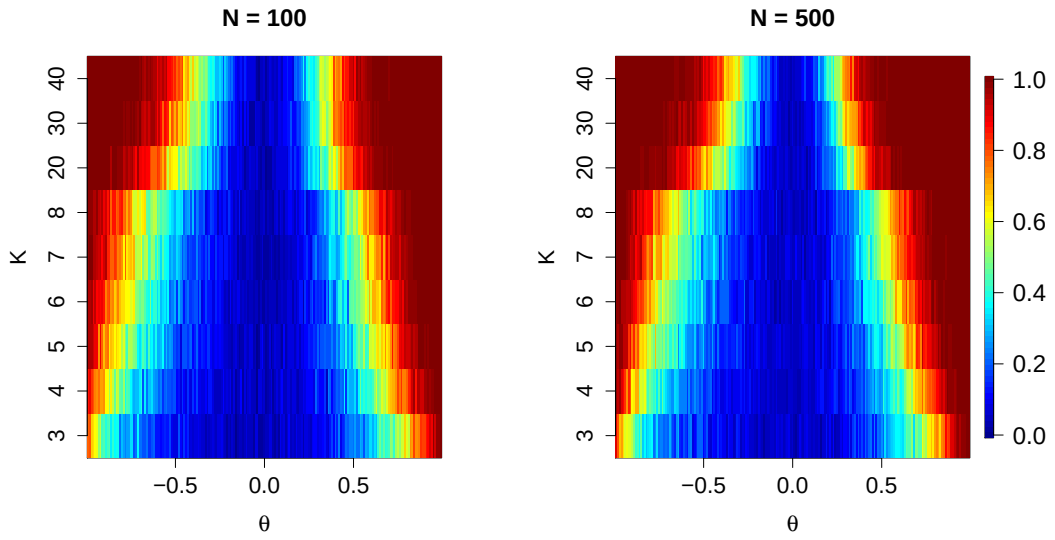

Figure 53: Simulated power functions of the K-depth tests for independence in an AR(1)-model.

Y_{n-1} and Y_n. For $\theta = 0$, $Y_1, \dots, Y_N$ are independent random variables and for $\theta > 0$ or $\theta < 0$, they are positively or negatively correlated, respectively. The hypotheses

$$H_0 : Y_1, \dots, Y_N \text{ are independent}, H_1 : Y_1, \dots, Y_N \text{ are not independent}$$

can be reformulated to the hypotheses

$$H_0 : \theta = 0, H_1 : \theta \neq 0.$$

The power functions are computed by 100 simulated time series of an AR(1)-model for each $\theta \in \{-0.99, -0.98, \dots, 0.98, 0.99\}$, $K \in \{3, \dots, 8, 20, 30, 40\}$ and $N \in \{100, 500\}$. The simulation studies show that the K-depth from Definition 2.2

only detect highly positive or negative correlations. Moreover, increasing the sample size does not improve the power under lower positive or negative correlations so that consistency properties cannot be followed. However, increasing K leads to more sensitive detections of lower correlations.
The supposed inconsistency in N of the K-depth test for fixed K can be explained as follows. The dependencies of the AR-model vanish exponentially and therefore, the AR-model is called short-memory process (Shumway and Stoffer, 2017, p. 241). I.e., these dependencies are only locally noticeable while the K-depth considers all K-tuples with distant entries from each other. Detecting local dependency structures with quickly vanishing autocorrelations based on K-tuples with distant entries does not make sense. This implies that short-memory dependencies cannot be detected suitably by the original K-depth for fixed K. Instead of increasing N, we need to increase also K as high as possible. Thus, we assume consistency properties for *simultaneously* increasing N and K instead. Moreover, recall the necessary sample sizes for rejecting the null hypothesis (cf. Chapter 6.3.1) which lists restrictions how high K can be chosen for given N.
Due to the missing improvements for increasing N and fixed K, we will consider a simplified version to compensate this weakness. This simplified version has already been defined in (Kustosz et al., 2016b) and considers only K-tuples with neighboring entries in windows of length K.

Definition 6.10. *Let* $\mathbf{R} = (R_1, \ldots, R_N)^\top$ *be a residual vector. For* $K \in \mathbb{N} \setminus \{1\}$*, we define the* ***simplified K-sign depth*** *as*

$$d_K^S(\mathbf{R}) := \frac{1}{N-K+1} \sum_{n=1}^{N-K+1} \left(\prod_{k=1}^{K} \mathbb{1}\{R_{n+k-1}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{R_{n+k-1}(-1)^k < 0\} \right).$$

The original K-sign depth from Definition 2.2 is referred as the **full K-sign depth** for the rest of this chapter. The simplified K-depth has been introduced to mimic the full K-depth due to the lack of efficient computations in the past research. The simplified K-depth can be directly computed in linear time by definition compared to the full K-depth which needed more effort as in Chapter 4 and 5 shown.
An asymptotic analysis of the simplified K-depth is given in Kustosz et al. (2016b) based on a Central Limit Theorem for m-dependent random variables, i.e., the random variables are only independent outside windows of length m (Hoeffding and Robbins, 1948).

Theorem 6.11. *For a random vector* $\mathbf{E} = (E_1, \ldots, E_N)^\top$ *satisfying Assumption 2.1*

$$T_K^S(\mathbf{E}) = \sqrt{N-K+1}\frac{d_K^S(\mathbf{E}) - \frac{1}{2^{K-1}}}{\sqrt{\frac{1}{2^{K-1}}\left(3 - \frac{K}{2^{K-2}} - \frac{3}{2^{K-1}}\right)}} \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

For the proof of Theorem 6.11, the expected value and variance of the simplified K-depth have to be derived. Then the assumptions of the Central Limit Theorem for m-dependent random variables have to be checked. The tests in Theorem 6.1 and 6.2 can be reformulate with the simplified K-sign depth by replacing the respective test statistic and the quantiles of a standard normal distribution.
Simulation studies in Kustosz et al. (2016b) and Falkenau (2016) consider testing model parameters by Theorem 6.1 and yield that the full K-depth is more powerful than the simplified K-depth. After the improvement of the computational tools for the full K-depth, the simplified depth seemed to be obsolete for application. However, the simulation studies in Dohme et al. (2021) show that the simplified 2-depth performs well for testing independence. Moreover, the simplified 2-depth test and the Wald-Wolfowitz runs-test are similar since both tests consider an equivalent test statistic. The only difference between these two tests are the critical values, since the Wald-Wolfowitz runs-test considers the conditional distribution based on the number of negative or positive signs in the complete residual vector (Gibbons and Chakraborti, 2003, p. 78). Dohme et al. (2021) conclude that the power of both tests is very similar and supposes that both tests are asymptotically equivalent. For $K > 2$, the simplified K-depths are inappropriate for testing independence based on the results in Dohme et al. (2021). The higher K is chosen, the worse positive and negative correlations can be detected.
Since the simplified 2-depth test is more meant to detect dependencies in AR(1)- or MA(1)-model situations, extending the window length in the simplified K-depth seems to be an intuitive generalization. We introduce a trade-off between the simplified and full K-depth as an outlook for future research.

Definition 6.12. *Let* $\mathbf{R} = (R_1, \ldots, R_N)^\top$ *be a residual vector. For* $K \in \mathbb{N}\setminus\{1\}$ *and* $K \le L \le N$*, we define the* ***simplified (K,L)-sign depth*** *as*

$$d_{K,L}^S(\mathbf{R}) := \frac{1}{(N-L+1)\binom{L-1}{K-1} + \binom{L-1}{K}} \sum_{\substack{1\le n_1<\ldots<n_K\le N \\ n_K - n_1 \le L-1}} \left(\prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(-1)^k < 0\} \right). \tag{81}$$

The main idea of Formula (81) is to consider all K-tuples of the residual vectors with entries in a window of length L. Therefore, the first and last entry of a K-tuple can have a maximum distance of $L-1$. Formula (81) yields the simplified K-depth for $L=K$ and the full K-depth for $L=N$. Note also that the normalization factor coincides for $L=K$ to

$$\begin{aligned}&(N-L+1)\binom{L-1}{K-1}+\binom{L-1}{K}\\ =&(N-K+1)\binom{K-1}{K-1}+\binom{K-1}{K}=N-K+1\end{aligned}$$

since $\binom{K-1}{K}=0$ and for $L=N$

$$(N-L+1)\binom{L-1}{K-1}+\binom{L-1}{K}=\binom{N-1}{K-1}+\binom{N-1}{K}=\binom{N}{K}.$$

In order to see that this number corresponds to the number of summands of the simplified (K,L)-depth, sort all K-tuples by the first index. If $n_1\in\{1,\dots,N-L+1\}$ is fixed, then there are $\binom{L-1}{K-1}$ combinations for the other $K-1$ indices $n_2,\dots,n_K$. Since the first index is fixed, we can count up to $(N-L+1)\binom{L-1}{K-1}$ combinations. However, we have not counted $\binom{L-1}{K}$ combinations from the remaining $L-1$ values which cannot occur as a first index.
Further discussion of the simplified (K,L)-depths is beyond this thesis. Their asymptotic derivation and efficient computation can be content for future research. The proof strategies from this thesis can be useful for analyzing this statistic. Moreover, the performance for longer temporal dependencies as in AR(p)-models for greater p is assumed to be better since longer windows are considered. We still expect that the case $K=2$ will have the best performance for testing independence. Note that this approach may have connections to the procedure in Dürre et al. (2015).

Can the sign function only detect correlations?

The author of this thesis wants to thank gratefully to the referee Prof. Dr. Carsten Jentsch for the following comment. He mentioned that sign depths may only detect correlations but not general dependency structures. E.g., GARCH-processes having autoregressive heteroskedasticity and dependency structures in the variance do not have correlations (Verbeek, 2012, p. 298-299). Dohme (2021) shows that the sign depths are not able to detect the dependencies of the GARCH-process. This may also be caused by the fact that the sign function is not able to detect different scaled variations in the data either. The literature offers other methods

as the distance correlations to detect general classes of dependencies which are not necessarily correlated Székely et al. (2007). An answer to the comment will not be given in the following. Nevertheless, we want to illustrate the following property of random variables having only two different values as the sign function.

Lemma 6.13. *Let $X_1, \ldots, X_N$ be uncorrelated random variables with*

$$\mathbb{P}(X_n = a) = \mathbb{P}(X_n = b) = \frac{1}{2} \text{ for } n = 1, \ldots, N \text{ and } a > 0, b < 0.$$

Then $X_1, \ldots, X_N$ are pairwise stochastically independent.

Proof of Lemma 6.13: Since $X_1, \ldots, X_N$ are uncorrelated and their expected value is $\mathbb{E}(X_n) = \frac{1}{2}(a+b)$, we obtain

$$\mathbb{E}(X_m X_n) = \tfrac{1}{4}(a+b)^2 \text{ for } m \neq n \in \{1, \ldots, N\}. \tag{82}$$

By defining $p := \mathbb{P}(X_m = a, X_n = a)$, we obtain the following probabilities:

$\mathbb{P}(X_m = x, X_n = y)$		y		
		a	b	$\mathbb{P}(X_m = x)$
x	a	p	$\frac{1}{2} - p$	$\frac{1}{2}$
	b	$\frac{1}{2} - p$	p	$\frac{1}{2}$
$\mathbb{P}(X_n = y)$		$\frac{1}{2}$	$\frac{1}{2}$	

Table 18: Probabilities based on X_m and X_n

By Formula (82) and the following representation

$$\mathbb{E}(X_m X_n) = \sum_{(x,y) \in \{a,b\}^2} xy \cdot \mathbb{P}(X_m = x, X_n = y),$$

we obtain with the probabilities from Table 18 the equation

$$\tfrac{1}{4}(a+b)^2 = p(a^2 + b^2) + 2\left(\tfrac{1}{2} - p\right) ab$$

which has the solution $p = \frac{1}{4}$. This implies that X_m and X_n are independent. □

It cannot be followed that $X_1, \ldots, X_N$ are mutually independent, since the intersection of three or more events is not clearly determined by the assumption of uncorrelated random variables. Thus, counter examples can be constructed.

This result shows that uncorrelated signs (for $a = 1$ and $b = -1$) lead to pairwise independence. This can be an indication that the sign structure cannot give relevant information about dependency structures without having correlations.

7 Generalization approaches based on the K-sign depth for further research

This chapter will illustrate approaches for generalizations of the K-sign depth for further research. In Chapter 7.1, we exchange the sign function by other robust score functions and understand the sign function as a special case. The sign function is outlier robust, but much information can be lost from the data in general. Under some additional assumptions, the generalized approaches have the same asymptotic distribution as the one in Chapter 3. Only an additional unknown scaling factor has to be estimated which is one for the sign function. If the scores are based on ranks, the scaling factor does not depend on the distribution of the errors and can be derived theoretically in advance. However, the asymptotic analysis is more difficult due to dependent random variables then. In Chapter 7.2, we consider Donsker's invariance principle for temporally dependent random variables with short memories. We obtain the same asymptotic distribution with an additional correction factor. The correction factor can be estimated from the data. This leads to the possibility to test models under particularly correlated errors.
The entire chapter should be understood as an outlook highlighting the impact of the previous results for further research. Therefore, the ideas are presented shortly since the execution of each detail is beyond this thesis.

7.1 Generalizations of the K-sign depth

According to Chapter 3, Theorem 3.8, p. 22, the K-sign depth can be represented by a sum of products with the signs of the residuals:

$$\begin{aligned} & N\left(d_K(\mathbf{E}) - \frac{1}{2^{K-1}}\right) \\ =& \frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le n_1<\ldots<n_K\le N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1\le i(1)<\ldots<i(2L)\le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(E_{n_{i(j)}}) \text{ almost surely.} \end{aligned}$$

This representation can be considered as an alternative definition of the K-sign depth under errors which are not zero almost surely. This formula yields positive summands for K-tuples with alternating signs and negative summands otherwise. Further, it delivers a new perspective how the K-sign depth can be generalized or modified. By replacing the sign function ψ with another score function for the residuals, we can extend the class of depth functions.

7.1.1 K-score depth

The sign function regularizes random variables with heavy-tailed distributions such that the moments exist afterwards and the Central Limit Theorem can be applied. Outliers and heteroscedasticity can be handled as well. However, the signs have the drawback that much information is lost compared to the original residuals. Thus, the efficiency towards other methods can suffer drastically. Therefore, we will consider generalizations of the K-sign depth by replacing ψ with score functions which are outlier robust and provide the existence of the moments of the residuals but also contain more information to obtain higher efficiency. This is the typical trade-off between *robustness* and *efficiency* in robust statistics (Maronna et al., 2006, p. 27).

Definition 7.1. *Let* $\mathbf{R} = (R_1, \dots, R_N)^\top$ *be a residual vector. For* $K \in \mathbb{N} \setminus \{1\}$ *and score function* $\psi : \mathbb{R} \to \mathbb{R}$*, we define the* K***-score depth*** *with score* ψ *by*

$$T_{K,\psi}(\mathbf{R}) := \frac{N}{2^{K-1}\binom{N}{K}} \sum_{1 \le n_1 < \dots < n_K \le N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \dots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi(R_{n_{i(j)}}).$$

Note that we use the symbol ψ for a more general class of functions compared to the previous chapters. The class of score functions should be identified such that an asymptotic distribution can be obtained. This question also depends on the distribution of the errors. In the following, we discuss some properties which the errors and ψ have to satisfy simultaneously and how the asymptotic distribution can be derived then. Compared to Assumption 2.1, p. 5, we will need more assumptions on the errors for the K-score depth.

Assumption 7.2. *Let* $E_1, \dots, E_N$ *be random variables in* $\mathbb{R}$ *which satisfy the following conditions:*

$$\psi(E_1), \dots, \psi(E_N) \text{ are independent,} \tag{A1}$$
$$\mathbb{P}(\psi(E_n) = 0) = 0, n = 1, \dots, N, \tag{A2*}$$
$$\psi(E_1), \dots, \psi(E_N) \text{ are identically distributed,} \tag{A3}$$
$$\mathbb{E}(\psi(E_n)) = 0, n = 1, \dots, N, \tag{A4}$$
$$\sigma_\psi^2 := \operatorname{var}(\psi(E_n)) < \infty, n = 1, \dots, N. \tag{A5}$$

Formula (A1) is the same as in Assumption 2.1. If ψ is the sign function, then (A2*) and (A4) imply (A2) from Assumption 2.1. Assuming (A2) implies Formula (A3) and thus (A5) directly if ψ is the sign function. For more general scores, (A3) - (A5) have to be assumed additionally in general.

E.g., for $\psi(x) = x$ and $E_n \sim Cau(0, \gamma n)$, $n = 1, \ldots, N$, the assumptions (A3) - (A5) are violated. The Assumption (A3) is usually inherited by assuming identically distributed errors $E_1, \ldots, E_N$. Assumption (A4) is usually given due to under symmetric errors and scores. For Assumption (A5), the scores have to be robustifying if the second moment of the errors do not exist. Note that each variance should be equal.

Derivation of the asymptotic distribution

The most part of the following derivation is similar to Chapter 3. Therefore, we will only highlight the differences between the derivation from Chapter 3 to keep this chapter as short as possible.

Theorem 7.3. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables and $\psi : \mathbb{R} \to \mathbb{R}$ a score function such that Assumption 7.2 is satisfied. Let $\beta_{K,N} := \frac{N^K K(K-1)}{4\langle N\rangle_K}$ and $\beta_K := \frac{K(K-1)}{4}$. We denote $\mathbf{E} = \mathbf{E}_N = (E_1, \ldots, E_N)^\top$. Then, we have*

$$(a)\ T_{K,\psi}(\mathbf{E}) = -\beta_{K,N} \sum_{1\le n_1\ne n_2\le N} \left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} \psi(E_{n_1})\psi(E_{n_2}) + o_{a.s.}(1),$$

$$(b)\ T_{K,\psi}(\mathbf{E}) = \beta_{K,N}\Bigg(-\sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \Upsilon_{J,N}(S^N_{\psi,\bullet}) - \left(-\frac{1}{2}\right)^{K-2} \left(S^N_{\psi,1}\right)^2 + \frac{1}{N2^{K-2}} \sum_{n=1}^{N} \psi(E_n)^2 \Bigg) + o_{a.s.}(1) \text{ almost surely}$$

with the same notation as in Theorem 3.22, p. 39, and

$$S^N_{\psi,t} = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt\rfloor} \psi(E_n) \text{ for } t \in [0,1].$$

$$(c)\ \frac{T_{K,\psi}(\mathbf{E})}{\sigma^2_\psi} \xrightarrow[N\to\infty]{\mathcal{D}} \beta_K\left(-\sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \Upsilon_J(B_\bullet) - \left(-\frac{1}{2}\right)^{K-2} B_1^2 + \frac{1}{2^{K-2}} \right)$$

where $(B_t)_{t\in[0,1]}$ denotes the standard Brownian motion. For $K \in \{2, 3\}$, the o-terms are equal to zero.

Assertion (a) is the more general version of Theorem 3.17, p. 35, which is also useful for an efficient computation similar to Chapter 4.1 and (b) and (c) are the pendants of Theorem 3.22, p. 39, and Theorem 3.29, p. 50, respectively.

Proof of Theorem 7.3: For (a), we need to prove Lemma 3.9, p. 23, for score functions ψ and errors $E_1, \ldots, E_N$ that satisfy (A1) - (A5) from Assumption 7.2.

Compared to the proof of this lemma, only Formula (22), p. 23, changes to

$$\mathbb{E}\left(\prod_{j=1}^{m}\psi(E_{n_{i(j)}})\psi(E_{\tilde{n}_{i(j)}})\right)=\begin{cases}\sigma_\psi^{2m}, & \text{if } n_{i(j)}=\tilde{n}_{i(j)} \text{ for } j=1,\dots,m,\\ 0, & \text{otherwise,}\end{cases}$$

for arbitrary indices $i(1),\dots,i(m)$ by (A4) and (A5). If ψ is the sign function, we have $\sigma_\psi^{2m}=1$. Since this expected value is constant and does not depend on $n_{i(1)},\dots,n_{i(m)}$ or N, the proof can be continued analogously. The rest of the derivation until Theorem 3.17, p. 35, can then be done completely identically for the remaining proof of (a).
For (b), we can obtain the same result as in Lemma 3.19, p. 36, except for the constant $\frac{1}{2^{K-2}}$ in Formula (36), p. 36:

$$\begin{aligned}&-\frac{1}{N}\sum_{1\le n_1\neq n_2\le N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\psi(E_{n_1})\psi(E_{n_2})\\ =&-\frac{1}{N}\sum_{n_1,n_2=1}^{N}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}\psi(E_{n_1})\psi(E_{n_2})+\frac{1}{N2^{K-2}}\sum_{n=1}^{N}\psi(E_n)^2\end{aligned}$$

since $\psi(E_n)^2$ cannot be simplified further in general. Apart from that, the other results up to Theorem 3.22, p. 39, can be derived identically for (b) by using (A2*).
For proving (c), we are going to use Donsker's invariance principle. According to Assumption 7.2 (A1) and (A3), $\psi(E_1),\dots,\psi(E_N)$ are i.i.d. random variables. Further, (A4) and (A5) imply $\mathbb{E}(\psi(E_n))=0$ and $\operatorname{var}(\psi(E_n))=\sigma_\psi^2$. For Donsker's invariance principle, we need to rescale the variance of the random walk S_t^N to one:

$$\tilde{S}_t^N=\frac{1}{\sigma_\psi\sqrt{N}}\sum_{n=1}^{\lfloor Nt\rfloor}\psi(E_n).$$

Then, Donsker's invariance principle implies

$$(\tilde{S}_t^N)_{t\in[0,1]}\xrightarrow[N\to\infty]{\mathcal{D}}B_\bullet$$

where $B_\bullet$ denotes the standard Brownian motion. According to (b)

$$\begin{aligned}\frac{T_{K,\psi}(\mathbf{E})}{\sigma_\psi^2}=&\beta_{K,N}\Bigg(\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(\tilde{S}_\bullet^N)\\ &-\left(-\frac{1}{2}\right)^{K-2}\left(\tilde{S}_1^N\right)^2+\frac{1}{N2^{K-2}\sigma_\psi^2}\sum_{n=1}^{N}\psi(E_n)^2\Bigg)+o_{a.s.}(1).\end{aligned}$$

Note that $c^2 \Upsilon_{J,N}(f) = \Upsilon_{J,N}(cf)$ for arbitrary $c \in \mathbb{R}$ and $f \in D[0,1]$ is used here. Since $\beta_{K,N} \xrightarrow[N\to\infty]{} \beta_K$ and by the strong law of large numbers:

$$\frac{1}{N} \sum_{n=1}^{N} \psi(E_n)^2 \xrightarrow[N\to\infty]{} \sigma_\psi^2 \text{ almost surely.}$$

Therefore, the assertion follows analogously to the proof of Theorem 3.29, p. 50. □

Estimation of the variance factor σ_ψ^2

The variance σ_ψ^2 is unknown in general and has to be estimated for a test based on the K-score depth. For the sign function ψ_{sgn}, this is not necessary since $\sigma_{\psi_{\text{sgn}}}^2 = 1$. Some estimators for σ_ψ^2 are listed in Table 19. The interquartile range (IQR) and

estimator	**relative efficiency**	**breakdown point**
empirical standard deviation	1	$\frac{1}{N}$
IQR	0.37	$\frac{1}{N}\left\lceil \frac{N}{4} \right\rceil$
MAD	0.37	$\frac{1}{N}\left\lfloor \frac{N-1}{2} \right\rfloor$
Q-estimator	0.82	$\frac{1}{N}\left\lfloor \frac{N}{2} \right\rfloor$

Table 19: Overview of several estimators for scaling compared with their relative efficiency to the empirical standard deviation (under i.i.d. normally distributed random variables) and their breakdown points.

median of the absolute derivations from the median (MAD) are multiplied by $\frac{1}{2\Phi^{-1}(\frac{3}{4})}$ or $\frac{1}{\Phi^{-1}(\frac{3}{4})}$, respectively, for a consistent estimation under i.i.d normally distributed random variables (Maronna et al., 2006, p. 33). Here, Φ denotes the distribution function of the standard normal distribution. The Q-estimator is introduced by Rousseeuw and Croux (1993) and based on the 25%-quantile of the pairwise absolute differences of the data. As the table shows, the Q-estimator is an improved estimator in both categories *efficiency* and *robustness*. The `R`-function `Qn()` from the package `robustbase` provides a linear implementation of the Q-estimator. For a consistent estimation under the normal distribution, the Q-estimator should be corrected with a factor that depends on the sample size with asymptotic value $\frac{1}{\sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right)}$.
The *relative efficiency* of two unbiased estimators T_1 and T_2 is defined by the quotient of the variances of the considered estimators and taking its limit for increasing sample sizes if it exists (Staudte and Sheather, 1990, p. 74):

$$\frac{\text{var}(T_2(X_1, \ldots, X_N))}{\text{var}(T_1(X_1, \ldots, X_N))} \xrightarrow[N\to\infty]{} e(T_1, T_2).$$

In Table 19, the estimator T_2 is always the empirical standard deviation since it is the best estimator for σ^2 under the normal distribution. Under this situation, $\operatorname{var}(T_2(X_1,\ldots,X_N)) < \operatorname{var}(T_1(X_1,\ldots,X_N))$ and $e(T_1,T_2) \in [0,1]$ if the limit exists. The closer $e(T_1,T_2)$ is to one, the closer the performance of estimator T_1 is to the empirical standard deviation under normally distributed random variables. The *breakdown point* for a scale estimation is defined by the minimum of the explosion and implosion point of the estimator (Maronna et al., 2006, p. 59). The explosion or implosion point are defined by the minimal relative number of data we have to modify arbitrarily such that the estimation can tend to ∞ or zero, respectively. For obtaining an explosion, several data points are modified to reach high values tending to infinity. For an implosion, multiple data points are usually chosen to have the same values since many scale estimators are zero then. Note that $\frac{1}{N}\left\lfloor \frac{N}{2} \right\rfloor$ is the highest breakdown point for scale-equivariant and location-invariant estimators.
In order to estimate σ_ψ^2, we propose considering the scored residuals under a well fitting parameter estimated in advance, similar to the MM-estimation (Yohai, 1987). This estimation can be based on the maximal K-score depth or on another robust estimator. Since the residuals should be robustified by applying the score function, the empirical standard deviation for the residuals can be used. However, the estimation of σ_ψ^2 can have two problems. If the model class is too simple then no parameter could explain the model reasonably well and σ_ψ^2 can be overestimated. The second problem is a possibly too high type-I-error of the asymptotic test for too small N since the estimation of σ_ψ^2 is not considered in the quantiles. If we assume that the variance factor is estimated consistently and the correct model class is chosen, then we have an asymptotic test which controls the level for sufficiently high N.

Efficient computation of the K-score depth

The computation of the K-score depth by Definition 7.1 leads to an algorithm with time complexity $\Theta(N^K)$. As in Lemma 4.1, p. 56, with the K-sign depth, we use Theorem 7.3 (b) for an approximation of the depth with lesser computational costs.

Lemma 7.4. *Let* $S_{n,\psi,\alpha}^N = \sum_{k=1}^n \left(\frac{k}{N}\right)^\alpha \psi(R_n)$ *for* $n \in \{1,\ldots,N\}$, $\alpha \geq 0$ *and let* $\mathbf{R} = (R_1,\ldots,R_N)^\top$ *such that* $\mathbb{P}(\psi(R_n) \neq 0) = 1$ *for* $n \in \{1,\ldots,N\}$. *We define* $\beta_{K,N} := \frac{N^{K-1}K(K-1)}{2\langle N\rangle_K}$. *Then*

$$T_{K,\psi}(\mathbf{R}) = -\beta_{K,N} \sum_{j=0}^{K-2} \binom{K-2}{j} \sum_{n=2}^{N} \left(\frac{1}{2} - \frac{n}{N}\right)^{K-2-j} \psi(R_n) S_{n-1,\psi,j}^N + o_{a.s.}(1)$$

almost surely. For $K \in \{2,3\}$, *the o-term is exactly zero.*

Lemma 7.4 leads to an algorithm in linear time for the K-score depth. The proof is identical to the proof of Lemma 4.1, p. 56. The only difference is that the proof of Lemma 7.4 is based on Theorem 3.17, p. 35, instead of Theorem 7.3 (b).

Examples of score functions

The choice of a score function can be focused on its robustness and also on the level of information it includes from the residual vector. Two extreme scores are the sign function and the identity:

$$\begin{aligned}\psi_{\text{sgn}}(x) &:= \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\},\\ \psi_{\text{id}}(x) &:= x.\end{aligned}$$

An idea for a better trade-off between these two functions is the Huber-score.

Example 7.5. *For some tuning constant $b > 0$, we define the **Huber-score** as*

$$\psi_{\text{Hub}}(x) := \begin{cases} x, & \text{for } |x| \leq b,\\ b \cdot \psi_{\text{sgn}}(x), & \text{for } |x| > b.\end{cases}$$

Errors with larger absolute values than b can be understood as outliers. Therefore, the influence of large outliers is reduced by this class of functions. Additionally, the main information of the residuals is included for the estimation. This score function is inspired by the Huber-M-estimator and has the same form as the weight function from the Huber-M-estimation (Maronna et al., 2006, p. 26).

Note that the score functions ψ_{sgn} and ψ_{id} are also considered as weight functions for M-estimators corresponding to the ordinary-least-squares estimator or the least-absolute distances, respectively (Maronna et al., 2006, p. 23). Other weight functions, such as in the Hampel-M-estimator, can be considered further. Note that some of them set residuals to zero. This property has to be avoided since Theorem 7.3 assumes that no residuals are equal to zero almost surely. E.g., the function can be modified or Theorem 7.3 has to be extended for residuals equal to zero otherwise. A crucial point using these scores is the choice of the tuning constants, e.g., b for the Huber-score. A detailed discussion of this choice is beyond this thesis but there exists research for similar situations, e.g., Kelly (1992, 1996). In the literature, rescaling the residuals is often done before applying some robustifying function, i.e.:

$$T_{K,\psi}(\mathbf{E}) = \frac{N}{2^{K-1}\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_k \leq N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi\left(\frac{E_{n(j)}}{\sigma}\right)$$

and σ is a scale parameter of the density function of $E_1, \dots, E_N$ (Maronna et al., 2006, p. 38). This formulation by a scale parameter avoids particularly to assume the existence of the second moment of the errors and general classes of distributions can still be considered. This approach can lead to easier rules for the choice of tuning constants which should be discussed in further research as well.

Further ideas for the K-score depth

The type-I-error of the K-score depth test should be studied since it only can be used with the asymptotic quantiles. By simulation studies, the necessary sample size N should be derived. Such cut off values for a sufficiently high sample size can depend on K (cf. Chapter 6.3.1) but also on the chosen score function and the estimation of σ_ψ^2. A detailed focus on the estimation of σ_ψ^2 should be discussed as well. For obtaining a faster approximation of the asymptotic distribution, the paths of the random walk $(S_t^N)_{t\in[0,1]}$ can be linearly interpolated before evaluating them by the functional. According to Donsker's invariance principle for linearly interpolated random walks, this statistic has the same asymptotic distribution (Billingsley, 1999, p. 90). The results from Chapter 6.2 can be helpful for considering the linearly interpolated versions since Lemma 7.4 cannot be used for the computation.

7.1.2 K-rank depth

Instead of choosing deterministic score functions, we can consider the *signed ranks* of the absolute residuals. This has the advantage that we do not have to estimate a scaling factor and have a *distribution free* test statistic as well so that the exact distribution can be computed. In particular, we consider for $\mathbf{R} = (R_1, \dots, R_N)^\top$:

$$rk_n^N(\mathbf{R}) := rk(R_n) := \tfrac{1}{N+1} \textstyle\sum_{j=1}^N \mathbb{1}\{|R_j| \le |R_n|\} \text{ for } n = 1, \dots, N. \qquad (83)$$

If N and $\mathbf{R}$ are fixed, we shortly write $rk(R_n)$. Note that the division by $N+1$ yields values in $(0,1)$. This avoids that the ranks tend to ∞ for $N \to \infty$.

Definition 7.6. *Let $\mathbf{R} = (R_1, \dots, R_N)^\top$ be a residual vector. For $K \in \mathbb{N} \setminus \{1\}$ and score $\varphi : (0,1) \to [0,\infty)$, we define the K-**rank depth** with score φ by*

$$\begin{aligned} &T_{K,\varphi}^{rk}(\mathbf{R}) \\ :=&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1\le n_1<\dots<n_K\le N} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1\le i(1)<\dots<i(2L)\le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi_{\mathrm{sgn}}(R_{n_{i(j)}}) \varphi(rk(R_{n_{i(j)}})). \end{aligned}$$

The function φ is a score function for the ranks. For $\varphi(u) = u$, we have the original ranks also known as the *Wilcoxon ranks*. Moreover, the *van der Waerden ranks* are defined by $\varphi(u) = \Phi^{-1}(\frac{u+1}{2})$ where Φ is the standard normal distribution function (Antille, 1979). Note that the van der Waerden ranks could not be applied if Formula (83) was rescaled by $\frac{1}{N}$ since $\lim_{x \to 1} \Phi^{-1}(x) = \infty$. Rank tests based on the van der Waerden scores often show a good performance under normally distributed errors and have also robust properties under other distributions (Hodges and Lehmann, 1961). Moreover, we can set $\varphi(u) \equiv 1$ to obtain the original K-sign depth. This K-rank depth can also be robust. Compared to the K-sign depth, the K-rank depth has the advantage that it contains more information of the residual vector such as the monotony structure. If a K-tuple has no alternating signs, the penalty is higher for absolutely larger values. Moreover, it has also the advantage that the distribution of the K-rank depth is distribution free under the true parameter and some additional assumptions compared to deterministic scores from Chapter 7.1.1. However, the current form in Definition 7.6 has to be simplified further since the time complexity is $\Theta(N^K)$. Before computational simplifications can be considered, we introduce some assumptions.

Assumption 7.7. *Let $E_1, \ldots, E_N$ be random variables in $\mathbb{R}$ which satisfy the following conditions:*

$$E_1, \ldots, E_N \text{ are independent,} \tag{A1}$$

$$\mathbb{P}(E_n = 0) = 0, n = 1, \ldots, N, \tag{A2*}$$

$$E_1, \ldots, E_N \text{ are identically distributed,} \tag{A3}$$

$$E_1, \ldots, E_N \text{ are symmetric around } 0. \tag{S}$$

Compared to Assumption 7.2, the symmetry assumption (S) is given, but (A4) and (A5) are not needed due to the ranks. Further, (S) and (A2*) imply (A2) in Assumption 2.1. Under Assumption 7.7, the ranks from Formula (83) and the sign of the residuals are independent as the next lemma shows:

Lemma 7.8. *Let $E_1, \ldots, E_N$ be random variables satisfying Assumption 7.7. Then*

$$(\psi_{\text{sgn}}(E_1), \ldots, \psi_{\text{sgn}}(E_N)) \text{ is stochastically independent from } (rk(E_1), \ldots, rk(E_N))$$

where $rk(E_n)$ denotes the rank of E_n of the absolute values from $E_1, \ldots, E_N$ as defined in Formula (83).

Proof of Lemma 7.8: First, consider E_n and $a \in \mathbb{R}$ and prove that $rk(E_n)$ and $\psi_{\text{sgn}}(E_n)$ are independent. Due to the symmetry property (S) of E_n, we have

$$\mathbb{P}(0 < E_n \leq a) = \mathbb{P}(-a \leq E_n < 0).$$

From (A2*), we obtain $\mathbb{P}(|E_n| \leq a) = \mathbb{P}(-a \leq E_n < 0) + \mathbb{P}(0 < E_n \leq a)$ and thus

$$\mathbb{P}(|E_n| \leq a) = 2 \cdot \mathbb{P}(0 < E_n \leq a) \quad \Leftrightarrow \quad \frac{1}{2} \cdot \mathbb{P}(|E_n| \leq a) = \mathbb{P}(0 < E_n \leq a).$$

Assumption (A2*) and (S) imply $\mathbb{P}(\psi_{\text{sgn}}(E_n) = 1) = \frac{1}{2}$ so that

$$\mathbb{P}(|E_n| \leq a, \psi_{\text{sgn}}(E_n) = 1) = \mathbb{P}(0 < |E_n| \leq a) = \mathbb{P}(|E_n| \leq a)\mathbb{P}(\psi_{\text{sgn}}(E_n) = 1). \quad (84)$$

Formula (84) can be shown for the event $\psi_{\text{sgn}}(E_n) = -1$ similarly by applying $\mathbb{P}(E_n \geq a) = \mathbb{P}(E_n \leq -a)$. Therefore, $|E_n|$ and $\psi_{\text{sgn}}(E_n)$ are independent. Assumption (A1) yields that $(|E_1|, \ldots, |E_N|)$ and $(\psi_{\text{sgn}}(E_1), \ldots, \psi_{\text{sgn}}(E_N))$ are independent as well. Since $(rk(E_1), \ldots, rk(E_N))$ is a measurable function of $(|E_1|, \ldots, |E_N|)$, the assertion follows. □

Note that Lemma 7.8 does not hold if the normal ranks of $R_1, \ldots, R_N$ are considered (i.e., not based on the absolute values) or the symmetry property (S) is not given.

Theorem 7.9. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 7.7, $\varphi : (0,1) \to [0,\infty)$ and $\beta_{K,N} := \frac{N^K K(K-1)}{4\langle N\rangle_K}$. We denote $\mathbf{E} = \mathbf{E}_N = (E_1, \ldots, E_N)^\top$. Assume that for each $m \in \{4, \ldots, K\}$ a sequence $(\gamma_N)_{N\in\mathbb{N}}$ exists such that*

$$\begin{aligned} &\varphi\left(\frac{n}{N+1}\right) \leq \gamma_N \text{ for } n = 1, \ldots, N, \\ &\text{with } \gamma_N^{2m} = o\left(N^{m-2}\right). \end{aligned} \quad \text{(R1)}$$

Then

$$(a)\ T_{K,\varphi}^{rk}(\mathbf{E}) = \beta_{K,N} \sum_{1\leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_1 - n_2|}{N}\right)^{K-2} \prod_{i=1}^{2} \psi_{\text{sgn}}(E_{n_i})\varphi(rk(E_{n_i})) + o_{a.s.}(1),$$

$$(b)\ T_{K,\varphi}^{rk}(\mathbf{E}) = \beta_{K,N}\left(-\sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \Upsilon_{J,N}(S_{rk,\varphi,\bullet}^N) \right.$$
$$\left. - \left(-\frac{1}{2}\right)^{K-2} \left(S_{rk,\varphi,1}^N\right)^2 + \frac{1}{N2^{K-2}} \sum_{n=1}^{N} \varphi(rk_n^N(\mathbf{E}_N))^2 \right) + o_{a.s.}(1)$$

with similar notations as in Theorem 3.22, p. 39, and

$$S^N_{rk,\varphi,t} = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \psi_{\text{sgn}}(E_n)\varphi(rk^N_n(\mathbf{E}_N)) \text{ for } t \in [0,1].$$

For $K \in \{2,3\}$, the o-terms are equal to zero.

Proof of Theorem 7.9: For (a), we need to modify the proof of Lemma 3.9, p. 23, for the signed ranks with score φ. Formula (22), p. 23, changes to

$$\begin{aligned} &\mathbb{E}\left(\prod_{j=1}^{m} \psi_{\text{sgn}}(E_{n_{i(j)}})\varphi(rk(E_{n_{i(j)}}))\psi_{\text{sgn}}(E_{\tilde{n}_{i(j)}})\varphi(rk(E_{\tilde{n}_{i(j)}}))\right) \\ &\leq \begin{cases} \gamma_N^{2m}, & \text{if } n_{i(j)} = \tilde{n}_{i(j)} \text{ for } j = 1,\dots,m, \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{85}$$

almost surely for arbitrary indices $i(1),\dots,i(m)$. For the first case in Formula (85), the inequality (R1) is used. The second case is also an equality and follows from the independence of the signs and signed ranks, cf. Lemma 7.8. A deterministic upper bound for the expected value exists for each m which does not depend on the indices $n_{i(1)},\dots,n_{i(m)}$. Therefore, the proof can be continued similarly to Lemma 3.9. The difference is here that the upper bound may depend additionally on N and thus, we will need $\gamma_N^{2m} = o(N^{m-2})$ in (R1) here. Based on the notation from Lemma 3.9, we have $M = K$ and $B = 1 - K$. Analogously to the proof of Lemma 3.9:

$$\begin{aligned} &\text{var}\left(\sum_{1\leq n_1<\ldots<n_K\leq N} \mathcal{O}(N^{1-K}) \prod_{j=1}^{m} \psi_{\text{sgn}}(E_{n_{i(j)}})\varphi(rk(E_{n_{i(j)}}))\right) \\ \leq &\mathcal{O}(N^{2-2K}) \sum_{\substack{1\leq n_1<\ldots<n_K\leq N \\ 1\leq \tilde{n}_1<\ldots<\tilde{n}_K\leq N}} \gamma_N^{2m} \cdot \mathbb{1}_{\mathcal{I}_{i(1),\ldots,i(m)}}(n_1,\dots,n_M,\tilde{n}_1,\dots,\tilde{n}_M) \end{aligned}$$

where $\mathcal{I}_{i(1),\ldots,i(m)}$ denotes the set of $2M$-tuples with $n_{i(j)} = \tilde{n}_{i(j)}$ for $j \in \{1,\dots,m\}$ according Formula (85). The remaining proof in Lemma 3.9 can be done identically up to Formula (24):

$$\begin{aligned} &\mathcal{O}(N^{2-2K}) \sum_{\substack{1\leq n_1<\ldots<n_K\leq N \\ 1\leq \tilde{n}_1<\ldots<\tilde{n}_K\leq N}} \gamma_N^{2m} \cdot \mathbb{1}_{\mathcal{I}_{i(1),\ldots,i(m)}}(n_1,\dots,n_M,\tilde{n}_1,\dots,\tilde{n}_M) \\ \overset{(24)}{\leq} &\mathcal{O}(N^{2-2K})\mathcal{O}(N^{2(K-m)})\mathcal{O}(N^m)\gamma_N^{2m} = \mathcal{O}(N^{2-m})\gamma_N^{2m} \xrightarrow[N\to\infty]{\text{(R1)}} 0. \end{aligned}$$

The last convergence is given if and only if (R1) is correct. The rest of the derivation up to Theorem 3.17, p. 35, can be continued identically for proving (a).
For (b), we can do completely the same derivation as in Chapter 3 using (A2*).

Compared to Theorem 3.22, p. 39, only the factor $\sum_{n=1}^{N} \varphi(rk(E_n))^2$ changes, since it cannot be simplified further. □

The condition (R1) in Theorem 7.9 is discussed in the next example.

Example 7.10. *The condition (R1) in Theorem 7.9 is satisfied for the Wilcoxon ranks with score $\varphi(u) = u$ and for the van der Waerden ranks with score $\varphi(u) = \Phi^{-1}(\frac{u+1}{2})$ where Φ is the standard normal distribution function.*

(a) For the Wilcoxon ranks, let $\gamma_N \equiv 1$ since $rk(e_n) \leq \frac{N}{N+1} < 1$ for arbitrary vectors $(e_1, \ldots, e_N)^\top$ and $n = 1, \ldots, N$.

(b) For the van der Waerden ranks, we can set $\gamma_N = \log(N+1)$ which is an asymptotic upper bound for the quantile function of the standard normal distribution. For proving this assertion, we will show for sufficiently large N

$$\Phi^{-1}\left(\frac{1+\frac{N}{N+1}}{2}\right) \leq \log(N+1) \quad \Leftrightarrow \quad 0 \leq 2\Phi(\log(N+1)) + \frac{1}{N+1} - 2. \tag{86}$$

Let $f : (0, \infty) \to \mathbb{R}, f(N) = 2\Phi(\log(N+1)) + \frac{1}{N+1} - 2$ and take its derivative:

$$\begin{aligned} f'(N) &= 2\Phi'(\log(N+1))\frac{1}{N+1} - \frac{1}{(N+1)^2} \stackrel{!}{=} 0 \\ &\Rightarrow \; 2\Phi'(\log(N+1)) \stackrel{!}{=} \frac{1}{N+1}. \end{aligned}$$

After substituting $x = \log(N+1)$, we obtain:

$$2\Phi'(x) \stackrel{!}{=} \exp(-x).$$

By solving a quadratic equation, we obtain the following solutions:

$$x_{1,2} = 1 \pm \sqrt{1 + \log\left(\tfrac{2}{\pi}\right)}$$

and obtain after resubstituting:

$$N_{1,2} = \exp\left(1 \pm \sqrt{1 + \log\left(\tfrac{2}{\pi}\right)}\right) - 1.$$

According to the previous calculation, f can only have for N_1 or N_2 local extrema. The solution $N_1 = \exp\left(1 + \sqrt{1 + \log\left(\frac{2}{\pi}\right)}\right) - 1$ is larger than N_2. Moreover, we can compute $f(N_1) \approx 0.094 > 0$ in R*. Since $\lim_{N\to\infty} f(N) = 0$ and no other extremum occurs after N_1, we obtain $f(N) \geq 0$ for $N \geq N_1$ which*

implies Formula (86) for sufficiently large N. Figure 54 presents the graph of the function f in order to follow the previous argumentation better.

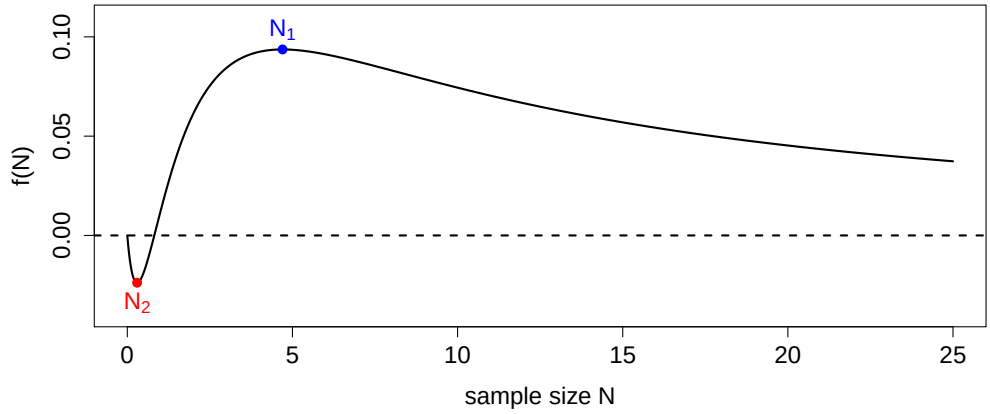

Figure 54: The graph of the function $f(N) = 2\Phi(\log(N+1)) + \frac{1}{N+1} - 2$

Completely analogously to Lemma 7.4, in Chapter 7.1, a representation for the K-rank depth in linear time can be derived.

Lemma 7.11. *Let $S^N_{n,rk,\varphi,\alpha} = \sum_{k=1}^{n} \left(\frac{k}{N}\right)^{\alpha} \psi_{\text{sgn}}(R_n)\varphi(rk(R_n))$ for $n \in \{1,\dots,N\}$, $\alpha \geq 0$ and let $\mathbf{R} = (R_1,\dots,R_N)^\top$ such that $\mathbb{P}(R_n = 0) = 0$ for $n \in \{1,\dots,N\}$. We define $\beta_{K,N} := \frac{N^{K-1}K(K-1)}{2\langle N\rangle_K}$. Then*

$$T^{rk}_{K,\varphi}(\mathbf{R}) = -\beta_{K,N} \sum_{j=0}^{K-2} \binom{K-2}{j} \sum_{n=2}^{N} \left(\frac{1}{2} - \frac{n}{N}\right)^{K-2-j} \times \psi_{\text{sgn}}(R_n)\varphi(rk(R_n)) S^N_{n-1,rk,\varphi,j} + o_{a.s.}(1)$$

almost surely. For $K \in \{2,3\}$, the o-term is exactly zero.

Testing, ties and the asymptotic distribution

In the context of statistics based on ranks, it is often assumed that $E_1,\dots,E_N$ have no ties almost surely in the literature (i.e., there does not exist any equal value between the errors). This holds in particular for a continuous error distribution. Moreover, this assumption is useful when applying the K-rank depth since the statistic is then *distribution free* under the true parameter. Thus, we can simulate the exact distribution similar to the way described in Chapter 6.2 for the K-sign depth by considering all $N! \cdot 2^N$ combinations for permutations of $(1,\dots,N)$ component wise multiplied by all vectors in $\{1,-1\}^N$ for the residuals.

However, ties can occur in real data applications because the data is actually discrete or due to restrictions of the measuring accuracy (Gibbons and Chakraborti, 2003, p. 194). For the computation of Formula (83) and Definition 7.6, ties can be handled without problems. The literature offers several of procedures such as considering the mean of the ranks with ties (midranks) or by randomizing the order. Note that the distribution under the null hypothesis is affected by using midranks.
For the discussions of the asymptotic analysis, we consider Theorem 7.9. First of all, we have to ensure that the following sum converges as $N \to \infty$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \varphi(rk_n^N(\mathbf{E}_N))^2 =: \sigma_\varphi^2 \tag{R2}$$

such that the corresponding part in Theorem 7.9 (b) can be handled asymptotically. For the Wilcoxon ranks and the van der Waerden ranks, (R2) is satisfied as the next example shows.

Example 7.12. *The condition (R2) is satisfied for the (a) Wilcoxon ranks with score $\varphi(u) = u$ with $\sigma_\varphi^2 = \frac{1}{3}$ and (b) for the van der Waerden ranks with score $\varphi(u) = \Phi^{-1}(\frac{u+1}{2})$ with $\sigma_\varphi^2 = 1$ when no ties occur almost surely.*

(a) For the Wilcoxon ranks, the sum in (R2) can be simplified to a polynomial of degree 3 according to Faulhaber's formula (Knuth, 1993):

$$\frac{1}{N} \sum_{n=1}^{N} \varphi(rk_n^N(\mathbf{E}_N))^2 = \frac{1}{N(N+1)^2} \sum_{n=1}^{N} n^2 = \frac{\frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}}{N(N+1)^2} \xrightarrow[N\to\infty]{} \frac{1}{3}.$$

(b) For the Van der Waerden ranks, we consider the limit of the sum as an integral:

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \varphi(rk_n^N(\mathbf{E}_N))^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Phi^{-1}\left(\frac{1+\frac{n}{N+1}}{2}\right)^2 \\ &= \int_0^1 \Phi^{-1}\left(\frac{1+u}{2}\right)^2 du = 2\int_{\frac{1}{2}}^1 \Phi^{-1}(v)^2\, dv = 2\int_0^\infty w^2 \Phi'(w)\, dw = 1. \end{aligned}$$

Note that the substitutions $u := 2v - 1$ and $v := \Phi(w)$ are used for the derivation. Furthermore, $\int_0^\infty w^2\Phi'(w)\, dw = \frac{1}{2}\mathbb{E}(Z^2)$ for $Z \sim \mathcal{N}(0,1)$ due to the symmetry around 0 of the standard normal distribution.

Example 7.12 (b) shows that (R2) implies the quadratic integrability of φ on $[0,1]$. Further, the asymptotic behavior of $\Upsilon(S^N_{rk,\varphi,\bullet})$ should be discussed. An asymptotic

distribution cannot be derived directly since the ranks are correlated and Donsker's invariance principle cannot be applied in its standard version as it is done in Theorem 3.29, p. 50, or Theorem 7.3 (c), p. 150.
Sen (1974), Sen (1981), and Hájek et al. (1999) provide a theory of invariance principles based on several type of ranks. However, the provided invariance principles have not exactly the form of $S^N_{rk,\varphi,\bullet}$ in Theorem 7.9 (b). An asymptotic analysis should be content for future research for proving the following conjecture:

Conjecture 7.13. *Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of random variables satisfying Assumption 7.7 and let $\varphi : (0,1) \to [0,\infty)$ be a score function that satisfies (R1) and (R2). We denote $\mathbf{E} = \mathbf{E}_N = (E_1, \ldots, E_N)^\top$. Then, we have*

$$\frac{T^{rk}_{K,\varphi}(\mathbf{E})}{\sigma^2_\varphi} \xrightarrow[N\to\infty]{\mathcal{D}} \beta_K \left(- \sum_{J=1}^{K-2} \binom{K-2}{J} \left(-\frac{1}{2}\right)^{K-2-J} \Upsilon_{J,N}(B_\bullet) - \left(-\frac{1}{2}\right)^{K-2} (B_1)^2 + \frac{1}{2^{K-2}} \right)$$

where σ^2_φ is defined in Formula (R2) and $(B_t)_{t\in[0,1]}$ is the standard Brownian motion.

If the convergence $\frac{1}{\sigma_\varphi} S^N_{rk,\varphi,\bullet} \xrightarrow[N\to\infty]{\mathcal{D}} B_\bullet$ with respect to the Skorokhod topology is proven, then Conjecture 7.13 follows directly. Dividing $T^{rk}_{K,\varphi}$ by the variance factor σ^2_φ has also been done for other invariance principles to have increments with variance 1 as in Sen (1981). The assumptions on φ may need to be fixed with more restrictions.

Remark 7.14. *A test based on Conjecture 7.13 for $K = 2$, $\varphi(u) = u$ and the model*

$$Y_n = \mu + E_n \text{ for } n = 1, \ldots, N$$

is equivalent to the one-sample Wilcoxon signed-rank test for testing the value of the median $\mu \in \mathbb{R}$ since both test statistics sum up the signed rank of the absolute valued data. The asymptotic result of this conjecture is especially the same as for the Wilcoxon signed-rank statistic (Gibbons and Chakraborti, 2003, p. 197-203). Thus, the K-rank depth tests can be understood as a generalization of the Wilcoxon signed-rank test as for the K-sign depth and the sign test.

7.1.3 Simulation study for the generalized depth approaches

The approaches in Chapter 7.1.1 and 7.1.2 will be compared in a short simulation study with the F-test for linear models and a robust Wald-test based on the MM-estimator (Yohai, 1987). We consider the model

$$Y_n = \theta \cdot x_n + E_n \text{ for } n = 1, \ldots, 25 \tag{87}$$

for equidistant $x_1, \ldots, x_{25}$ in the interval $[-3, 3]$. The errors are i.i.d. with distribution $E_n \sim \mathcal{N}(0, 4)$ or $E_n \sim Cau(0, 1)$. For each parameter $\theta \in \{-1, -0.95, \ldots, 2.95, 3\}$ (for normal distribution) or $\theta \in \{-3, -2.95, \ldots, 4.95, 5\}$ (for Cauchy distribution), 1,000 repetitions are considered to compute the power of testing $H_0 : \theta = 1$.
In Figure 55, we compare the resulting power functions of the 3-sign depth from Definition 2.2, the 3-score depth with Huber-scores from Definition 7.1 and Example 7.5 with tuning constants $b = 3$ and the 3-rank depths in Definition 7.6 for the Wilcoxon ranks and van der Waerden ranks. For the distribution free depth notions, the exact distributions are computed by 1,000,000 repetitions of samples under the required assumptions. The 3-score depth test with Huber-scores uses the asymptotic quantiles. The true scaling factors $\sigma^2_{\psi_{\text{Hub}}} \in \{1.75, 1.7\}$ (for the particular normal and Cauchy distribution, respectively) under H_0 are used in order to see the performance of the score-depth tests without bias due to the scale estimation. This scaling factor is obtained by a simulation study of 1,000,000 repetitions of the empirical standard deviation of scored residuals under H_0. (For real data application, an estimation is necessary in advance.) For the depths from Chapter 7, an exact or asymptotic test can be completely analogously constructed as in Theorem 6.1. The MM-estimator and the covariance matrix for the robust Wald-test are computed by `lmRob` from the package `robust` using the default settings (Wang et al., 2020).

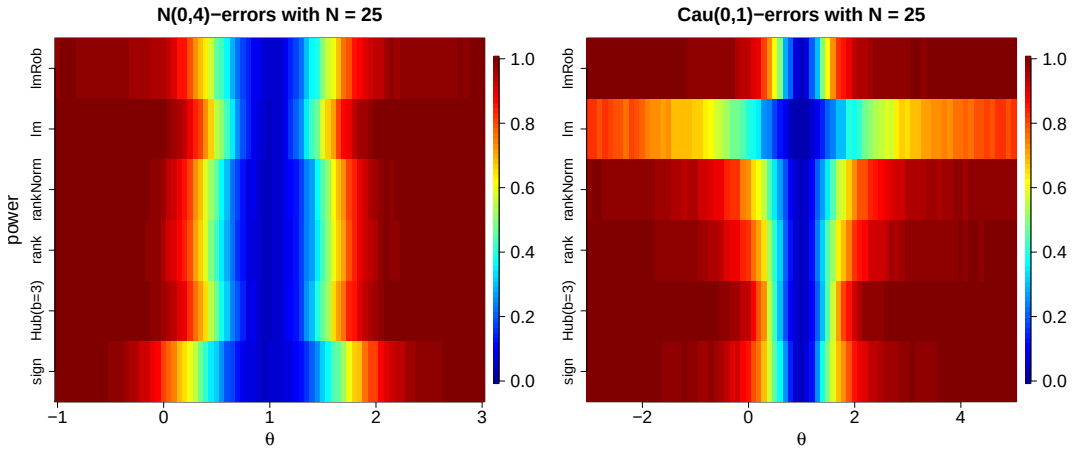

Figure 55: A comparison of the power functions testing $H_0 : \theta = 1$ among the F-test (`lm`), the robust Wald-test (`lmRob`), the sign depth test (`sign`), the Huber-score depth test with tuning constant $b = 3$ (`Hub(b=3)`) and the rank depth tests based on the Wilcoxon ranks (`rank`) and the van der Waerden ranks (`rankNorm`)

The F-test delivers the best power under the normal distribution and the worst power under the Cauchy distribution among the considered tests. The robust Wald-test has a slightly lower power than the Huber-score depth test under the normal

distribution and a slightly higher power under the Cauchy distribution. In each case, the Huber-score depth test is better than the other depth tests. However, the power of the Wilcoxon rank depth is still very high in both cases. The van der Waerden rank depth is under normality a bit better than the Wilcoxon ranks but struggles more under the Cauchy distribution. The sign depth performs the worst under the normal distribution but is better than the rank approaches under the Cauchy distribution. Note also that the optimization of the M-estimator has not always converged using the default settings which could be tuned to obtain higher power values.
Based on this results, the Huber-score depth and the rank depth (as a distribution free statistic) show the best performance among the depth tests. The Huber-score depth has the drawback that the variance has actually to be estimated which has not been done here. Therefore, the comparison with the Wald-test, which estimates the variance as well, should only be understood as a demonstration of the potential of the Huber-score depth but is in the current execution not fair. The construction of a detailed estimation procedure of σ_ψ^2 will be investigated in further research. This drawback can be compensated by the rank depth which performs consistently well.
In a second simulation study, we consider the model

$$Y_n = \theta_1 x_n + \theta_2 x_n^2 + E_n \text{ for } n = 1, \ldots, 25$$

with $\theta_1 = 1$ fixed in all scenarios and $\theta_2 \in \{-3, -2.95, \ldots, 2.95, 3\}$. However, we still

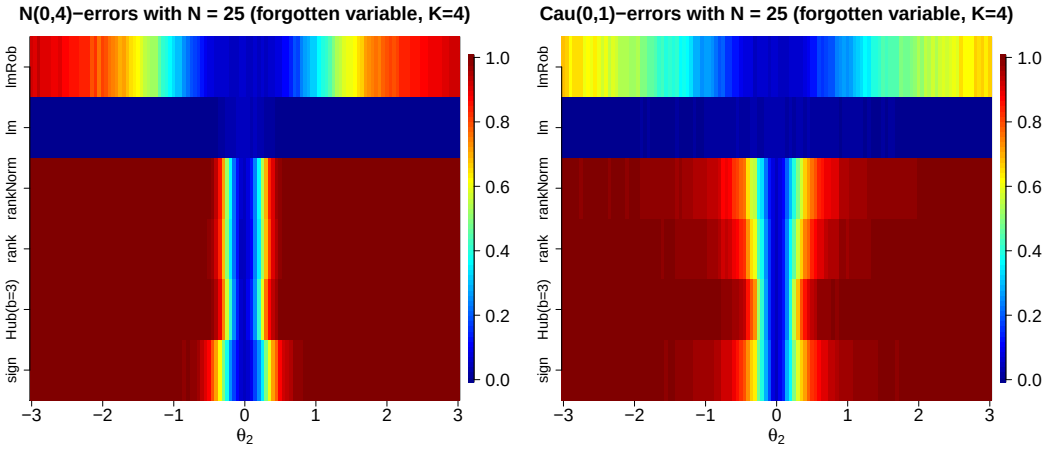

Figure 56: A comparison of the power functions testing $H_0 : \theta_1 = 1$ under varying θ_2 among the F-test (`lm`), the robust Wald-test (`lmRob`), the sign depth test (`sign`), the Huber-score depth test with tuning constant $b = 3$ (`Hub(b=3)`) and the rank depth tests based on the Wilcoxon ranks (`rank`) and the van der Waerden ranks (`rankNorm`) for $K = 4$

assume that the model in Formula (87) is correct and test $H_0 : \theta_1 = 1$ which can

be understood as an underfitted model or a forgotten variable. For the depths, we consider the cases $K = 4$. The other settings are the same as in the previous case. Figure 56 shows that the depth function are very sensitive for rejecting the model. In contrast, the robust Wald-test rejects only under absolutely higher values of θ_2 and the F-test never rejects the model. The F-test and robust Wald-test are based on comparing the suggested (aspect of the) parameter with an estimation of it. Moreover, the estimated covariance matrix depends on the suggest model class which can lead to overestimated variances under forgotten variables. The estimation for θ_1 is here especially for the F-test always very close to one so that no rejection happen for varying θ_2 from zero. Compared to the F-test, the Wald-test rejects the null hypothesis because the estimation of θ_1 gets unstable for a strong variation of θ_2. On the other hand, the depth tests only depend on the residuals so that actually $H_0 : \theta_1 = 1 \wedge \theta_2 = 0$ is tested although we do not know that other parameters can be given possibly. This behavior seems to be more fitting for tests which should check the complete model and not only the parameter. An interesting point here is that we have *parametric* assumptions on the model but we can test hypotheses in a *nonparametric* manner. For application, this knowledge is very important in order to know what rejecting H_0 actually implies.
Figure 57 shows the noticeably worse results for $K = 3$. According to Chapter 6.4,

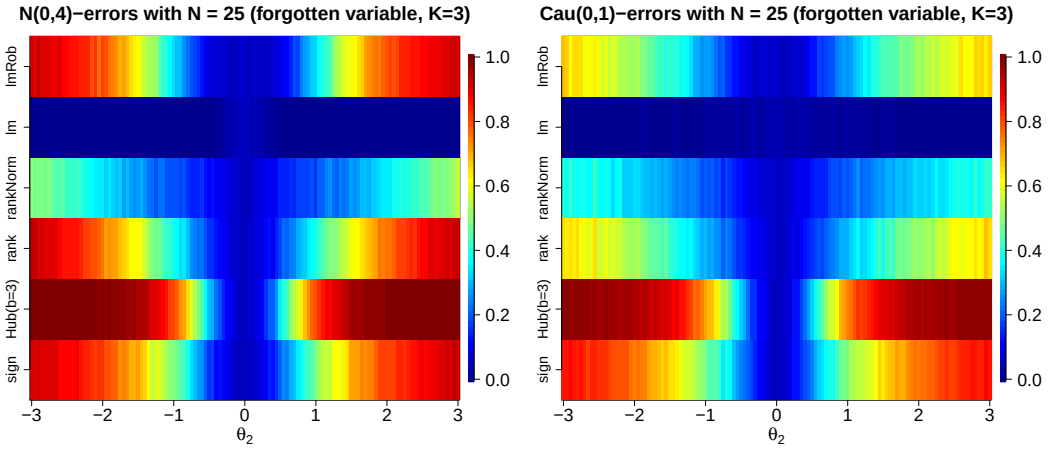

Figure 57: A comparison of the power functions testing $H_0 : \theta_1 = 1$ under varying θ_2 among the F-test (`lm`), the robust Wald-test (`lmRob`), the sign depth test (`sign`), the Huber-score depth test with tuning constant $b = 3$ (`Hub(b=3)`) and the rank depth tests based on the Wilcoxon ranks (`rank`) and the van der Waerden ranks (`rankNorm`) for $K = 3$

the number of intersections between the suggested model and the alternative affects the power for different K. Between a linear and quadratic regression line, we can have up to two intersections such that the difference function can have two sign changes.

Therefore, $K = 4$ should lead to the highest power. However, the choice of K is usually based on the model assumption (cf. Chapter 6.4.2) so that we would except to choose $K = 3$. Therefore, the results compared to the robust Wald-test may not be that impressive on the first glance (expect of the Huber-score depth which may have worse variance estimations) if we choose strictly the hyper-parameter based on the model assumption. Varying the hyper-parameter K for the depth tests is nevertheless a more convenient method for detecting bad model suggestions than extending the model with more unknown parameters or variables for the F-test or robust Wald-test, especially in higher dimensions. An odd and even candidate for K can be considered and the α-level controlled by a Bonferroni correction.

7.2 Errors with weak correlations

Assumption (A1) can be weakened by applying a more general version of Donsker's invariance principle allowing short memory dependencies (Billingsley, 1999, p. 196).

Theorem 7.15. *Let $(X_N)_{N\in\mathbb{N}}$ be a stationary stochastic process, i.e., for each $j \in \mathbb{N}$ the distribution of $(X_k, \ldots, X_{k+j})$ is the same for arbitrary $k \in \mathbb{N}$ (Billingsley, 1999, p. 88), with $\mathbb{E}(X_n) = 0$ and $\mathbb{E}(X_n^2) < \infty$. Moreover, we define*

$$\rho_m := \mathbb{E}(X_1 X_m) \text{ for } m \in \mathbb{N} \text{ and suppose } \sum_{m=2}^{\infty} \rho_m < \infty.$$

Then, the following convergence in distribution holds in $(D[0,1], \mathfrak{D}[0,1])$ with respect to the Skorokhod topology

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} X_n \xrightarrow[N\to\infty]{\mathcal{D}} B_\bullet$$

where $B_\bullet = (B_t)_{t\in[0,1]}$ is the standard Brownian motion on $D[0,1]$ with $B_\bullet \sim W_H$ (cf. Chapter 3.3.1 for more details) and

$$\sigma^2 := \mathbb{E}(X_1^2) + 2\sum_{m=2}^{\infty} \rho_m.$$

In Theorem 7.15, the variance factor σ^2 also contains the autocorrelations ρ_m. Note that Theorem 7.15 is a generalization of Theorem 3.25. For an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$

of random variables with $\mathbb{E}(X_1) = 0$, it follows

$$\sum_{m=2}^{\infty} \mathbb{E}(X_1 X_m) = \sum_{m=2}^{\infty} \mathbb{E}(X_1)\mathbb{E}(X_m) = 0 < \infty,$$

i.e., the autocorrelations vanish and the assumptions of Theorem 7.15 hold. The scaling factor σ^2 also coincides with the variance in this case:

$$\sigma^2 = \mathbb{E}(X_1^2) + 2\sum_{m=2}^{\infty} \mathbb{E}(X_1 X_m) = \operatorname{var}(X_1).$$

Analogous to the proof of Theorem 3.29, p. 50, we can derive the asymptotic distribution of the K-sign depth for a model under more general assumptions as in Assumption 2.1:

Assumption 7.16. *Let $(E_N)_{N\in\mathbb{N}}$ be a stationary stochastic process with values in $\mathbb{R}$ which satisfy the following conditions:*

$$\sum_{m=2}^{\infty} \mathbb{E}(\psi(E_1)\psi(E_m)) < \infty, \tag{A1*}$$

$$\mathbb{P}(\psi(E_n) > 0) = \mathbb{P}(\psi(E_n) < 0) = \frac{1}{2}, n = 1, \ldots, N. \tag{A2}$$

In Assumption 7.16, $\psi(x) = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ can be considered as the sign function but other scores from Chapter 7.1 can be considered with additional assumptions we will not specify here. Note especially that the existence of each expected value $\mathbb{E}(\psi(E_1)\psi(E_m))$ in Assumption (A1*) for each $m \in \mathbb{N}$ has to be ensured for general scores.
Formula (A1*) is satisfied for some *short memory processes* $(E_n)_{n\in\mathbb{N}}$ with vanishing autocorrelations. E.g., MA models (moving-average models) have autocorrelations jumping to zero as the lag is sufficiently high so that almost all summands are zero (Hamilton, 1994, p. 51). In practice, this dependency structure can describe some shortly temporal effects. These temporal effects should disappear sufficiently quickly as Formula (A1*) requires. If ψ denotes the sign function, the summands in Formula (A1*) can be rewritten by

$$\begin{aligned}&\mathbb{E}(\psi(E_1)\psi(E_m)) = \mathbb{E}\left((\mathbb{1}\{E_1 > 0\} - \mathbb{1}\{E_1 < 0\})(\mathbb{1}\{E_m > 0\} - \mathbb{1}\{E_m < 0\})\right)\\ =&\mathbb{P}(E_1 \text{ and } E_m \text{ have the same signs}) - \mathbb{P}(E_1 \text{ and } E_m \text{ have different signs}).\end{aligned}$$

In the next limit theorem, we consider directly the asymptotic representation of the K-sign depth given in Theorem 3.22, p. 39, in order to avoid the discussion

whether the rest terms are asymptotically negligible. This theorem can be applied by considering the asymptotic representation for computing the test statistic, as Lemma 4.1 delivers, and ignoring the rest terms.

Theorem 7.17. *Let $(E_N)_{N\in\mathbb{N}}$ be a stochastic process satisfying Assumption 7.16 and $\mathbf{E} = \mathbf{E}_N = (E_1, \ldots, E_N)^\top$. Consider the approximated K-sign depth:*

$$\tilde{T}_K(\mathbf{E}) := \mathbf{\Psi}_{K,N}(S_\bullet^N) \tag{88}$$

where $\mathbf{\Psi}_{K,N}$ is defined by

$$\mathbf{\Psi}_{K,N}(S_\bullet^N) = \beta_{K,N}\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_{J,N}(S_\bullet^N) - \left(-\frac{1}{2}\right)^{K-2}\left(S_1^N\right)^2 + \frac{1}{2^{K-2}}\right)$$

with $\beta_{K,N} = \frac{N^K K(K-1)}{4\langle N\rangle_K}$ and Notation 3.18 (ii), p. 35, and Notation 3.21, p. 39. Then

$$\frac{\tilde{T}_K(\mathbf{E})}{\sigma^2} \xrightarrow[N\to\infty]{\mathcal{D}} \beta_K\left(-\sum_{J=1}^{K-2}\binom{K-2}{J}\left(-\frac{1}{2}\right)^{K-2-J}\Upsilon_J(B_\bullet) - \left(-\frac{1}{2}\right)^{K-2}B_1^2 + \frac{1}{2^{K-2}}\right)$$

where $(B_t)_{t\in[0,1]}$ denotes the standard Brownian motion, $\beta_K := \frac{K(K-1)}{4}$ and

$$\sigma^2 = 1 + 2\sum_{m=2}^{\infty}\mathbb{E}(\psi(E_1)\psi(E_m)).$$

Note that we have to estimate σ^2 to compute the asymptotic distribution under dependencies since σ^2 is usually unknown. Furthermore, we require assumptions on the dependency structure of the errors to estimate σ^2. However, Theorem 7.17 justifies the application of the K-sign depth under the occurrence of weak dependency structures of the errors and how the necessary correction has to be. In particular, we can test the fit of a parameter and allow the occurrence of particularly given class of dependency structures. E.g., we can consider the null hypothesis of the form:

$$H_0 : \boldsymbol{\theta} \in \mathbf{\Theta}_0 \wedge \rho \in [0, \rho_{max}]$$

where ρ denotes the true value of the series of autocorrelations between the errors and ρ_{max} is the maximal allowed positive correlation. However, the discussion of this application and other ones as simultaneous testing the model and independence is beyond this thesis and content of further research.

8 Summary and outlook for future research

Until the year 2018, asymptotic results for the K-sign depth were limited to the cases $K \in \{2, 3\}$. Additionally, the lack of efficient computational tools hindered strongly research based on simulation studies for performance analysis or real data applications with high sample sizes. This thesis yields solutions for both problems and the basis for further research.

The derivation of the asymptotic distribution for all K delivers an applicable tool for testing the fit of parameters for data with large sample sizes. Generalized versions of the K-sign depth (cf. Chapter 7.1) also have the same asymptotic distribution except for a scaling factor. Therefore, these results have major impact on future research. Furthermore, the theoretical results deliver an approximative algorithm in linear time. For $K \leq 5$, this algorithm can also compute the K-sign depth exactly with some additional effort for the cases $K \in \{4, 5\}$ (cf. Chapter 4).

Analyzing the block structure of the signs of the residual vector yields a deeper understanding of the K-sign depth and leads to a highly efficient and exact algorithm in linear time for all K. Moreover, this algorithm is even faster if the number of blocks is small. While studying several efficient algorithms in Chapter 4, 5 or 6.2, it is noticeable that all of them have in common that they store terms in advance. This strategy may also be useful for other statistics with similar structures to the K-sign depth as U-statistics (Lee, 1990). Besides that, theoretical analyses of the K-sign depth based on considering the block structures were done in the last years as well. From this research, the conjecture in Chapter 5.4 about maximality properties of the K-sign depth was obtained, which may have implications for consistency properties but its detailed analysis is beyond this thesis. Suggestions for applications and further research are given instead such as the choice of the hyper-parameter K in polynomial models (cf. Chapter 6.4), an example of a two-sample relevance test based on the K-sign depth (cf. Chapter 6.5), how to modify the K-sign depth for nonparametric tests for independence and the connections to the runs-test (cf. Chapter 6.6) or generalizations by score functions, ranks or weakening the independence assumption (cf. Chapter 7).

The applications in this thesis are only mentioned for the univariate case, i.e., the models contain one explanatory variable. Horn (2021b) proposes applications based on the K-sign depth for models with multiple explanatory variables in lower and higher dimensions. By computing paths through the explanatory variables in higher dimensional spaces based on distance measures (shortest Hamiltonian path, nearest neighbors, hierarchical clustering), an order of the explanatory variable can be found

which is used as the order of the residuals evaluated for the K-sign depth. Note that these ordering approaches increase the computational costs in dependence of the dimension. Especially the shortest Hamiltonian path algorithm has exponential time complexity in worst case (Held and Karp, 1962). Horn (2021b) shows in simulation studies that distance based measures are a suitable approach for tests based on the depth for several arrangments in the higher dimensional space. Compared to other popular methods, evaluating the K-sign depth of a parameter neither depends on the design matrix (full model) nor estimations of the parameter and covariance matrix are done. Therefore, sparse linear models can be evaluated as well since no matrices needs to be inverted. Further research should combine the results of Horn (2021b) and this thesis. Furthermore, (Horn, 2021b) provides an analysis based on the K-sign depth in a real data application for bridge monitoring where a model is discussed which describes the relationship between the current crack width and the temperature, traffic, time and the crack width 24 hours ago Abbas et al. (2019). A real data application is not considered in this thesis since Horn (2021b) provides the necessary tools for multiple regression analysis based on the K-sign depth.

The generalized depth notions from Chapter 7.1 should be investigated in several model classes due to their potentially higher efficiency. Furthermore, more flexible null hypotheses than hypotheses with only a single point (cf. Chapter 6.1,6.4, 6.5) are important to construct for the applications, e.g., for having ANOVA-type tests. Research for finding an efficient estimator based on the K-sign depth and generalizations (cf. Chapter 6.5.1) or the discussion of the robustness based on the breakdown point or by simulation studies with contaminations can be focused as well. Modern robust statistic considers outliers in higher dimension in all explanatory variables than in just some contaminated explanatory variables or only the explained variables (Raymaekers and Rousseeuw, 2021) which can lead to discussions for robustness based on finding robust orderings of the explanatory variables. Moreover, the performance in other real data examples with outliers for outlier detection should be investigated. Note that also nonparametric models based on kernel smoothing estimations, neural networks or random forests (Hastie et al., 2009, p. 191, p. 389, p. 587), can be tested by the K-sign depth as long as the models are residual-based. This class of models can be considered for future research as well.

This thesis cannot answer the question if the K-sign depth and its generalizations will have a large impact in applied statistics, but the results of this thesis and Horn (2021b) deliver tools to answer this question in future research. At this point, the variety for applications of the K-sign depth is notable. Its large flexibility due to the small number of assumptions on the model leads to many possible research.

Bibliography

Abbas, S., Fried, R., Heinrich, J., Horn, M., Jakubzik, M., Kohlenbach, J., Maurer, R., Michels, A., and Müller, C. (2019). Detection of anomalous sequences in crack data of a bridge monitoring. *Applications in Statistical Computing - From Music Data Analysis to Industrial Quality Improvement*, pages 251–269.

Antille, A. (1979). On the invariance principle for signed-rank statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, (47):315–324.

Aragón, F. J., Goberna, M. A., López, M. A., and Rodríguez, M. M. L. (2018). *Nonlinear Optimization*. Springer.

Aubin, J. (2000). *Applied Functional Analysis*. Wiley, second edition.

Audet, C. and Hare, W. (2017). *Derivative-Free and Blackbox Optimization*. Springer, Cham, first edition.

Bartels, R. (1982). The Rank Version of von Neumann's Ratio Test for Randomness. *Journal of the American Statistical Association*, 77(377):40–46.

Baudin, M. (2009). *Nelder-Mead User's Manual*. `https://www.scilab.org/sites/default/files/neldermead.pdf`.

Beutner, E. and Zähle, H. (2016). Functional delta-method for the bootstrap of quasi-hadamard differentiable functions. *Eletronic Journal of Statistics*, 10(1):1181–1222.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Bogachev, V. I. and Smolyanov, O. G. (2020). *Real and Functional Analysis*. Springer.

Chambers, L. (2001). *The Practical Handbook of Genetic Algorithms*. Chapman & Hall/CRC, second edition.

Clark, B. R. (2008). *Linear Models. The theory and application of Analysis of Variance*. John Wiley & Sons.

Conn, A. R., Scheinberg, K., and Vicente, L. N. (2009). *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia, PA, USA.

Corder, G. W. and Foreman, D. I. (2009). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. Wiley.

Dekkers, A. and Aarts, E. (1991). Global optimization and simulated annealing. *Mathematical Programming*, 50:367–393.

Dette, H. and Kutta, T. (2021). Detecting structural breaks ineigensystemsof functional time series. *Eletronic Journal of Statistic*, 15:944–983.

Dohme, H. (2021). Vergleich der Vorzeichentiefetests mit anderen Tests zur Überprüfung von Unabhängigkeitsannahmen in Zeitreihen. *Master Thesis, TU Dortmund*. `https://www.statistik.tu-dortmund.de/1255.html`.

Dohme, H., Malcherczyk, D., Leckey, K., and Müller, C. H. (2021). K-depth tests for testing simultaneously independence and other model assumptions in time series. *SFB Discussion Paper 24/21*. `https://www.statistik.tu-dortmund.de/2942.html`.

Dürre, A., Fried, R., and Liboschik, T. (2015). Robust estimation of (partial) autocorrelation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7:205–222.

Dueck, G. (1993). New optimization heuristics the great deluge algorithm and the record-to-record travel. *Journal of Computational Physics*, 104(1):86–92.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, 40(8):1–18.

Falkenau, C. P. (2016). Depth based estimators and tests for autoregressive processes with application on crack growth and oil prices. *Dissertation, TU Dortmund*.

Frazier, P. I. (2018). A tutorial on bayesian optimization. *arXiv*. `https://arxiv.org/abs/1807.02811`.

Gibbons, J. and Chakraborti, S. (2003). *Nonparametric statistical inference*. Statistics, textbooks and monographs. Marcel Dekker Incorporated.

Gomez, K. and Gallón, S. (2011). Comparison among high dimensional covariance matrix estimation methods. *Revista Colombiana de Estadistica*, 34(3):567–588.

Gould, H. W. and Srivastava, H. M. (1997). Some combinatorical identities associated with the vandermonde convolution. *Appl. Math. Comput.*, 84(2-3):97–102.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer, second edition.

Held, M. and Karp, R. M. (1962). Dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1).

Hájek, J., Sidak, Z., and Sen, P. (1999). *Theory of Rank Tests.* ISSN. Elsevier Science.

Hodges, J. and Lehmann, E. (1961). Comparison of the normal scores and wilcoxon tests. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.*

Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.*, 15(3):773–780.

Horn, M. (2021a). *GSignTest: Robust Tests for Regression-Parameters via Sign Depth.* R package version 1.0.8.

Horn, M. (2021b). *Sign Depth for Parameter Tests in Multiple Regression.* Dissertation, TU Dortmund.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics.* Wiley, second edition.

Kelly, G. (1992). Robust regression estimators - the choice of tuning constants. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(3):303–314.

Kelly, G. (1996). Adaptive choice of tuning constant for robust regression estimators. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45(1):35–40.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. *IEEE international conference on neural networks*, pages 1942–1948.

Knuth, D. E. (1993). Johann faulhaber and sums of powers. *Mathematics of Computation*, 61(203):277–294.

Kustosz, C. P., Leucht, A., and Müller, C. H. (2016a). Tests based on simplicial depth for AR(1) models with explosion. *Journal of Time Series Analysis*, 37:763–784.

Kustosz, C. P. and Müller, C. H. (2014). Analysis of crack growth with robust, distribution-free estimators and tests for non-stationary autoregressive processes. *Statistical Papers*, 55(1):125–140.

Kustosz, C. P., Müller, C. H., and Wendler, M. (2016b). Simplified simplicial depth for regression and autoregressive growth processes. *Journal of Statistical Planning and Inference*, 173:125–146.

Lagarias, J. C., Reeds, J., Wright, M. H., and E. Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147.

Leckey, K., Malcherczyk, D., and Müller, C. H. (2020). Powerful generalized sign tests based on sign depth. *SFB Discussion Paper 12/20*. `https://www.statistik.tu-dortmund.de/2630.html`.

Lee, A. J. (1990). *U-Statistics. Theory and Practice*. Taylor & Francis.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, 3rd edition.

Liu, R. Y. (1988). On a notion of simplicial depth. *Proceedings of the National Academy of Sciences of the United States of America*, 85:1732–1734.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of t in time series models. *Biometrika*, 65(2):297–303.

Malcherczyk, D. (2018a). Asymptotische Verteilung von vollen Datentiefen. *Master Thesis, TU Dortmund*. `https://www.statistik.tu-dortmund.de/1255.html`.

Malcherczyk, D. (2018b). Vergleich von Zwei-Stichproben-Relevanz-Tests basierend auf t-Tests und Datentiefen. *Bachelor Thesis, TU Dortmund*. `https://www.statistik.tu-dortmund.de/1255.html`.

Malcherczyk, D., Leckey, K., and Müller, C. H. (2021). K-sign depth: From asymptotics to efficient implementation. *Journal of Statistical Planning and Inference*, 215:344–355.

Marcinko, T. (2014). Consequences of assumption violations regarding one-way anova. *The 8th International Days of Statistics and Economics, Prague, September 11-13*.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust statistics*. Wiley.

McKinnon, K. (1998). Convergence of the nelder–mead simples method to a non-stationary point. *SIAM Journal on Optimization*, 9(1):148–158.

Mizera, I. (2002). On depth and deep points: A calculus. *The Annals of Statistics*, 30(6):1681–1736.

Moreau, L. and Aeyels, D. (2000). Optimization of discontinuous functions: A generalized theory of differentation. *SIAM Journal on Optimization*, 11(1):56–69.

Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *Journal of Multivariate Analysis*, 95(1):153–181.

Neusser, K. (2016). *Time Series Econometrics*. Springer.

Paolella, M. S. (2018). *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. Wiley.

Peña, D., Tiao, G. C., and Tsay, R. S. (2001). *A Course in Time Series Analysis*. Wiley.

Qian, W. and Li, M. (2018). Convergence analysis of standard particle swarm optimization algorithm and its improvement. *Soft Computing*, 22:4047–4070.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymaekers, J. and Rousseeuw, P. (2021). *cellWise: Analyzing Data with Cellwise Outliers*. R package version 2.2.5.

Rousseeuw, P. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283.

Rousseeuw, P. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8:193–203.

Rousseeuw, P. J. and Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association*, 94(446):388–402.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.

Scariano, S. M. and Davenport, J. M. (1987). The effects of violations of independence assumptions in the one-way anova. *The American Statistician*, 41(2):123–129.

Sedgewick, R. and Flajolet, P. (2012). *An Introduction to the Analysis of Algorithms*. Addison-Wesley.

Sen, P. (1974). The invariance principle for one-sample rank-order statistics. *Annals of Statistics*, 2(1):49–62.

Sen, P. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. Wiley.

Shumway, R. and Stoffer, D. (2017). *Time Series Analysis and Its Applications*. Springer, fourth edition.

Siarry, P. (2016). *Simulated Annealing*. Metaheuristics. Springer.

Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory of Probability and its Applications*, 1(3):261–290.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation & Testing*. Wiley Series in Probability and Statistics. Wiley & Sons.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794.

Törn, A. and Žilinskas, A. (1989). *Global Optimization*. Springer-Verlag.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 2:523–531.

Van Aelst, S., Rousseeuw, P. J., Hubert, M., and Struyf, A. (2002). The deepest regression method. *Journal of Multivariate Analysis*, 81.1:138–166.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Verbeek, M. (2012). *A Guide to Modern Econometrics*. Wiley, 4th edition.

Wang, J., Zamar, R., Marazzi, A., Yohai, V., Salibian-Barrera, M., Maronna, R., Zivot, E., Rocke, D., Martin, D., Maechler, M., and Konis., K. (2020). *robust: Port of the S+ Robust Library*. R package version 0.5-0.0.

Wellmann, R. and Müller, C. H. (2010). Tests for multiple regression based on simplicial depth. *Journal of Multivariate Analysis*, 101(4):824 – 838.

Wolfram Research, I. (2019). *Mathematica, Version 12.0*. Champaign, IL.

Yan, M. (2016). *Introduction to Topology*. de Gruyter.

Yohai, V., J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, 15(20):642–656.

Eidesstattliche Versicherung (Affidavit)

_____ _____

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

_____

_____

_____

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)