SFB
823

K-sign depth: From asymptotics to efficient implementation

Dennis Malcherczyk, Kevin Leckey,
Christine H. Müller

Discussion Paper

Nr. 14/2020

SFB
823

K-sign depth: From asymptotics to efficient implementation

Dennis Malcherczyk*† Kevin Leckey† Christine H. Müller†

March 30, 2020

Abstract

The K-sign depth (K-depth) of a model parameter θ in a data set is the relative number of K-tuples among its residual vector that have alternating signs. The K-depth test based on K-depth, recently proposed by Leckey et al. (2019), is equivalent to the classical residual-based sign test for $K = 2$, but is much more powerful for $K \geq 3$. This test has two major drawbacks. First, the computation of the K-depth is fairly time consuming, and second, the test requires knowledge about the quantiles of the test statistic which previously had to be obtained by simulation for each sample size individually. We tackle both of these drawbacks by presenting a limit theorem for the distribution of the test statistic and deriving an (asymptotically equivalent) form of the K-depth which can be computed efficiently. For $K = 3$, such a limit theorem was already derived in Kustosz et al. (2016a) by mimicking the proof for U-statistics. We provide here a much shorter proof based on Donsker's theorem and extend it to any $K \geq 3$. As part of the proof, we derive an asymptotically equivalent form of the K-depth which can be computed in linear time. This alternative and the original implementation of the K-depth are compared with respect to their runtimes and absolute difference.

MSC 2010 Subject classifications: Primary 62E20; secondary 62F05, 62F35

Keywords: distribution-free test, K-depth test, K-sign depth, linear runtime, outlier robustness, residual-based test

1 Introduction

Let $R_1(\theta), \dots, R_N(\theta)$ be real-valued residuals for a model with model parameter θ satisfying $\mathbb{P}(R_n(\theta) = 0) = 0$ for the true underlying parameter. For $K \in \mathbb{N} \setminus \{1\}$, the K-sign depth (K-depth) of $R_1(\theta), \dots, R_N(\theta)$ is defined as

$$d_K(R_1(\theta), \dots, R_N(\theta)) = \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(\theta)(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{R_{n_k}(\theta)(-1)^k < 0\} \right), \tag{1}$$

where $\mathbb{1}\{\}$ denotes the indicator function. Hence, the K-depth is the relative number of K-tuples of the residuals with alternating signs. In several models for time series and several generalized

*Corresponding author, Email: `dennis.malcherczyk@tu-dortmund.de`

†Faculty of Statistics, TU Dortmund University, D-44227 Dortmund, Germany

linear models, it can be assumed that the residuals $R_1(\theta), \ldots, R_N(\theta)$ are independent with median $\text{Med}(R_n(\theta)) = 0$ if θ is the true underlying parameter. Under such assumptions, the K-depth can be used for testing hypotheses of the form $H_0 : \theta \in \Theta_0$ where Θ_0 is an arbitrary subset of the parameter space: the so-called K-depth test rejects H_0 if the maximum K-depth of $R_1(\theta), \ldots, R_N(\theta)$ with $\theta \in \Theta_0$ is smaller than a critical value, i.e. there are not enough sign changes in the residuals of the null hypothesis. Leckey et al. (2019) showed that this K-depth test is equivalent to the classical residual-based sign test for $K = 2$ and is much more powerful for $K > 2$. Since it is based on signs, it is outlier robust.

Moreover, a two-sided version can be used to test simultaneously the independence of the residuals and $\text{Med}(R_n(\theta)) = 0$ since a too large K-depth, i.e. too many sign changes in the residuals, indicates negatively correlated residuals, whereas a too small K-depth indicates not only $\text{Med}(R_n(\theta)) \neq 0$ but also positively correlated residuals. In particular, Leckey et al. (2019) noted that a modification of the two-sided version can be considered as a generalization of the runs test given by Wald and Wolfowitz (1940), see e.g. Gibbons and Chakraborti (2003), pp. 78-86.

An important condition for a powerful K-depth test is an appropriate ordering of the residuals $R_1(\theta), \ldots, R_N(\theta)$. For time series or for generalized linear models with one quantitative explanatory variable, the natural ordering is appropriate. If no natural ordering exists, appropriate orderings can be defined as Horn and Müller (2020) demonstrated for multiple regression.

There has been two open problems: One is the determination of the critical values of the test for large sample sizes N and the other problem is the $\Theta(N^K)$ time complexity when calculating the K-depth. Both are solved in this paper.

At first note that K-depth for residuals belongs to the big class of depth notions developed after Tukey (1975) introduced the half space depth. Many of these depth notions concern the deepness of data points in $\mathbb{R}^p$ like Oja depth (Oja, 1983; Chen et al., 2013), simplical depth (Liu, 1988, 1990), Mahalanobis depth (Liu and Singh, 1993; Hu et al., 2011), projection depth (Zuo, 2003, 2006), or zonoid depth (Mosler, 2002; Liu et al., 2019). Other depth notions concern the depth of distributions as considered by Dong and Lee (2014) or of functional data as considered by López-Pintado and Romo (2007, 2009); Claeskens et al. (2014); López-Pintado et al. (2014); Nagy and Ferraty (2019).

However, the K-depth for residuals belongs to the depth notions for model parameters first introduced by Rousseeuw and Hubert (1999). They defined the regression depth of a regression parameter $\theta \in \mathbb{R}^p$ via the residuals given by θ. Instead of residuals, other quality measures like general likelihood functions can be used as proposed by Mizera (2002); Mizera and Müller (2004); Müller (2005); Denecke and Müller (2011, 2012). Unfortunately, the complicated computation of the regression depth and their extensions is a crucial drawback.

The K-depth first appeared as a simplification of the so-called simplicial regression depth. This depth notion is a combination of the regression depth of Rousseeuw and Hubert (1999) and the simplicial depth of Liu (1988, 1990). It defines the depth of a p-dimensional parameter θ as the relative number of $(p+1)$-tuples of residuals that have a positive regression depth. Under certain conditions given in Kustosz et al. (2016b), a positive regression depth of a $(p+1)$-tuple of residuals is equivalent to alternating signs. Hence the simplicial regression depth coincides with the $(p+1)$-depth given by (1) in many applications. Note that the K-depth in its general form also works for other choices than $K = p + 1$ when considering a p dimensional parameter θ.

When using the K-depth for testing, a critical value is needed. For small sample sizes N, the critical value of the K-depth test can be calculated by determining the K-depth for all of the 2^N possible sign constellations. For larger sample sizes, it is way more efficient to use the quantiles of an asymptotic distribution of the K-depth. The derivation of asymptotic distributions of depth notions and depth estimators is not easy, see e.g. Bai and He (1999); Wang (2019). However, almost all simplicial depth notions have the advantage that the depth statistic is a U-statistic which was used in Dümbgen (1992); Arcones and Gine (1993); Arcones et al. (1994). Unfortunately, the U-statistic is degenerated in several cases which makes it harder to derive limit theorems, see Müller (2005); Wellmann et al. (2009); Wellmann and Müller (2010b,a). Some simplicial depth notions based on likelihood depth are not degenerated U-statistics but with the price of providing biased estimators, see Denecke and Müller (2012). The statistic given by the K-depth has the additional disadvantage that it is not a classical U-statistic since it is influenced by the order of the residuals and thus its kernel is not symmetric for $K \geq 3$. In the particular case of an explosive AR(1) regression with two unknown parameters, Kustosz et al. (2016a) provide a limit theorem for the 3-depth by mimicking the classical proof for U-statistics. The resulting asymptotic distribution is given as an integrated two-dimensional Gaussian process. Note that the limit theorem provided by Kustosz et al. (2016a) is not restricted to AR(1) models since the proof only relies on having independent residuals with signs that are uniformly distributed on $\{-1, 1\}$. However, the proof is restricted to the 3-depth, i.e. K-depths with $K \neq 3$ are not covered.

Here, we provide a much shorter proof by using Donsker's theorem (see e.g. Billingsley, 1999, Theorem 14.1). This also yields a slightly different representation of the asymptotic distribution based on a standard Brownian motion rather than the two-dimensional Gaussian process in Kustosz et al. (2016a). Moreover, we extend this proof to any $K \geq 3$. Finally, our proof yields an asymptotically equivalent form of K-depth which can be computed in linear time for all $K \geq 3$ instead of the $\Theta(N^K)$ time needed for any simplicial depth based on subsets with K observations.

The paper is organized as follows. Section 2 provides the derivation of this asymptotic distribution. In Section 3, the asymptotically equivalent form, which can be computed in linear time, is derived and the exact K-depth is compared to the asymptotically equivalent form. The comparison is based on the absolute differences between the two statistics and their runtimes when considering randomly generated signs as an input. In Section 3.2, we present a plot of the densities of the asymptotic distributions for $K = 3, 4, 5$ and describe how the quantiles can be computed.

Notation. $\mathbb{1}_A(t) = \mathbb{1}\{t \in A\}$ denotes the indicator function, i.e. $\mathbb{1}_A(t) = 1$ if $t \in A$ and $\mathbb{1}_A(t) = 0$ if $t \notin A$. For real numbers x, y, let $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. Moreover, the sign of a real number x is denoted by $\psi(x) = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$. We use the standard Bachmann-Landau symbols: For sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, $a_n = O(b_n)$ denotes the existence of a constant $C > 0$ and an integer n_0 such that $|a_n| \leq C|b_n|$ for all $n \geq n_0$. Moreover, $a_n = o(b_n)$ denotes that $a_n/b_n \to 0$ as $n \to \infty$. Finally, $a_n = \Theta(b_n)$ denotes that both $a_n = O(b_n)$ and $b_n = O(a_n)$.

2 The asymptotic distribution of K-sign depth

Since only the asymptotic distribution of the residuals for the true underlying parameter θ is needed for the critical value of the K-depth tests, we now use the notation $E_n = R_n(\theta)$ for $n = 1, \ldots, N$. Hence we have the following assumptions:

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

$$\mathbb{P}(E_n > 0) = \mathbb{P}(E_n < 0) = \frac{1}{2} \quad \text{for all } n \in \mathbb{N}. \tag{A2}$$

The main goal of this article is to derive a limit theorem for the K-depth of $E_1, \ldots, E_N$ as N tends to infinity. As already stated in (1), the K-depth, $K \geq 2$, is defined as

$$d_K(E_1, \ldots, E_N) = \\ \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} \right).$$

Section 2.1 provides the main theorem (Theorem 2.2) with the asymptotic distribution and a second theorem (Theorem 2.3) from which Theorem 2.2 is deduced. Since the proof of Theorem 2.3 needs several lemmas, its proof is given in Section 2.2.

Remark 2.1. Note that a well defined ordering of the residuals is crucial for the definition of the K-depth. The chosen ordering has a large impact on the power of the resulting test and thus should be adjusted to the model. For time series, usually the chronological ordering of the observations in the process is the canonical choice. If a regression model is given only by one univariate explanatory variable then often the ordering with respect to this variable is a good choice. If the explanatory variable is multi-dimensional then data driven orderings as proposed by Horn and Müller (2020) can be used.

2.1 Asymptotic distribution

A limit theorem for the normalized K-depth can be achieved under fairly general assumption on the residuals. Our limit theorem holds for random variables (residuals) $(E_n)_{n \geq 1}$ satisfying the assumptions (A1) and (A2) from the previous section:

Theorem 2.2. *Let $K \geq 3$. If $E_1, \ldots, E_N$ satisfy* (A1) *and* (A2) *then, as $N \to \infty$,*

$$N \left(d_K(E_1, \ldots, E_N) - \left(\frac{1}{2}\right)^{K-1} \right) \xrightarrow{d} \Psi_K(W)$$

where $W = (W_t)_{t \in [0,1]}$ denotes a standard Brownian motion and

$$\Psi_3(W) = \frac{3}{4} \left(1 - \int_0^1 (W_1 - 2W_t)^2 \mathrm{d}t \right),$$

$$\Psi_K(W) \\ = -\frac{K!}{4(K-4)!} \int_{-0.5}^{1} \int_{t \vee 0}^{t+0.5} \left(\frac{1}{2} + t - s \right)^{K-4} \left((W_{s \wedge 1} - W_{t \vee 0})^2 - ((s \wedge 1) - (t \vee 0)) \right) \mathrm{d}s \mathrm{d}t \\ - \frac{K!}{2(K-4)!} \int_{0.5}^{1} \int_{0}^{t-0.5} \left(\frac{1}{2} + s - t \right)^{K-4} W_s \left(W_1 - W_t \right) \mathrm{d}s \mathrm{d}t, \quad K \geq 4.$$

The proof of Theorem 2.2 is done in two steps. In a first step, we rewrite the normalized K-depth as a function of the associated random walk given by

$$\mathcal{W}_t^N := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor tN \rfloor} \psi(E_n), \quad \psi(E_n) = \mathbb{1}\{E_n > 0\} - \mathbb{1}\{E_n < 0\}, \quad t \in [0,1], \tag{2}$$

with the convention $\mathcal{W}_t^N = 0$ for $t < 1/N$. Afterwards we apply Donsker's invariance principle in combination with the continuous mapping theorem to obtain the limit law of the K-depth. Since the first step is much more tedious, we only state the resulting representation in the theorem below and defer its proof to the next section.

Theorem 2.3. *Let $E_1, \ldots, E_N$ be random variables with $\mathbb{P}(E_n \neq 0) = 1$ and let $\mathcal{W}^N$ be as in* (2)*. Moreover, let Ψ_K be as in Theorem 2.2. Then, almost surely,*

$$N\left(d_3(E_1, \ldots, E_N) - \frac{1}{4}\right) = \frac{N^2}{(N-1)(N-2)} \Psi_3\left(\mathcal{W}^N\right).$$

If $E_1, \ldots, E_N$ satisfy (A1) *and* (A2) *then, for all $K \geq 4$, as $N \to \infty$,*

$$N\left(d_K(E_1, \ldots, E_N) - \left(\frac{1}{2}\right)^{K-1}\right) = \frac{N^K (N-K)!}{N!} \Psi_K\left(\mathcal{W}^N\right) + o_P(1),$$

where $o_P(1)$ denotes a random variables which converges to zero in probability.

We end this section by deducing Theorem 2.2 from Theorem 2.3. Some potential generalizations of the limit theorem are given in Remark 2.4 at the end of the section.

Proof of Theorem 2.2. Recall that $W = (W_t)_{t\in[0,1]}$ denotes a standard Brownian motion. By Theorem 2.3 and Slutsky's Theorem, it is sufficient to show that

$$\Psi_K(\mathcal{W}^N) \xrightarrow{d} \Psi_K(W), \quad \text{as } N \to \infty,$$

where $\mathcal{W}^N = (\mathcal{W}_t^N)_{t\in[0,1]}$ is defined as in (2). First note that Donsker's invariance principle for càdlàg processes (Billingsley, 1999, Theorem 14.1) yields that

$$\mathcal{W}^N \xrightarrow{d} W, \quad \text{as } N \to \infty,$$

with respect to the Skorokhod topology. By the continuous mapping theorem and the fact that W is almost surely continuous, it is therefore sufficient to show that the function $\Psi_K : \mathcal{D}[0,1] \to \mathbb{R}$ is continuous in f for all $f \in \mathcal{C}[0,1]$, where $\mathcal{D}[0,1]$ denotes the set of all càdlàg functions on $[0,1]$ and $\mathcal{C}[0,1]$ denotes the set of all continuous functions on $[0,1]$.

To this end, let $(f_n)_{n\geq 1}$ be a convergent sequence in $\mathcal{D}[0,1]$ with a continuous limit f. Note that if f is continuous then the convergence to f in the Skorokhod topology implies (Billingsley, 1999, p. 124)

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0, \quad \text{where } \|f_n - f\|_\infty = \max_{t\in[0,1]} |f_n(t) - f(t)|.$$

Since f is a bounded function, $\|f_n\|_\infty \leq C$ for some constant $C > 0$ and all sufficiently large n. In particular, the dominated convergence theorem implies

$$\lim_{n\to\infty} \int_0^1 (f_n(1) - 2f_n(t))^2 \mathrm{d}t = \int_0^1 \lim_{n\to\infty} (f_n(1) - 2f_n(t))^2 \mathrm{d}t = \int_0^1 (f(1) - 2f(t))^2 \mathrm{d}t$$

and hence $\Psi_3(f_n) \to \Psi_3(f)$ as $n \to \infty$. Essentially the same arguments yield $\Psi_K(f_n) \to \Psi_K(f)$ as $n \to \infty$ for $K \geq 4$. Therefore Ψ_K is continuous in f for $f \in \mathcal{C}[0,1]$ and the assertion follows. □

Remark 2.4. Note that the first identity in Theorem 2.3 for $K = 3$ holds without assuming (A1) and (A2). Hence, if $(E_n)_{n\geq 1}$ is an arbitrary sequence of random variables such that

$$P(E_n \neq 0) = 1 \quad \text{for } n \geq 1,$$
$$\mathcal{W}^N \xrightarrow{d} \widetilde{W} \quad \text{as } N \to \infty$$

with $\mathcal{W}^N$ as in (2) and a process $\widetilde{W}$ with continuous paths, then the convergence in Theorem 2.2 still holds for $K = 3$ after replacing $\Psi_3(W)$ with $\Psi_3(\widetilde{W})$. In particular, we may discard the assumption (A1) and may consider a stationary process $(E_n)_{n\geq 1}$ with some (sufficiently fast decreasing) dependence structure. Then $\mathcal{W}^N$ still converges to a Brownian motion under suitable assumptions, see (Billingsley, 1999, Theorem 19.1 and 19.2) for more details. Hence the limit quantiles are still a fairly good approximation in applications where some local dependence structure between errors $(E_n)_{n\geq 1}$ can occur. Note that a correct approximation might require a proper rescaling by a factor σ^2 though, see (Billingsley, 1999, Theorem 19.1).

2.2 Proof of Theorem 2.3

This section contains the missing proof of Theorem 2.3, that is the transformation of the rescaled K-depth to a function of the associated random walk. We start with an outline of the proof. Recall that the rescaled K-depth is given by

$$N\left(d_K(E_1, \ldots, E_N) - \left(\frac{1}{2}\right)^{K-1}\right). \tag{3}$$

Step 1. Rewrite the rescaled K-depth as a polynomial in $(\psi(E_1), \ldots, \psi(E_N))$. This part of the proof is fairly easy when observing that $\mathbb{1}\{x > 0\} = (\psi(x) + 1)/2$ for all $x \neq 0$. The resulting representation is given in Lemma 2.5.

Step 2. Show that all terms of the polynomial (Step 1) with degree larger than two are asymptotically negligible as $N \to \infty$. This reduces the representation to a sum of pairs $\psi(E_{n_1})\psi(E_{n_2})$ over n_1, n_2. The convergence of the higher degree monomials is stated in Lemma 2.6 whereas the resulting representation in terms of a degree two polynomial is given in Lemma 2.7.

Step 3. After simplifying the polynomial from the previous step (Lemma 2.8), it only remains to transform

$$\sum_{1\leq n_1\neq n_2\leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}$$

into a function of the associated random walk $\mathcal{W}^N$. This mainly involves transforming the leading factor $\left(\frac{1}{2} - \frac{|n_2-n_1|}{N}\right)$ in order to split the double sum into a product of two sums (which can be represented via $\mathcal{W}^N$). More precisely, we rewrite this leading factor using an integral representation stated in Lemma 2.12. This leads to a new representation (Corollary 2.13 for $K \geq 4$ or (12) for $K = 3$) in which the integrand can be expressed in terms of $\mathcal{W}^N$, yielding the assertion stated in Theorem 2.3.

The remainder of this section contains the full proof details of the outline above. Some calculations (for steps 1 and 2) coincide with proofs in Leckey et al. (2019). In that case, proof details are omitted. Rewriting the K-depth in terms of the signs $(\psi(E_1), \ldots, \psi(E_n))$ is based on the following identity.

Lemma 2.5. *For every $K \geq 2$ and $x_1, \ldots, x_K \in \mathbb{R} \setminus \{0\}$,*

$$\prod_{k=1}^{K} \mathbb{1}\{x_k(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{x_k(-1)^k < 0\} - \left(\frac{1}{2}\right)^{K-1}$$
$$= \left(\frac{1}{2}\right)^{K-1} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} (-1)^{i(1)+\ldots+i(2L)} \prod_{j=1}^{2L} \psi(x_{i(j)})$$

where $\psi(x)$ denotes the sign of x, that is $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$.

Proof. The identity is based on $\mathbb{1}\{x > 0\} = (\psi(x) + 1)/2$ and $\psi(-x) = -\psi(x)$ for all $x \neq 0$. The full proof for Lemma 2.5 is given in (Leckey et al., 2019, Lemma 2). □

Note that if $\mathbb{P}(E_n \neq 0) = 1$ for all $n \in \mathbb{N}$ then Lemma 2.5 yields almost surely

$$\begin{aligned} &N\left(d_K(E_1, ..., E_N) - \left(\frac{1}{2}\right)^{K-1}\right) \\ =&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} (-1)^{i(1)+\ldots+i(2L)} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \prod_{j=1}^{2L} \psi(E_{n_{i(j)}}). \end{aligned} \tag{4}$$

The next step is to show that the summands for $L \geq 2$ are asymptotically negligible (note that these summands only occur for $K \geq 4$).

Lemma 2.6. *Let $K \geq 4$. If $(E_n)_{n \geq 1}$ satisfy* (A1) *and* (A2) *then, as $N \to \infty$,*

$$\frac{N}{2^{K-1}\binom{N}{K}} \sum_{L=2}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \leq i(1) < \ldots < i(2L) \leq K} \sum_{1 \leq n_1 < \ldots < n_K \leq N} (-1)^{i(1)+\ldots+i(2L)} \prod_{j=1}^{2L} \psi(E_{n_{i(j)}}) \xrightarrow{P} 0.$$

Proof. The proof is based on computing the variance and using Chebyshev's inequality to obtain the convergence in probability. A full proof is given in (Leckey et al., 2019, Theorem 1). Also see Lemma 2.9 below for a similar proof strategy. □

Lemma 2.6 in combination with (4) yields the following representation:

$$\begin{aligned} &N\left(d_K(E_1, ..., E_N) - \left(\frac{1}{2}\right)^{K-1}\right) \\ =&\frac{N}{2^{K-1}\binom{N}{K}} \sum_{1 \leq i(1) < i(2) \leq K} \sum_{1 \leq n_1 < \ldots < n_K \leq N} (-1)^{i(1)+i(2)} \psi(E_{n_{i(1)}}) \psi(E_{n_{i(2)}}) + o_P(1). \end{aligned}$$

Note that for $K = 3$, this equation holds almost surely even without the $o_P(1)$ term. We continue the proof by replacing the sum over $n_1 < \ldots < n_K$ with a double sum involving only $n_{i(1)}$ and $n_{i(2)}$. This can be done by counting the number of choices for $\{n_j; j \notin \{i(1), i(2)\}\}$ when $n_{i(1)}$ and $n_{i(2)}$ are fixed. We obtain the following identity.

1	2	$\dots$	$n_{i(1)}-1$	$n_{i(1)}$	$n_{i(1)}+1$	$\dots$	$n_{i(2)}-1$	$n_{i(2)}$	$n_{i(2)}+1$	$\dots$	N

$\binom{n_{i(1)}-1}{i(1)-1}$ choices $\quad$ $\binom{n_{i(2)}-n_{i(1)}-1}{i(2)-i(1)-1}$ choices $\quad$ $\binom{N-n_{i(2)}}{K-i(2)}$ choices

Figure 1: Number of possible choices for $\{n_j; j < i(1)\}$ (left), $\{n_j; i(1) < j < i(2)\}$ (middle), and $\{n_j; j > i(2)\}$ (right) for fixed $i(1), i(2), n_{i(1)}, n_{i(2)}$.

Lemma 2.7. *Let $K \geq 3$. Then*

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

Proof. In the first part of the proof, we fix $1 \leq i(1) < i(2) \leq K$. The aim is to reduce the inner sum over $n_1, \ldots, n_K$ to a sum with only two indices of the following form:

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

with suitable constants $\kappa(n_{i(1)}, n_{i(2)}) := \kappa(N, K, i(1), i(2), n_{i(1)}, n_{i(2)})$. These constants can be obtained by counting the possible choices for $\{n_j; j \notin \{i(1), i(2)\}\}$ when $n_{i(1)}$ and $n_{i(2)}$ are fixed. After counting these possibilities as depicted in Figure 1, we obtain using $J = i(2) - i(1)$,

$$\begin{aligned}\kappa(n_{i(1)}, n_{i(2)}) &= \binom{n_{i(1)}-1}{i(1)-1}\binom{n_{i(2)}-n_{i(1)}-1}{i(2)-i(1)-1}\binom{N-n_{i(2)}}{K-i(2)} \\ &= \binom{n_{i(1)}-1}{i(1)-1}\binom{n_{i(2)}-n_{i(1)}-1}{J-1}\binom{N-n_{i(2)}}{K-i(1)-J}. \end{aligned} \tag{5}$$

Recall that $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$ for $x, y \in \mathbb{R}$. Then (5) implies

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

where the last equation holds by symmetry between n_1, n_2 and by $(-1)^{a+b} = (-1)^{a-b}$ for $a, b \in \mathbb{N}$. It remains to sum this result over $i(1) < i(2)$ to obtain the assertion. Note that rather than summing over $i(2)$ directly, we may sum over $J = 1, \dots, K-1$ and $i(1) = 1, \dots, K-J$ and set $i(2) = i(1) + J$. We obtain

$$\begin{aligned}
&\sum_{1 \le i(1) < i(2) \le K} \sum_{1 \le n_1 < \dots < n_K \le N} (-1)^{i(1)+i(2)} \psi(E_{n_1}) \psi(E_{n_2}) \\
&= \sum_{J=1}^{K-1} \sum_{i(1)=1}^{K-J} \frac{(-1)^J}{2} \sum_{1 \le n_1 \ne n_2 \le N} \binom{(n_1 \wedge n_2) - 1}{i(1) - 1} \binom{|n_1 - n_2| - 1}{J - 1} \times \\
&\qquad \binom{N - (n_1 \vee n_2)}{K - i(1) - J} \psi(E_{n_1}) \psi(E_{n_2}).
\end{aligned}$$

To simplify this expression, note that Vandermonde's convolution (see e.g. Gould and Srivastava, 1997) yields

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}, \quad \text{for all integers } m, n, r \ge 0. \tag{6}$$

After changing the order of summation by swapping the sum over $n_1 \ne n_2$ with the sum over $i(1)$, we obtain for any fixed J, n_1, n_2 as an inner sum

$$\begin{aligned}
&\sum_{i(1)=1}^{K-J} \binom{n_1 \wedge n_2 - 1}{i(1) - 1} \binom{|n_1 - n_2| - 1}{J - 1} \binom{N - (n_1 \vee n_2)}{K - i(1) - J} \psi(E_{n_1}) \psi(E_{n_2}) \\
=& \psi(E_{n_1}) \psi(E_{n_2}) \binom{|n_1 - n_2| - 1}{J - 1} \sum_{i(1)=1}^{K-J} \binom{(n_1 \wedge n_2) - 1}{i(1) - 1} \binom{N - (n_1 \vee n_2)}{K - i(1) - J} \\
=& \psi(E_{n_1}) \psi(E_{n_2}) \binom{|n_1 - n_2| - 1}{J - 1} \binom{N - (n_1 \vee n_2) + (n_1 \wedge n_2) - 1}{K - J - 1},
\end{aligned}$$

in which the last equality uses (6) and the substitution $k = i(1) - 1$. Since $(n_1 \vee n_2) - (n_1 \wedge n_2) = |n_1 - n_2|$, the summation over J and n_1, n_2 yields that

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

The assertion follows after substituting $\tilde{J} = J - 1$ and swapping the order of summation. □

The next step is to simplify the inner sum over J in Lemma 2.7. To this end, let the K-th falling factorial of N be denoted by $\langle N \rangle_K$, that is

$$\langle N \rangle_K = \frac{N!}{(N-K)!}.$$

The previous transformations and the asymptotic $\binom{N}{k} = \frac{N^k}{k!} + O(N^{k-1})$ for bounded k as $N \to \infty$ yield the following representation of the normalized K-depth:

Lemma 2.8. *Let $K \geq 3$. If* (A1) *and* (A2) *hold then, as $N \to \infty$,*

$$
\begin{aligned}
&N\left(d_K(E_1, ..., E_N) - \left(\frac{1}{2}\right)^{K-1}\right) \\
&= -\frac{N^K}{\langle N\rangle_K}\frac{K(K-1)}{4}\sum_{1\leq n_1\neq n_2\leq N}\left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right)^{K-2}\frac{\psi(E_{n_1})\psi(E_{n_2})}{N} + o_P(1). \qquad (7)
\end{aligned}
$$

For $K = 3$, equation (7) *holds for any sequence $(E_n)_{n\geq 1}$ with $\mathbb{P}(E_n \neq 0) = 1$ and is an almost sure equality without the $o_P(1)$ term.*

Proof. First recall that a combination of (4), Lemma 2.6, and Lemma 2.7 yields

$$
\begin{aligned}
&N\left(d_K(E_1, ..., E_N) - \left(\frac{1}{2}\right)^{K-1}\right) \\
&= -\frac{N}{2^K\binom{N}{K}}\sum_{1\leq n_1\neq n_2\leq N}\psi(E_{n_1})\psi(E_{n_2})\times \\
&\qquad \sum_{J=0}^{K-2}(-1)^J\binom{|n_1 - n_2| - 1}{J}\binom{N - |n_1 - n_2| - 1}{K - 2 - J} + o_P(1).
\end{aligned}
$$

Note that $\binom{N}{k} = N^k/k! + O(N^{k-1})$ for bounded k. This can be seen, e.g., after rewriting $\binom{N}{k} = \langle N\rangle_k/k!$ and using $N^k \geq \langle N\rangle_k \geq (N - k)^k$, and, by Bernoulli's inequality (Carothers, 2000, p. 9),

$$(N - k)^k = N^k(1 - k/N)^k \geq N^k(1 - k^2/N) = N^k - k^2N^{k-1}.$$

In combination with $|n_1 - n_2| \leq N$, we thus obtain

$$
\begin{aligned}
\binom{|n_1 - n_2| - 1}{J} &= \frac{(|n_1 - n_2| - 1)^J}{J!} + O(N^{J-1}), \\
\binom{N - |n_1 - n_2| - 1}{K - 2 - J} &= \frac{(N - |n_1 - n_2| - 1)^{K-2-J}}{(K - 2 - J)!} + O(N^{K-3-J}),
\end{aligned} \qquad (8)
$$

for all $J = 0, \ldots, K - 2$ and $n_1, n_2 = 1, \ldots, N$. Note that the constant in the O-term can be chosen to be, e.g., K^2 and hence does not depend on J, n_1, n_2. In particular, (8) implies

$$
\begin{aligned}
&\binom{|n_1 - n_2| - 1}{J}\binom{N - |n_1 - n_2| - 1}{K - 2 - J} \\
=&\frac{(|n_1 - n_2| - 1)^J}{J!}\frac{(N - |n_1 - n_2| - 1)^{K-2-J}}{(K - 2 - J)!} + O(N^{K-3}) \\
=&\frac{1}{(K - 2)!}\binom{K - 2}{J}(|n_1 - n_2| - 1)^J(N - |n_1 - n_2| - 1)^{K-2-J} + O(N^{K-3}).
\end{aligned}
$$

Plugging this approximation back into the sum over J yields

$$\begin{aligned}
&\sum_{J=0}^{K-2}(-1)^J\binom{|n_1-n_2|-1}{J}\binom{N-|n_1-n_2|-1}{K-2-J}\\
=&\frac{1}{(K-2)!}\sum_{J=0}^{K-2}\binom{K-2}{J}\left(-(|n_1-n_2|-1)\right)^J\left(N-|n_1-n_2|-1\right)^{K-2-J}+O(N^{K-3})\\
=&\frac{(N-2|n_1-n_2|)^{K-2}}{(K-2)!}+O(N^{K-3}),
\end{aligned}$$

in which the last equality holds by the binomial theorem. Thus

$$\begin{aligned}
&N\left(d_K(E_1,...,E_N)-\left(\frac{1}{2}\right)^{K-1}\right)\\
&=-\frac{N}{2^K\binom{N}{K}}\sum_{1\leq n_1\neq n_2\leq N}\psi(E_{n_1})\psi(E_{n_2})\frac{(N-2|n_1-n_2|)^{K-2}}{(K-2)!}+\mathcal{R}_N+o_P(1) \qquad (9)
\end{aligned}$$

where

$$\mathcal{R}_N=\sum_{1\leq n_1\neq n_2\leq N}\psi(E_{n_1})\psi(E_{n_2})\varepsilon_N(n_1,n_2)\quad\text{with}\quad\varepsilon_N(n_1,n_2)=O(N^{-2}).$$

As proven formally in Lemma 2.9 below, one can bound the variance of $\mathcal{R}_N$ to show that $\mathcal{R}_N\overset{P}{\to}0$ as $N\to\infty$. Equation (7) then follows from (9) after rewriting

$$\binom{N}{K}=\frac{\langle N\rangle_K}{K!}\quad\text{and}\quad(N-2|n_1-n_2|)^{K-2}=2^{K-2}N^{K-2}\left(\frac{1}{2}-\frac{|n_1-n_2|}{N}\right)^{K-2}.$$

Finally note that for $K=3$, all equalities hold almost surely (without the $o_P(1)$ error) since only the summand for $L=1$ in (4) occurs (hence we do not need to apply Lemma 2.6) and since (8) holds without the O-terms if $K=3$ and $J\in\{0,1\}$. □

Lemma 2.9. *Let $\{\varepsilon_N(n_1,n_2);N\in\mathbb{N},n_1,n_2\in\{0,\dots,N\}\}$ be a family of real numbers and let*

$$\mathcal{R}_N=\sum_{1\leq n_1\neq n_2\leq N}\psi(E_{n_1})\psi(E_{n_2})\varepsilon_N(n_1,n_2)$$

If $\varepsilon_N(n_1,n_2)=O(N^{-2})$ and $(E_n)_{n\geq1}$ satisfies (A1) *and* (A2)*, then $\mathcal{R}_N\overset{P}{\to}0$ as $N\to\infty$.*

Proof. First note that (A1) and (A2) imply $\mathbb{E}[\psi(E_{n_1})\psi(E_{n_2})]=0$ for $n_1\neq n_2$. Hence

$$\mathbb{E}[\mathcal{R}_N]=0.$$

Thus, by Chebyshev's inequality it is sufficient to show that

$$\lim_{n\to\infty}\operatorname{Var}(\mathcal{R}_N)=0. \qquad (10)$$

To this end, note that (A1), (A2) and $(\psi(x))^2 = 1$ imply for $n_1 \neq n_2$ and $\tilde{n}_1 \neq \tilde{n}_2$ that

$$\mathbb{E}[\psi(E_{n_1})\psi(E_{n_2})\psi(E_{\tilde{n}_1})\psi(E_{\tilde{n}_2})] = \begin{cases} 1, & \text{if } \{n_1, n_2\} = \{\tilde{n}_1, \tilde{n}_2\}, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, $\varepsilon_N(n_1, n_2) = O(N^{-2})$ implies $\varepsilon_N(n_1, n_2)^2 = O(N^{-4})$ and thus

$$\mathbb{E}[\mathcal{R}_N^2] = 2 \sum_{1 \leq n_1 \neq n_2 \leq N} (\varepsilon_N(n_1, n_2))^2 \to 0, \quad \text{as } N \to \infty.$$

Therefore (10) holds and Chebyshev's inequality yields the assertion. □

For $K = 3$, the identity in Theorem 2.3 can be deduced from (7) in the same way it is done in Kustosz et al. (2016a). We will end the section by first summarizing how the case $K = 3$ works. Afterwards, we prepare some additional transformations in Lemma 2.12 and Corollary 2.13 to finally prove Theorem 2.3 for $K \geq 4$.

To this end, recall that the associated random walk $\mathcal{W}^N$ is defined as

$$\mathcal{W}_t^N := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor tN \rfloor} \psi(E_n), \quad t \in [0, 1], \tag{11}$$

and that the statement of Theorem 2.3 for $K = 3$ is as follows:

Theorem 2.10. *Let $\mathcal{W}^N$ be as above and let*

$$\Psi_3(\mathcal{W}^N) = \frac{3}{4}\left(1 - \int_0^1 (\mathcal{W}_1^N - 2\mathcal{W}_t^N)^2 \mathrm{d}t\right).$$

Furthermore, let $E_1, \ldots, E_N$ be random variables with $\mathbb{P}(E_n \neq 0) = 1$. Then

$$N\left(d_3(E_1, \ldots, E_N) - \frac{1}{4}\right) = \frac{N^2}{(N-1)(N-2)} \Psi_3(\mathcal{W}^N) \quad \textit{almost surely.}$$

Proof. By Lemma 2.8 it is sufficient to show that, almost surely,

$$-\frac{3}{2} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right) \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} = \Psi_3(\mathcal{W}^N).$$

First note that since $(\psi(E_n))^2 = 1$ almost surely for all n,

$$\sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right) \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} = \sum_{n_1, n_2 = 1}^{N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right) \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} - \frac{1}{2}.$$

In order to separate the sum over n_1 and n_2, note that for all $n_1, n_2 \leq N$

$$\frac{|n_1 - n_2|}{N} = 1 - \int_{-0.5}^{1.5} \mathbb{1}_{[-0.5,0.5)}(t - n_1/N) \mathbb{1}_{[-0.5,0.5)}(t - n_2/N) \mathrm{d}t, \tag{12}$$

see also Kustosz et al. (2016a). Plugging this identity into the sum and exchanging sum and integral yields

$$\sum_{n_1,n_2=1}^{N}\left(\frac{1}{2}-\frac{|n_2-n_1|}{N}\right)\frac{\psi(E_{n_1})\psi(E_{n_2})}{N}$$
$$=-\frac{1}{2}\left(\mathcal{W}_1^N\right)^2+\int_{-0.5}^{1.5}\sum_{n_1,n_2=1}^{N}\mathbb{1}_{[-0.5,0.5)}(t-n_1/N)\mathbb{1}_{[-0.5,0.5)}(t-n_2/N)\ \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}\ \mathrm{d}t.$$

Finally note that $\mathbb{1}_{[-0.5,0.5)}(t-n_1/N)\mathbb{1}_{[-0.5,0.5)}(t-n_2/N)=1$ if and only if

$$\lfloor((t-0.5)\vee 0)N\rfloor+1\leq n_1\leq\lfloor((t+0.5)\wedge 1)N\rfloor,$$
$$\lfloor((t-0.5)\vee 0)N\rfloor+1\leq n_2\leq\lfloor((t+0.5)\wedge 1)N\rfloor.$$

Hence we obtain

$$\int_{-0.5}^{1.5}\sum_{n_1,n_2=1}^{N}\mathbb{1}_{[-0.5,0.5)}(t-n_1/N)\mathbb{1}_{[-0.5,0.5)}(t-n_2/N)\ \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}\ \mathrm{d}t$$
$$=\int_{-0.5}^{1.5}\sum_{n_1,n_2=\lfloor((t-0.5)\vee 0)N\rfloor+1}^{\lfloor((t+0.5)\wedge 1)N\rfloor}\frac{\psi(E_{n_1})\psi(E_{n_2})}{N}\mathrm{d}t$$
$$=\int_{-0.5}^{1.5}(\mathcal{W}_{(t+0.5)\wedge 1}^N-\mathcal{W}_{(t-0.5)\vee 0}^N)^2\mathrm{d}t$$
$$=\int_0^1\left(\mathcal{W}_t^N\right)^2+\left(\mathcal{W}_1^N-\mathcal{W}_t^N\right)^2\mathrm{d}t$$

in which the last equality holds by splitting the integral into the ranges $[-0.5,0.5]$ and $[0.5,1.5]$, substituting $\tilde{t}=t+0.5$ in the first part and $\tilde{t}=t-0.5$ in the second part. In summary, we have shown that

$$-\frac{3}{2}\sum_{1\leq n_1\neq n_2\leq N}\left(\frac{1}{2}-\frac{|n_2-n_1|}{N}\right)\frac{\psi(E_{n_1})\psi(E_{n_2})}{N}$$
$$=\frac{3}{4}+\frac{3}{4}\left(\mathcal{W}_1^N\right)^2-\frac{3}{2}\int_0^1\left(\mathcal{W}_t^N\right)^2+\left(\mathcal{W}_1^N-\mathcal{W}_t^N\right)^2\mathrm{d}t$$
$$=\frac{3}{4}\left(1-\int_0^1 2\left(\mathcal{W}_t^N\right)^2+2\left(\mathcal{W}_1^N-\mathcal{W}_t^N\right)^2-\left(\mathcal{W}_1^N\right)^2\mathrm{d}t\right).$$

Therefore it only remains to show that

$$2\left(\mathcal{W}_t^N\right)^2+2\left(\mathcal{W}_1^N-\mathcal{W}_t^N\right)^2-\left(\mathcal{W}_1^N\right)^2=\left(\mathcal{W}_1^N-2\mathcal{W}_t^N\right)^2,\quad t\in[0,1].$$

This equality can be obtained by simplifying the left-hand-side using the binomial theorem. □

It remains to prove Theorem 2.3 for $K\geq 4$. Recall that the statement for $K\geq 4$ is as follows:

Theorem 2.11. *Let $\mathcal{W}^N$ be as in* (11) *with $(E_n)_{n\geq 1}$ satisfying* (A1) *and* (A2)*. Let*

$$\begin{aligned}
&\Psi_K(\mathcal{W}^N)\\
&= -\frac{K!}{4(K-4)!}\int_{-0.5}^{1}\int_{t\vee 0}^{t+0.5}\left(\frac{1}{2}+t-s\right)^{K-4}\left((\mathcal{W}^N_{s\wedge 1}-\mathcal{W}^N_{t\vee 0})^2-((s\wedge 1)-(t\vee 0))\right)\mathrm{d}s\mathrm{d}t\\
&\quad -\frac{K!}{2(K-4)!}\int_{0.5}^{1}\int_{0}^{t-0.5}\left(\frac{1}{2}+s-t\right)^{K-4}\mathcal{W}^N_s\left(\mathcal{W}^N_1-\mathcal{W}^N_t\right)\mathrm{d}s\mathrm{d}t, \quad K\geq 4.
\end{aligned}$$

Then, for all $K\geq 4$, as $N\to\infty$,

$$N\left(d_K(E_1,\ldots,E_N)-\left(\frac{1}{2}\right)^{K-1}\right)=\frac{N^K(N-K)!}{N!}\Psi_K\left(\mathcal{W}^N\right)+o_P(1).$$

The only major difference in the proof compared to $K=3$ is that a transformation of $(0.5-|n_2-n_1|/N)^{K-2}$ based on (12) becomes more tedious if $K\geq 4$. We will therefore replace (12) by another identity based on the following observation.

Lemma 2.12. *Let $c<d$ and let $J\geq 2$. Then, for all $a,b\in[c,d]$,*

$$(b-a)^J=J(J-1)\int_c^d\int_c^t(t-s)^{J-2}\mathbb{1}\{s\geq a\}\mathbb{1}\{t<b\}+(s-t)^{J-2}\mathbb{1}\{s\geq b\}\mathbb{1}\{t<a\}\mathrm{d}s\mathrm{d}t.$$

Proof. We first consider the case $a\leq b$. By computing the integral on the right hand side it is easy to check that

$$(b-a)^J=J(J-1)\int_a^b\int_a^t(t-s)^{J-2}\mathrm{d}s\mathrm{d}t. \tag{13}$$

Note that $\mathbb{1}\{s\geq b\}\mathbb{1}\{t<a\}=0$ for all $s\leq t$ since $a\leq b$. Hence

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

and the assertion for $a\leq b$ follows from (13). Now consider the case $a>b$. Then $\mathbb{1}\{s\geq a\}\mathbb{1}\{t<b\}=0$ for all $s\leq t$ and thus

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

The assertion once again follows from (13) after exchanging the roles of a and b and noting that $(b-a)^J=(-1)^J(a-b)^J$. □

Corollary 2.13. *Let $N \in \mathbb{N}$, $n_1, n_2 \in \{1, \dots, N\}$ and $J \geq 2$. Then*

$$\begin{aligned}
&\left(\frac{1}{2} - \frac{n_2 - n_1}{N}\right)^J \\
&= J(J-1) \int_{-0.5}^{1} \int_0^{t+0.5} (0.5 + t - s)^{J-2} \, \mathbb{1}\{s \geq n_2/N\} \mathbb{1}\{t < n_1/N\} \mathrm{d}s\mathrm{d}t \\
&\quad + J(J-1) \int_{0.5}^{1} \int_0^{t-0.5} (0.5 + s - t)^{J-2} \, \mathbb{1}\{s \geq n_1/N\} \mathbb{1}\{t < n_2/N\} \mathrm{d}s\mathrm{d}t.
\end{aligned}$$

Proof. Let $b = 0.5 + n_1/N$ and $a = n_2/N$. Since $a, b \in [0, 1.5]$, Lemma 2.12 yields

$$\begin{aligned}
\left(\frac{1}{2} - \frac{n_2 - n_1}{N}\right)^J &= J(J-1) \int_0^{1.5} \int_0^t (t-s)^{J-2} \, \mathbb{1}\{s \geq n_2/N\} \mathbb{1}\{t - 0.5 < n_1/N\} \mathrm{d}s\mathrm{d}t \\
&\quad + J(J-1) \int_0^{1.5} \int_0^t (s-t)^{J-2} \, \mathbb{1}\{s - 0.5 \geq n_1/N\} \mathbb{1}\{t < n_2/N\} \mathrm{d}s\mathrm{d}t.
\end{aligned}$$

The assertion follows after substituting $\tilde{t} = t - 0.5$ in the first integral and $\tilde{s} = s - 0.5$ in the second integral and adjusting the integration limits to areas where the integrand is nonzero. □

Proof of Theorem 2.11. By Lemma 2.8 it is sufficient to study the sum in (7), that is it only remains to show that

$$-\frac{K(K-1)}{4} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} = \Psi_K(\mathcal{W}^N) + o_P(1).$$

Note that by (a) symmetry and (b) Corollary 2.13

$$\begin{aligned}
&\sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N}\right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} \\
&\overset{(a)}{=} 2 \sum_{1 \leq n_1 < n_2 \leq N} \left(\frac{1}{2} - \frac{n_2 - n_1}{N}\right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} \\
&\overset{(b)}{=} 2(K-2)(K-3) \int_{-0.5}^{1} \int_0^{t+0.5} (0.5 + t - s)^{K-4} \, \mathcal{S}_1^N(s,t) \mathrm{d}s\mathrm{d}t \qquad (14) \\
&\quad + 2(K-2)(K-3) \int_{0.5}^{1} \int_0^{t-0.5} (0.5 + s - t)^{K-4} \, \mathcal{S}_2^N(s,t) \mathrm{d}s\mathrm{d}t \qquad (15)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{S}_1^N(s,t) &= \sum_{1 \leq n_1 < n_2 \leq N} \mathbb{1}\{s \geq n_2/N\} \mathbb{1}\{t < n_1/N\} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}, \\
&\quad \text{with } -0.5 \leq t \leq 1, \, 0 \leq s \leq t + 0.5, \\
\mathcal{S}_2^N(s,t) &= \sum_{1 \leq n_1 < n_2 \leq N} \mathbb{1}\{s \geq n_1/N\} \mathbb{1}\{t < n_2/N\} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}, \\
&\quad \text{with } 0.5 \leq t \leq 1, \, 0 \leq s \leq t - 0.5.
\end{aligned}$$

Next note that $\mathcal{S}_1^N(s,t) = 0$ for $s \le t$ and that for $0 \le t < s \le 1$, again using symmetry in (a) and using $\psi(E_n)^2 = 1$ a.s. in (b),

$$\begin{aligned}\mathcal{S}_1^N(s,t) &= \sum_{n_1=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \sum_{n_2=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \mathbb{1}\{n_1 < n_2\} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} \\ &\overset{(a)}{=} \frac{1}{2} \sum_{n_1=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \sum_{n_2=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \mathbb{1}\{n_1 \neq n_2\} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} \\ &\overset{(b)}{=} \frac{1}{2} \sum_{n_1=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \sum_{n_2=\lfloor tN\rfloor+1}^{\lfloor sN\rfloor} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} - \frac{\lfloor sN\rfloor - \lfloor tN\rfloor}{2N}.\end{aligned}$$

Finally note that $\mathcal{S}_1^N(s,t) = \mathcal{S}_1^N(1,t)$ for $s > 1$ and $\mathcal{S}_1^N(s,t) = \mathcal{S}_1^N(s,0)$ for $t < 0$. Hence, we obtain

$$\mathcal{S}_1^N(s,t) = \mathbb{1}\{s > t\}\frac{1}{2}\left(\left(\mathcal{W}_{s\wedge 1}^N - \mathcal{W}_{t\vee 0}^N\right)^2 - ((s\wedge 1) - (t \vee 0))\right) + o(1). \tag{16}$$

For $\mathcal{S}_2^N(s,t)$ note that $s < t$, $n_1/N \le s$ and $n_2/N > t$ already imply $n_1 < n_2$. Hence,

$$\mathcal{S}_2^N(s,t) = \sum_{n_1=1}^{\lfloor sN\rfloor} \sum_{n_2=\lfloor tN\rfloor+1}^{N} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N} = \mathcal{W}_s^N\left(\mathcal{W}_1^N - \mathcal{W}_t^N\right). \tag{17}$$

Plugging (16) and (17) back into (14) and (15) yields the assertion. □

3 Implementation and quantiles

This section focuses on practical aspects of the K-depth. Firstly, we present an algorithm that computes the asymptotic equivalent version of the K-depth in linear time and compare the runtimes of this algorithm and a "naive" implementation. In the second part, we describe the limit distribution and explain how its quantiles can be computed for tests based on the K-depth. We use the software `R` in the following computational applications, see R Core Team (2019).

3.1 Implementation of the K-depth in linear time

A major drawback of the K-depth is its high computational effort when implementing the sum over $1 \le n_1 < \ldots, n_K \le N$ directly using K nested loops, hence leading to a runtime of order $\Theta(N^K)$. This implementation will be referred to as the *naive implementation*.

The representation in Theorem 2.3 can be used to speed up the computation significantly: Instead of computing the depth $d_K(E_1,\ldots,E_N)$ or its rescaled version in (3) directly, it is sufficient to compute $\Psi_K(\mathcal{W}^N)$ instead and obtain the K-depth (asymptotically) using the equality stated in Theorem 2.3. The value for $\Psi_K(\mathcal{W}^N)$, however, can be computed in linear time as discussed below. We start with the case $K = 3$ to keep the implementation as simple as possible.

First note that the random walk $\mathcal{W}^N$ is piecewise constant by definition. Hence, the integral in Theorem 2.10 becomes

$$\int_0^1 (\mathcal{W}_1^N - 2\mathcal{W}_t^N)^2\,\mathrm{d}t = \frac{1}{N}\sum_{k=0}^{N-1}(\mathcal{W}_1^N - 2\mathcal{W}_{k/N}^N)^2.$$

The remaining sum can be computed in linear time as follows. At first, compute the vector $(\mathcal{W}^N_{k/N})_{k=0,\dots,N}$ which is essentially the cumulative sum of $\psi(E_1),\dots,\psi(E_k)$ and thus can be derived in linear time. Then compute $(\mathcal{W}^N_1 - 2\mathcal{W}^N_{k/N})_{k=0,\dots,N}$ and finally take the sum of the squared entries of that vector. Each step requires a total of $O(N)$ operations, adding up to an algorithm with linear runtime. Since $\Psi_3(\mathcal{W}^N)$ can be computed from this integral in constant time, we thus obtain an algorithm for $\Psi_3(\mathcal{W}^N)$ (and hence $d_3(E_1,\dots,E_N)$) with linear runtime.

Linear implementation of the K-depth (general K). For $K \geq 4$, the calculations in this paper can be used to compute an asymptotically equivalent form of the K-depth in linear time. We first summarize how such an algorithm can be implemented and discuss its correctness afterwards. When neglecting the $o_P(1)$-term in Lemma 2.8, an asymptotically equivalent value of $N\left(d_K(E_1,\dots,E_N) - (1/2)^{K-1}\right)$ can be computed as follows:

1. For $j = 0,\dots,K-2$, compute:
$$S_j = (S_j(1),\dots,S_j(N)) \quad \text{with} \quad S_j(n) := \sum_{i=1}^{n} \left(\frac{i}{N}\right)^j \psi(E_i) \quad \text{for} \quad n = 1,\dots,N.$$
2. Compute: $\widetilde{S} := \sum_{j=0}^{K-2} \sum_{n=2}^{N} \binom{K-2}{j} \left(\frac{1}{2} - \frac{n}{N}\right)^j \psi(E_n) S_{K-2-j}(n-1).$
3. Return: $-\frac{N^{K-2}}{\prod_{n=1}^{K-1}(N-n)} \cdot \frac{K(K-1)}{2} \cdot \widetilde{S}.$

Note that computing S_j for all $j = 1,\dots,K-2$ in advance enables us to compute $\widetilde{S}$ in Step 2 in linear time and thus leads to an algorithm with time complexity $O(N)$. Due to its linear time complexity, we refer to this implementation as the *linear implementation.* In order to see that this algorithm indeed yields $N\left(d_K(E_1,\dots,E_N) - (1/2)^{K-1}\right)$ asymptotically, we first recall that if $E_1,\dots,E_N$ satisfy (A1) and (A2) then, according to Lemma 2.8, $N\left(d_K(E_1,\dots,E_N) - (1/2)^{K-1}\right)$ is asymptotically equivalent to

$$\widetilde{d}_K(E_1,\dots,E_N) := -\frac{N^K}{\langle N\rangle_K}\frac{K(K-1)}{4} \sum_{1\leq n_1\neq n_2\leq N} \left(\frac{1}{2} - \frac{|n_2-n_1|}{N}\right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}. \qquad (18)$$

We claim that the calculations in Steps 1-3 indeed yield $\widetilde{d}_K(E_1,\dots,E_N)$. By symmetry, the sum over all $n_1 \neq n_2$ is equal to twice the sum over all $n_1 < n_2$, hence

$$\widetilde{d}_K(E_1,\dots,E_N) = -\frac{N^{K-1}}{\langle N\rangle_K}\frac{K(K-1)}{2} \sum_{1\leq n_1< n_2\leq N} \left(\frac{1}{2} - \frac{n_2-n_1}{N}\right)^{K-2} \psi(E_{n_1})\psi(E_{n_2}).$$

Since the leading factor coincides with the factor in Step 3 of the algorithm (note that $\langle N\rangle_K = N \cdot \prod_{n=1}^{K-1}(N-n)$), it only remains to show

$$\sum_{1\leq n_1< n_2\leq N} \left(\frac{1}{2} - \frac{n_2-n_1}{N}\right)^{K-2} \psi(E_{n_1})\psi(E_{n_2}) = \widetilde{S}.$$

This can be done by summing over n_1 and n_2 separately and by using the binomial theorem to expand $\left(\frac{1}{2} - \frac{n_2-n_1}{N}\right)^{K-2}$:

$$\begin{aligned}&\sum_{1\le n_1<n_2\le N}\left(\frac{1}{2}-\frac{n_2-n_1}{N}\right)^{K-2}\psi(E_{n_1})\psi(E_{n_2})\\&=\sum_{n_2=2}^{N}\sum_{n_1=1}^{n_2-1}\sum_{j=0}^{K-2}\binom{K-2}{j}\left(\frac{1}{2}-\frac{n_2}{N}\right)^{j}\left(\frac{n_1}{N}\right)^{K-2-j}\psi(E_{n_1})\psi(E_{n_2}),\end{aligned}$$

which equals $\widetilde{S}$ after rearranging the order of summation. Hence the algorithm based on Steps 1-3 indeed returns $\widetilde{d}_K(E_1, \ldots, E_N)$ and thus approximates the rescaled K-depth.

The runtimes of the K-depth are compared under various sample sizes up to 100 for $K \in \{3, 4, 5\}$ in Figure 2. The median of the runtimes (in milliseconds) from one hundred measured repetitions are considered. Since (nested) loops are extremely inefficient when using the software `R`, we implemented the naive algorithm in `C++` and only used a wrapper function to call this code in `R`. However, even with this speedup via a `C++` implementation, the difference between the performances of the naive- and the linear implementation is still immense. Therefore, the runtimes presented in Figure 2 are given on a logarithmic scale (with base 10). The three plots representing the 3-, 4-, and 5-depth each contain the logarithmic runtime of the naive implementation in black (solid lines) and the one of the linear implementation in red (dashed lines). As expected from an algorithm with runtime $\Theta(N^K)$, the effort to calculate the K-depth with the naive implementation increases significantly as K gets larger. On the one hand, the graphics show this increase. On the other hand, differences between the log-runtimes from the linear implementation can barely be seen since it has linear time complexity for all K. The linear implementation should be chosen especially for high sample size. These results confirm our theoretical analysis of the time complexity.

Since the linear implementation only yields the K-depth up to an asymptotically negligible error, we end the section with a discussion on how large this error is. To this end, we consider an input $E_1, \ldots, E_N$ satisfying (A1) and (A2). The aim is to study the absolute difference between

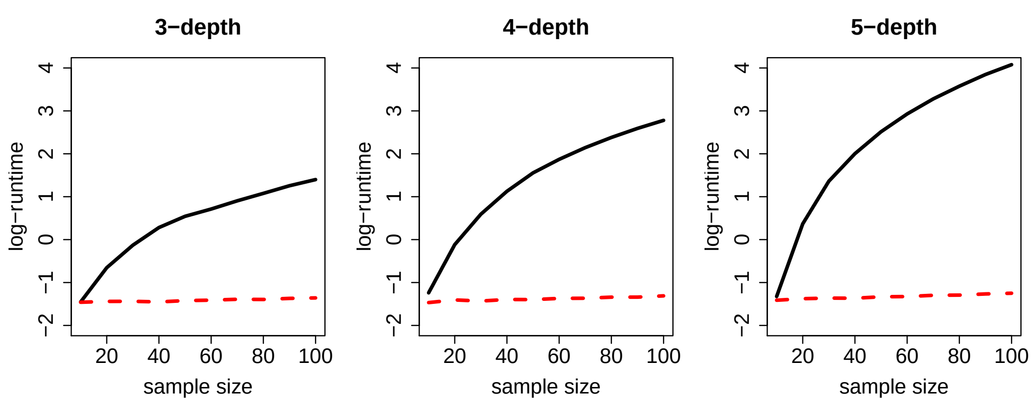

Figure 2: Logarithms of the runtimes for K-depth with $K \in \{3, 4, 5\}$ (black solid lines: naive C++ implementation, red dashed lines: linear implementation).

the exact rescaled K-depth given in (3) and the value of the linear implementation given in (18). More formally, if $X_{N,K}$ denotes the exact rescaled K-depth and $Y_{N,K}$ denotes the value of the linear implementation then the aim is to study the behavior of the random variable $|X_{N,K} - Y_{N,K}|$. Figure 3 contains box plots of this absolute difference for $K \in \{4, 5\}$ and various values of N based on one hundred repetitions of $X_{N,K}$ and $Y_{N,K}$. These box plots reveal that, unless N is fairly small, the absolute difference between the exact K-depth and the value of the linear implementation is negligibly small. Hence, in order to avoid larger errors, one should only use the linear implementation for sufficiently large N, e.g. $N \geq 25$ for $K = 4$ or $N \geq 50$ for $K = 5$.

Remark 3.1. An alternative approach to avoid any errors is to directly use

$$\widetilde{d}_K(E_1, \ldots, E_N) := -\frac{N^K}{\langle N \rangle_K} \frac{K(K-1)}{4} \sum_{1 \leq n_1 \neq n_2 \leq N} \left(\frac{1}{2} - \frac{|n_2 - n_1|}{N} \right)^{K-2} \frac{\psi(E_{n_1})\psi(E_{n_2})}{N}$$

as a test statistic rather than the rescaled K-depth itself. This statistic still converges in distribution to $\Psi_K(W)$ if assumptions (A1) and (A2) hold. Hence it has the same asymptotic quantiles as the rescaled K-depth. Although its usefulness for applications is not immediately apparent, its efficient computation via the linear implementation makes it more practical to use.

3.2 Quantiles and limit distribution

Unfortunately, it is unclear how to compute the quantiles of the limit distributions $\Psi_K(W)$ in Theorem 2.2 analytically. We therefore add simulations to approximate their quantiles and also provide estimations of their density functions to highlight some of the properties of $\Psi_K(W)$. Since simulating a Brownian motion and computing the integral of its transformation numerically still leads to a (small) approximation error, we decided to take the simpler route of simulating the K-depth of a large number N of residuals $E_1, \ldots, E_N$ instead.

More precisely, for $K \in \{3, 4, 5\}$, we consider 10^6 realizations of the K-depth of realizations from $N = 500$ uniformly distributed independent random variables in [-1,1] using the linear imple-

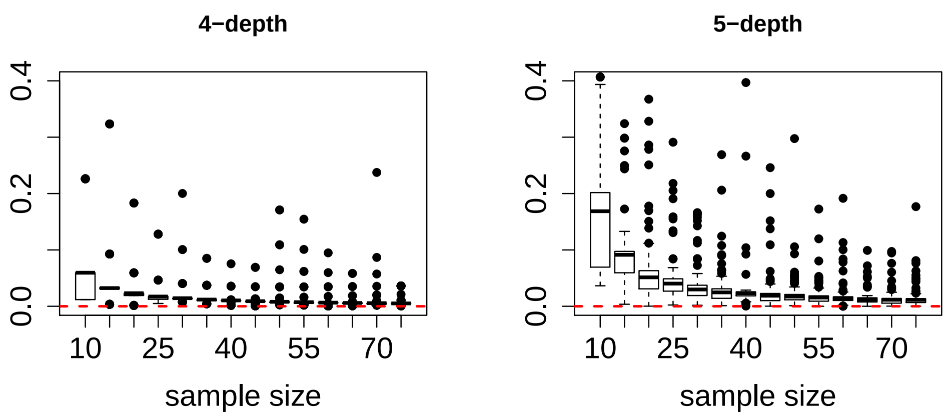

Figure 3: Absolute errors between the exact rescaled K-depth and its approximated version.

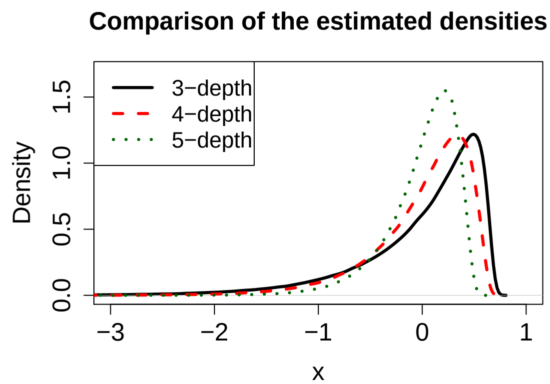

Figure 4: Density plots of the asymptotic distribution of the K-depth for $K \in \{3, 4, 5\}$.

mentation. Note that the uniform distribution is a somewhat arbitrary choice among distributions that satisfy (A2). In Figure 4, the estimated densities of $\Psi_K(W)$ are presented by using those 10^6 realizations. The kernel density estimation is done by the `R`-function `density()` with the default settings. The distributions are asymmetric and have the upper bound $\frac{K(K-1)}{2^K}$, see Leckey et al. (2019). Also note that the distributions tend to be more concentrated around zero as K increases.

Table 1 shows various α-quantiles of the asymptotic distribution. According to the estimated densities, the quantiles are closer to zero for higher K. Assuming that N is sufficiently large, these quantiles can be used as critical values for testing hypotheses with the K-depth, see also Leckey et al. (2019).

Table 1: α-quantiles $q_{\alpha,K}$ of the limit distribution for $K \in \{3, 4, 5\}$.

α	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
$q_{\alpha,3}$	-2.200	-1.645	-1.222	-0.800	0.557	0.603	0.633	0.658
$q_{\alpha,4}$	-1.637	-1.234	-0.931	-0.624	0.470	0.525	0.563	0.597
$q_{\alpha,5}$	-1.122	-0.856	-0.653	-0.445	0.349	0.396	0.430	0.463

Note that for small sample sizes, the asymptotic quantiles can be a poor approximation of the correct critical values for tests based on the K-depth. In this case, one should rather compute the exact quantiles of the rescaled K-depth $N(d_K(E_1, \ldots, E_N) - (1/2)^{K-1})$ by considering all possible 2^N sign constellations in order to compute all point masses of the distribution. This can also be done for the test statistic in Remark 3.1 which has a slightly different distribution than (3) although both converge to the same asymptotic distribution if (A1) and (A2) hold. In particular if K is large, we recommend to always use the test statistic in Remark 3.1 instead of the exact K-depth, even for small sample sizes N, in order to speed up the computation significantly.

4 Conclusion

The K-depth is a general tool for analyzing residuals where an ordering is possible. It is not limited to time series and models with univariate regressors since approaches of ordering multivariate data exist, see Horn and Müller (2020). Due to the fact that the K-depth is distribution-free under independent residuals with median zero, we do not need additional distributional assumptions on the errors of the model and thus we can use it for testing hypotheses in many situations. The sign function in the K-depth robustifies the test statistic against outliers, and moreover, changes of variances do not effect the test statistic.

For large sample sizes N, there are two problems: the $\Theta(N^K)$ complexity of calculating the K-depth, and the asymptotic distribution of the test statistic is necessary to get critical values of the test. This paper considers a derivation of the limit distribution of the K-depth by improving and generalizing the proof idea in Kustosz et al. (2016a) for $K = 3$. Moreover, the derivation of the asymptotics provides a possibility to compute an asymptotic equivalent version of the K-depth in linear time. Simulations indicate that the differences between the exact and asymptotic version of the K-depth are already for medium sample sizes irrelevantly small.

Further research will consider how the parameter K effects the power of the tests depending on various models. In Kustosz et al. (2016a) and Leckey et al. (2019), we have already done such simulation studies. Now, we have the possibility for much faster computations.

Note that the sign function ψ in the random walk $\mathcal{W}^N$ in (2) can be replaced by any other score function and the convergence of $\Psi_K(\mathcal{W}^N)$ to $\Psi_K(W)$ would still hold as long as Donsker's invariance principle is still applicable. Hence one could also consider a test statistics $\Psi_K(\widetilde{\mathcal{W}}^N)$ where $\widetilde{\mathcal{W}}^N$ is defined as in (2) but with a different score function $\widetilde{\psi}$ than the sign function ψ. Considering ranks or other scores for $\widetilde{\psi}$ is a work in progress. Moreover, the independence assumption on the residuals can be weakened, see Remark 2.4. Other aspects, we look ahead to, concern the investigation of tests for independence based on the K-depth, the influence of the ordering in the multivariate case, or studies on the robustness, breakdown points etc.

Generally, the idea to define a new depth concept by the asymptotically equivalent representation in Lemma 2.8 (see Remark 3.1) should be mentioned. This can also lead to other depth concepts by replacing Ψ_K with a different functional $\widetilde{\Psi}$. This can also be combined with different choices for the score function $\widetilde{\psi}$ as mentioned in the previous paragraph. Hence the test statistic in Remark 3.1 can be considered as an example of a general framework for defining residual-based tests where the robustness can be adjusted by the choice of the score function $\widetilde{\psi}$ and a model-based choice for $\widetilde{\Psi}$ can help increasing the power of the test.

Acknowledgements

The authors gratefully acknowledge support from the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823, B5) of the German Research Foundation (DFG).

References

Arcones, M. A., Chen, Z., and Gine, E. (1994). Estimators related to U-processes with applications to multivariate medians: Asymptotic normality. *The Annals of Statistics*, 22(3):1460–1477.

Arcones, M. A. and Gine, E. (1993). Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542.

Bai, Z.-D. and He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *The Annals of Statistics*, 27(5):1616–1637.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Carothers, N. (2000). *Real Analysis*. Cambridge University Press.

Chen, D., Devillers, O., Iacono, J., Langerman, S., and Morin, P. (2013). Oja centers and centers of gravity. *Computational Geometry*, 46(2):140 – 147.

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109:411–423.

Denecke, L. and Müller, C. H. (2011). Robust estimators and tests for copulas based on likelihood depth. *Computational Statistics and Data Analysis*, 55:2724–2738.

Denecke, L. and Müller, C. H. (2012). Consistency and robustness of tests and estimators based on depth. *Journal of Statistical Planning and Inference*, 142(9):2501 – 2517.

Dong, Y. and Lee, S. M. S. (2014). Depth functions as measures of representativeness. *Statistical Papers*, 55(4):1079–1105.

Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statistics & Probability Letters*, 14:119–128.

Gibbons, J. and Chakraborti, S. (2003). *Nonparametric statistical inference*. Statistics, textbooks and monographs. Marcel Dekker Incorporated.

Gould, H. W. and Srivastava, H. M. (1997). Some combinatorical identities associated with the vandermonde convolution. *Appl. Math. Comput.*, 84(2-3):97–102.

Horn, M. and Müller, C. H. (2020). Tests based on sign depth for multiple regression. *SFB Discussion Paper 07/20*. `https://www.statistik.tu-dortmund.de/2630.html`.

Hu, Y., Wang, Y., Wu, Y., Li, Q., and Hou, C. (2011). Generalized Mahalanobis depth in the reproducing kernel Hilbert space. *Statistical Papers*, 52(3):511–522.

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., 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.

Leckey, K., Malcherczyk, D., and Müller, C. H. (2019). Powerful generalized sign tests based on sign depth. *Submitted*.

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.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18:405–414.

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260.

Liu, X., Rahman, J., and Luo, S. (2019). Generalized and robustified empirical depths for multivariate data. *Statistics & Probability Letters*, 146:70 – 79.

López-Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51(10):4957–4968.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.

López-Pintado, S., Sun, Y., Lin, J. K., and Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification*, 8(3):321–338.

Mizera, I. (2002). On depth and deep points: A calculus. *The Annals of Statistics*, 30(6):1681–1736.

Mizera, I. and Müller, C. H. (2004). Location-scale depth (with discussion). *Journal of the American Statistical Association*, 99:949–966.

Mosler, K. (2002). *Multivariate dispersion, central regions and depth. The Lift Zonoid Approach.* Lecture Notes in Statistics, 165, Springer, New York.

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.

Nagy, S. and Ferraty, F. (2019). Data depth for measurable noisy random functions. *Journal of Multivariate Analysis*, 170:95–114.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1(6):327–332.

R Core Team (2019). *R: A Language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rousseeuw, P. J. and Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association*, 94(446):388–402.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 2:523–531.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162.

Wang, J. (2019). Asymptotics of generalized depth-based spread processes and applications. *Journal of Multivariate Analysis*, 169:363–380.

Wellmann, R., Harmand, P., and Müller, C. H. (2009). Distribution-free tests for polynomial regression based on simplicial depth. *Journal of Multivariate Analysis*, 100(4):622 – 635.

Wellmann, R. and Müller, C. H. (2010a). Depth notions for orthogonal regression. *Journal of Multivariate Analysis*, 101(10):2358 – 2371.

Wellmann, R. and Müller, C. H. (2010b). Tests for multiple regression based on simplicial depth. *Journal of Multivariate Analysis*, 101(4):824 – 838.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics*, 31(5):1460–1490.

Zuo, Y. (2006). Multidimensional trimming based on projection depth. *The Annals of Statistics*, 34(5):2211–2251.