Measuring dependence between a scalar response and a functional covariate

Siegfried Hörmann* and Daniel Strenger†
Institute of Statistics, Graz University of Technology

May 14, 2024

Abstract

We extend the scope of a recently introduced dependence coefficient between a scalar response Y and a multivariate covariate X to the case where X takes values in a general metric space. Particular attention is paid to the case where X is a curve. While on the population level, this extension is straight forward, the asymptotic behavior of the estimator we consider is delicate. It crucially depends on the nearest neighbor structure of the infinite-dimensional covariate sample, where deterministic bounds on the degrees of the nearest neighbor graphs available in multivariate settings do no longer exist. The main contribution of this paper is to give some insight into this matter and to advise a way how to overcome the problem for our purposes. As an important application of our results, we consider an independence test.

1 Introduction

Assume that Y is a real random variable with distribution function $F(t)$. For some covariate X, let $G_X(t) = P(Y \geq t|X)$. Dette et al. (2013), Chatterjee (2020) and Azadkia and Chatterjee (2021) have studied a dependence coefficient between Y and X, which, for continuous F, can be expressed as

$$T(X,Y) := 6 \times \int \mathrm{Var}(G_X(t))dF(t).$$

Note that if X and Y are independent, then $G_X(t) = G(t) = P(Y \geq t)$, which is non-random, and hence $T(X,Y) = 0$. On the other hand, if $Y = f(X)$ for some measurable function f, then $G_X(t) = 1\{Y \geq t\}$ and hence $T(X,Y) = 6\int F(t)(1-F(t))dF(t) = 1$. Azadkia and Chatterjee (2021) show that $0 \leq T(X,Y) \leq 1$. Moreover, $T(X,Y) = 0$ if and only if X and Y are independent and $T(X,Y) = 1$ if and only if $Y = f(X)$.

In their seminal paper, Dette et al. (2013) have handled the case of univariate X, where (X,Y) has continuous marginal distributions. Chatterjee (2020) has extended their work to arbitrary marginals and proposed a new tuning parameter-free estimator, while Azadkia and Chatterjee (2021) have treated a further generalization allowing for multivariate X and conditional dependence. Since then, the dependence coefficient $T = T(X,Y)$ has attracted a lot of attention. Cao and Bickel (2020) relate it to the maximal correlation coefficient and study similar measures that are able to detect functional relationships of

*shoermann@tugraz.at
†daniel.strenger@tugraz.at

prespecified shape. Much research has been conducted on the asymptotic behavior of related estimators and independence tests: Shi et al. (2024) prove the asymptotic normality of the estimator proposed by Azadkia and Chatterjee (2021) when X and Y are independent and reveal a weakness in the test's capability to identify local alternatives, a matter that is also treated by Shi et al. (2022), Bickel (2022) and Lin and Han (2023). The latter two papers also advise a way to overcome these issues. Auddy et al. (2024) investigate for which kind of contiguous alternatives the test possesses non-trivial power. As T is non-symmetric, Zhang (2023) studies the asymptotic behavior of a symmetrized version proposed by Chatterjee (2020) under the null hypothesis of independence. For a detailed review of recent developments, we refer to Chatterjee (2023).

Motivated by our research in functional data analysis (FDA), the purpose of this article is to extend the scope of the dependence coefficient T to covariates X taking values in a function space. What we have in mind are curves $X = (X(u)\colon u \in \mathcal{U})$, where $\mathcal{U}$ is some continuum, typically an interval. A very common assumption in FDA is that $X \in L^2(\mathcal{U})$, the space of square integrable functions on $\mathcal{U}$, but our main results below will remain true for a general separable metric space (H, d). In fact, it is not hard to see that the population properties of T quoted above and qualifying it as a dependence measure remain valid in any metric space.

While some nice theoretical properties of T are easily seen to remain valid in very general settings, the measure is only operational if we can find an empirical version which is able to consistently estimate T from a sample. Hence, in this paper we pursue the following goals:

1. Extend the consistency of the estimator suggested in Azadkia and Chatterjee (2021) when X takes values in a separable metric space.
2. Derive the limiting law of this estimator when Y and X are independent.
3. Use the result in 2. to establish a statistical test for independence of the covariate X and the response Y.

Attempts to generalize Azadkia and Chatterjee (2021) to settings where X and Y can take values in a broad class of topological spaces have been made previously by Deb et al. (2020) (unconditional dependence) and Huang et al. (2022) (conditional dependence). Those authors consider measures of association based on kernels and geometric graphs, the most prominent example being the k-nearest neighbor graph. Deb et al. (2020) show that the measure T can be viewed as a special case (see their Proposition 8.2). However, their results rely on assumptions on the underlying graphs (see Remarks 3 and 4) which may be difficult to verify and which generally may fail in infinite-dimensional spaces. The bottleneck typically arises from bounds on the maximum degree. Our Theorem 3 below gives some insight in this matter and shows that nearest neigbor graphs for functional data can have maximal degrees which do not just diverge with sample size, but their rate of divergence can be rather fast. This result easily extends to k-nearest neighbor graphs and minimal spanning trees (see Remark 5) and illustrates that assumptions that have been used previously in the literature may fail in infinite-dimensional settings. In addition, it shows that the arguments of Azadkia and Chatterjee (2021) do not generalize to infinite-dimensional spaces.

The subsequent sections are organized as follows: In Section 2, we will briefly outline the estimation approach in Azadkia and Chatterjee (2021) and explain the difficulties when it comes to studying the asymptotic properties of this estimator in infinite-dimensional spaces. In Section 3, we prove consistency and obtain the limiting distribution under independence. The resulting independence test is shown to be universally consistent. Moreover, for any given covariance operator, we construct corresponding functional random samples and give upper and lower bounds on the maximal degree of the resulting nearest neighbor graphs. We illustrate the empirical performance of our theory in Section 4 and give the proofs in Section 5. Some additional tables and graphs from comprehensive simulations are provided in Appendix A.

2 An estimator for $T(X,Y)$

For the remainder of the paper, we impose some mild technical requirements that streamline our presentation. To this end, define $H(t) := P(d(X,X') \leq t)$, where X' an independent copy of X.

Assumption 1. *We have*

(a) $H(t)$ *is continuous;*

(b) $F(t)$ *is continuous;*

(c) for $P_Y := P \circ Y^{-1}$ *almost all* t*, the mapping* $x \mapsto G_x(t)$ *is continuous* $P_X := P \circ X^{-1}$ *almost everywhere.*

Assumption 1 constitutes a set of continuity conditions on the distribution of (Y,X). Continuity assumptions have also been used in Dette et al. (2013), while Azadkia and Chatterjee (2021) work under general distributional assumptions. We note that Assumption 1 could be relaxed, but in the context of functional data, our assumptions are reasonably general and common. For example, a violation of (a) would arise if P_X is a discrete measure, which is quite uncommon for a functional data model.

Assumption 1 (b) implies that $\mathrm{Var}(G_X(t)) = \mathbb{E}\, G_X^2(t) - G^2(t)$ and $\int G^2(t)dF(t) = 1/3$. Hence, estimation of $T(X,Y)$ reduces to estimation of

$$Q(X,Y) := \int \mathbb{E}\, G_X^2(t)dF(t). \tag{1}$$

Consider a random sample $\{(X_i,Y_i), 1 \leq i \leq n\}$ with $(X_i,Y_i) \sim (X,Y)$. For $i \in \{1,\dots,n\}$, let $N(i) = N_n(i)$ be the index of the nearest neighbor of X_i in the sample of covariates $X_1,\dots,X_n$, that is, $d(X_i, X_{N(i)}) \leq d(X_i,X_j)$ for all $j \neq i$. Assumption 1 (a) implies that $N(i)$ is unique with probability one. If n is large, we expect that $X_{N(i)}$ is close to X_i and therefore, considering Assumption 1 (c), we get by some heuristics that for almost all t

$$\mathbb{E}\, G_X^2(t) \approx \mathbb{E}\, G_{X_i}(t) G_{X_{N(i)}}(t). \tag{2}$$

Moreover, noting that

$$\begin{aligned} \mathbb{E}\, G_{X_i}(t) G_{X_{N(i)}}(t) &= \mathbb{E}\left(\mathbb{E}\left[1\{Y_i \geq t\}1\{Y_{N(i)} \geq t\} | X_1,\dots,X_n\right]\right) \\ &= \mathbb{E}\, 1\{Y_i \geq t\}1\{Y_{N(i)} \geq t\}, \end{aligned}$$

we obtain the approximation

$$Q(X,Y) \approx E \int 1\{Y_i \geq t\}1\{Y_{N(i)} \geq t\} dF(t) = E \min\{F(Y_i), F(Y_{N(i)})\}.$$

This motivates the estimators

$$\widehat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} \min\{F_n(Y_i), F_n(Y_{N(i)})\} \quad \text{and} \quad \widehat{T}_n = 6\,\widehat{Q}_n - 2, \tag{3}$$

with F_n being the empirical distribution function of $Y_1, \dots, Y_n$.

Remark 1. *We note that $\widehat{T}_n$ is, in essence, the estimator given in Azadkia and Chatterjee (2021), where we have been taking into account that Y is supposed to have a continuous distribution.*

Remark 2. *Let $\mathcal{G}_n = \mathcal{G}_n(X_1, \dots, X_n)$ be the nearest neighbor graph related to the sample $X_1, \dots, X_n$. Then $\widehat{T}_n$ is a functional of the responses $Y_1, \dots, Y_n$ and of $\mathcal{G}_n$.*

A non-trivial question is whether the heuristics leading to $\widehat{T}_n$ can be rigorously justified, i.e. whether $\widehat{T}_n$ is a consistent estimator of $T(X,Y)$. Azadkia and Chatterjee (2021) have shown that the answer is affirmative in the case of $H = \mathbb{R}^p$. A closer inspection of their paper reveals that the proof is crucially based on the fact that within a set of arbitrary points $\mathcal{S}_n = \{x_1, \dots, x_n\} \subset \mathbb{R}^p$, an element $x_i \in \mathcal{S}_n$ can be the nearest neighbor to at most $k(p)$ points in $\mathcal{S}_n$, where $k(p)$ is some finite constant that is independent of n. Formally, define the maximum degree $L_n = L_n(\mathcal{S}_n) := \max_{1\leq\, i \leq\, n} L_{i,n}$, where $L_{i,n}$ is the number of elements in $\mathcal{S}_n \setminus x_i$ that have x_i as their nearest neighbor. Then

$$L_n \leq\ k(p) \quad \text{for all } n \geq\ 1. \tag{4}$$

For example, if we generate a sequence in $\mathbb{R}$ then trivially $L_n \leq\ 2$. In $\mathbb{R}^2$ it is not hard to see that $L_n \leq\ 6$. Kabatjanski and Levenstein (1978) have shown that $k(p) \leq\ \text{const} \times \gamma^p$ for a specific $\gamma > 1$ and for all $p \geq\ 1$.

In this paper, we are interested in an infinite-dimensional covariate space. Here we generally cannot bound L_n by a constant. Consider, for example, an orthonormal sequence of elements x_k, $k \geq\ 2$, in $L^2(\mathcal{U})$, i.e. $\int_{\mathcal{U}} x_k(u)x_\ell(u)du = \delta_{k,\ell}$, where $\delta_{k,\ell}$ denotes the Kronecker delta. Then $d(x_k, x_\ell) = \sqrt{2}(1 - \delta_{k,\ell})$. If we set $x_1 = 0$, then x_1 is the nearest neighbor to all other elements, and thus $L_n = n - 1$. In this example, however big is n, $d(x_i, x_{N(i)})$ is not becoming small, and the heuristics that lead to (2) are no longer applicable. This illustrates that in infinite-dimensional spaces, we are not just facing an extra technical challenge for proving convergence of $\widehat{T}_n$, but that consistency may be at stake if we cannot control L_n. For insightful discussions on the problematic usage of nearest neighbor methods in high dimension, we refer to Beyer et al. (1997) and Durrant and Kabán (2009).

Remark 3. *The empirical dependence measures proposed in Deb et al. (2020) and Huang et al. (2022) are based on general graph functionals $\mathcal{G}_n = \mathcal{G}_n(X_1, \dots, X_n)$. Both require that the maximal and the minimal degrees of $\mathcal{G}_n$ be of the same order of magnitude (see their Assumptions (A3) and (12), respectively). In the case of a directed nearest neighbor graph, this means that L_n has to be bounded (since the minimum degree cannot be greater than 1), which is generally not fulfilled for infinite-dimensional data. See Theorem 3 below.*

Let us illustrate this problem on the data example that will be presented in Section 4.4. These data consist of the age distributions of $n = 2117$ Austrian municipalities. In Figure 1 we see one curve which turns out to be the nearest neighbor of 66 other curves, which corresponds to $\approx 3\%$ of the sample size. The example confirms that our theoretical issue is also relevant in practice and that we may get nearest neighbor graphs with some rather large degrees.

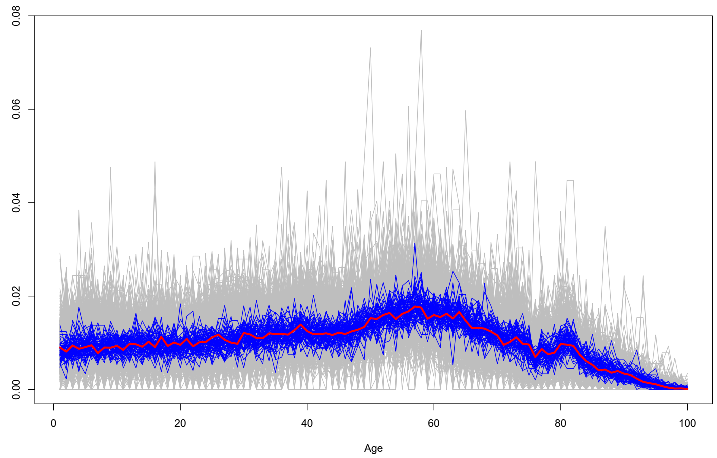


Figure 1: Age distribution curves for 2117 municipalities in Austria. On the y-axis we see the proportions. The curve related to Wolfsberg (solid red) is the nearest neighbor of the age distribution curves of 66 municipalities (blue).

3 Main results

Our first main result establishes weak consistency of $\widehat{T}_n$ for i.i.d. data under the continuity properties stated in Assumption 1.

Theorem 1. *Assume that X takes values in a separable metric space (H, d). Let Assumption 1 hold. Then we have that*

$$\widehat{T}_n \xrightarrow{\mathcal{P}} T(X, Y) \quad as \quad n \to \infty. \tag{5}$$

We note that Azadkia and Chatterjee (2021) have shown that in the finite-dimensional setup almost sure convergence can be obtained in (5). We do not pursue an improvement in this direction, but are rather targeting for weak convergence to a limiting distribution. This will be important in the realm of statistical applications, particularly in its utilization to develop an independence test.

Assuming that X has a density and is independent of Y, Shi et al. (2024) have shown that $\sqrt{n}\widehat{T}_n$ will be asymptotically normally distributed. The necessity for X to possess a density function serves as an initial indication that their result does not directly generalize to infinite-dimensional X. Also, the limiting variance constitutes a rather involved expression and crucially depends on the geometry of the Euclidean space. It is deduced from results of Henze (1987) and involves the limiting law of $L_{1,n}$. Inspired by Deb et al. (2020) and Lin and Han (2022), we overcome the latter problem, by using a data-dependent self-normalization. In order to account for potentially diverging L_n, we are imposing the following high-level assumption.

Assumption 2. *We have $L_n = o_P(n^{\frac{1}{4}})$.*

Before we discuss Assumption 2 in more detail, we will state our next result.

Theorem 2. *Let Assumptions 1 and 2 hold. Assume that X_i and Y_i are independent. Then there is a random variable $W_n = W_n(X_1, \ldots, X_n)$ such that*

$$\sqrt{\frac{n}{36\, W_n}}\widehat{T}_n \xrightarrow{\mathcal{D}} N(0, 1), \quad n \to \infty.$$

The variable W_n is explicitly defined in (12).

Remark 4. *Both, Deb et al. (2020) and Lin and Han (2022), have obtained a CLT for $\widehat{T}_n$ (or generalizations) using a data-dependent self-normalization step. Although both results are impressive in their generality, they are not suitable for our setting: As mentioned above, Assumption (A3) in Deb et al. (2020) implies that L_n is bounded in case of nearest neighbor graphs. But even if other graph functionals are used, which might allow for growing L_n, this growth is limited only to a polylogarithmic rate, which may be quite limiting as our Theorem 3 suggests. Lin and Han (2022), on the other hand, show normality even under dependence of Y on X, but since their result is based on the nearest neighbor CLT by Chatterjee (2008), L_n is again required to be bounded by a constant.*

If the target is to test

$$\mathcal{H}_0 \colon Y_i \text{ and } X_i \text{ are independent} \quad \text{v.s.} \quad \mathcal{H}_A \colon \mathcal{H}_0 \text{ doesn't hold,}$$

we can use the test statistics $\mathcal{I}_n := \sqrt{\frac{n}{36\,W_n}}\widehat{T}_n$. If n is sufficiently large, we reject at significance level α if $\mathcal{I}_n > z_{1-\alpha}$, where z_α is the α-quantile of a standard normal variable. The following corollary gives conditions when this test is consistent.

Corollary 1. *Let Assumptions 1 and 2 hold. Assume that X_i and Y_i are not independent. Then $\mathcal{I}_n \to \infty$ in probability as $n \to \infty$.*

As presented in Section 4, the suggested test for independence shows reasonable power against fixed alternatives in practice. However, the lack of power against local alternatives as mentioned above remains a potential issue. A modification of the estimator to make use of several nearest neighbors, as done by Lin and Han (2023), seems promising, but is beyond the scope of this paper.

The data points from the counterexample in Section 2 form an infinite series of orthogonal functions, and thus are distributed much differently from a random sample. In our next results, we show that there are functional random samples where on the one hand L_n diverges and on the other hand Assumption 2 can be verified. To this end, we consider random elements X_i taking values in $L^2([0,1])$. The space $L^2([0,1])$ is equipped with the inner product $\langle x, y\rangle = \int_0^1 x(u)y(u)du$ and the corresponding norm $\|x\|^2 = \langle x, x\rangle$, which gives rise to the distance $d(x,y) = \|x - y\|$. Let $\Sigma = \mathrm{Var}(X_1)$ be the covariance operator of X_1, that is

$$\Sigma(v) = \mathbb{E}\,(X_1 - EX_1)\langle X_1 - EX_1, v\rangle$$

and denote by $\lambda_1 \geq \lambda_2 \geq \cdots$ its eigenvalues and by $e_1, e_2, \ldots$ corresponding eigenfunctions. Moreover, recall that a function $\ell(x)$ is called slowly varying at ∞ if $\ell(yx)/\ell(x) \to 1$ for any $y > 0$ and $x \to \infty$. We write $\ell \in R_0$. It is a well known fact that $\ell \in R_0$ implies that $\ell(x) = o(x^\delta)$ for any $\delta > 0$. We use the notion $a_n = \omega(b_n)$ if $|a_n/b_n| \to \infty$.

Theorem 3. *Let Σ be a symmetric, positive semidefinite operator on $L^2([0,1])$ having eigenvalues λ_k. Suppose that Σ is trace-class, i.e., $\sum_{k\geq 1}\lambda_k < \infty$. Then there is a random variable $X \in L^2([0,1])$ with covariance operator $\mathrm{Var}(X) = \Sigma$, such that for a random sample $X_1, X_2, \ldots$ with $X_i \overset{iid}{\sim} X$ we have the following:*

(i) If $\lambda_k = o\left(\sum_{j\geq k}\lambda_j\right)$, then $L_n = L_n(X_1, \ldots, X_n) \to \infty$ in probability.

(ii) If $\lambda_k = \ell(k)k^{-a}$ with $\ell \in R_0$, then there is an $h \in R_0$, such that $L_n = \omega\Big(h(n)n^{\frac{1}{1-2a}}\Big)$ in probability.

(iii) If in (ii) we have $a > 9/2$, then $L_n = o(n^{1/4})$, and hence Assumption 2 is fulfilled.

Remark 5. *It is easy to see from the proof of the Theorem 3, that in our construction of the random sample the maximum degree L_n diverges even faster if, instead of the 1-nearest neighbor graph, one uses k-nearest neighbor graphs with $k > 1$ (possibly growing) or minimum spanning trees. These are the examples discussed by Deb et al. (2020).*

Combining (ii) and (iii) in Theorem 3 implies that there exist non-trivial examples, in which L_n diverges at a polynomial rate, but where $\widehat{T}_n$ is asymptotically normal by our Theorem 2.

Next, we give a result that shows that Assumption 2 can be verified under some general conditions. To this end, we need another assumption, which holds e.g. for Gaussian processes.

Assumption 3. *The scores $Z_{1k} := \langle X_1, e_k \rangle$, $k \geq 1$, are independent and have density functions $f_k(s)$. Moreover, $f_1(s)$ and $f_2(s)$ are uniformly bounded over $s \in \mathbb{R}$.*

Theorem 4. *Let Assumption 3 hold and assume that the eigenvalues λ_k of $\mathrm{Var}(X)$ satisfy $\sum_{j \geq q} \lambda_j = O(\lambda_q)$ and $\lambda_{q_n} = o(n^{-6})$ when $q_n = \lfloor 0.899 \log n \rfloor$. Then Assumption 2 is fulfilled.*

The proof of Theorem 4 will follow from our general but slightly more technical Proposition 5 in Section 5.4.

4 Empirical investigations

In this section, we apply the dependence measure and the resulting test for independence to simulated and real world data. In Subsection 4.1 we compute $\widehat{T}_n$ in different scenarios and compare to the popular *distance correlation*, whose empirical version is here denoted by $\widehat{R}_n$ (see Definition 5 in Székely et al. (2007)). We refer to Dehling et al. (2020) for a detailed discussion of the applicability of distance correlation to discretized functional data. We follow this approach with $p = 200$ equidistant sampling points per curve. We will compare our independence test $\mathcal{I}_n$ to a permutation test based on distance correlation ($\mathcal{I}_n^{\mathrm{DC}}$) as well as to an independence test for functional data developed by García-Portugués et al. (2014) ($\mathcal{I}_n^{\mathrm{CvM}}$). The latter is a bootstrap test utilizing a Cramér-von-Mises type statistic. See Sections 4.2 and 4.3. We refer to Székely et al. (2007) and García-Portugués et al. (2014) for details on these competing approaches. Finally, in Section 4.4 we give an illustration of real-world data.

For our simulation study, we consider two different types of covariates X. In Setup (a), we let $X = \sum_{k=1}^{20} Z_k e_k(u)$ where $e_k(u) = \sqrt{2}\sin((k - 1/2)\pi u)$ and $Z_k \overset{\text{ind.}}{\sim} N(0, 0.3^k)$, modeling fast decaying eigenvalues (in fact $\lambda_k = 0$ for $k > 20$) and hence being in line with the assumptions in Theorem 4. In our Setup (b), we let X be a standard Brownian motion, which has the same principal components e_k as in Setup (a) and eigenvalues $\lambda_k = \frac{1}{\pi^2 (k-1/2)^2}$. Hence, the process does not satisfy the assumption of Theorem 4. Tables and graphics belonging to Setup (b) are deferred to Appendix A.

Our implementation of the test for independence is available via the R package `FDEP` from the second author's GitHub profile, Strenger (2024).

4.1 Dependence measure

In our first simulation exercise, we mimic several types of relationships between X and Y. We use sample sizes $n = 20$ (small) $n = 100$ (medium) and $n = 1000$ (large). Each time, we generate $B = 500$ samples to compute $\widehat{T}_n$ and $\widehat{R}_n$. The relation between Y and X is determined as $Y = f(X) + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ independent of X. We chose σ^2 such that $r^2 := \frac{\mathrm{Var}(f(X))}{\mathrm{Var}(Y)}$ is 100%, 90%, 50% and 10%, respectively. The functions $f : H \to \mathbb{R}$ are chosen as follows:

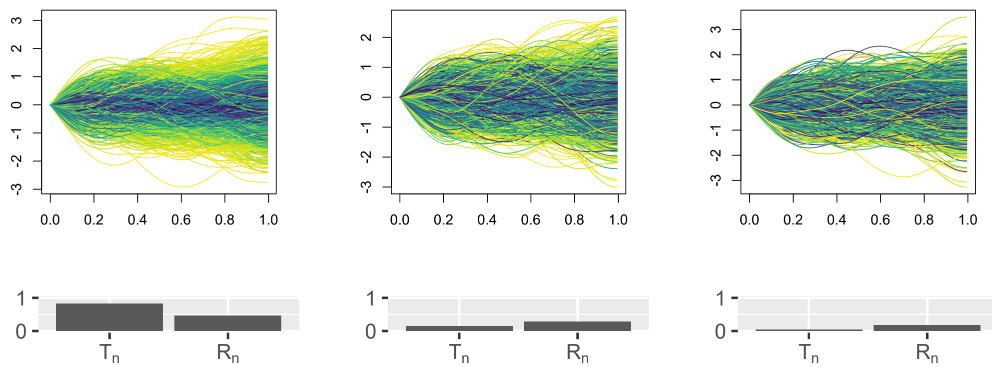
- $f(X) = 0$ (`ind`),
- $f(X) = \int_0^1 X(t)dt$ (`int`),
- $f(X) = \int_0^1 X(t)^2 dt$ (`sqnorm`),
- $f(X) = \int_0^1 t^2 X(t)dt$ (`weight`),
- $f(X) = \sin\left(2\pi \int_0^1 X(t)dt\right)$ (`sin`),
- $f(X) = \max_{t\in[0,1]} X(t)$ (`max`),
- $f(X) = \max_{t\in[0,1]} X(t) - \min_{t\in[0,1]} X(t)$ (`range`), and
- $f(X) = X(0.5)$ (`eval`).

For the Brownian motion, the scaling coefficients σ are computable in all cases (see Feller (1951) for `range`). If we do not have a closed form expression (as in Setup (a)), we can simply estimate it from the $n \times B$ signals $f(X_i^b)$ (i-th observation in the b-th sample). For calculating $\widehat{R}_n$, we use the R-package `energy` by Rizzo and Szekely (2022). Table 1 shows the means and standard deviations of the calculated values of $\widehat{T}_n$ and $\widehat{R}_n$ for sample size $n = 100$. The results for $n = 20$ and $n = 1000$ as well as for Setup (b) are provided in Tables 3-7 in Appendix A. In Figure 2 the functional relationships between Y and X under Setups (`sqnorm`) and (`sin`) are visualized. The visualizations of the other types of relationships can be found in Figures 7 and 8 in AppendixA.

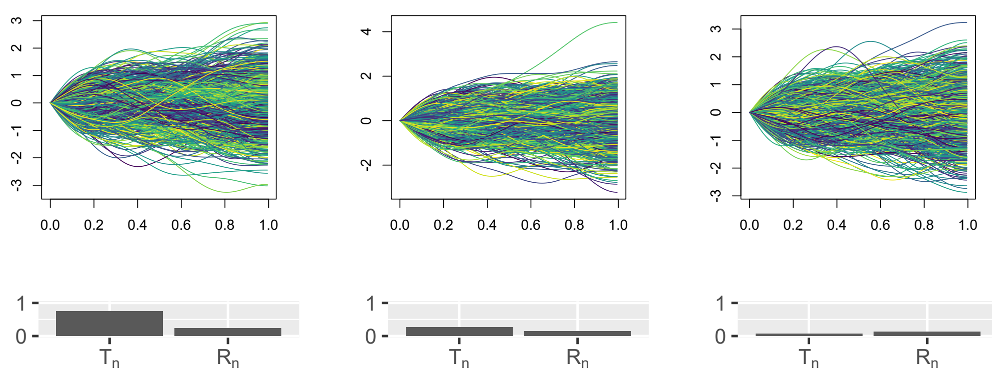
We can observe that $\widehat{R}_n$ takes larger values than $\widehat{T}_n$ in most cases, especially for high levels of noise. This indicates a higher ability to detect noisy relationships. A notable exception is the sine of integral relationship `sin`, for which $\widehat{T}_n$ takes higher values than $\widehat{R}_n$ at all levels of noise. This is in line with the observation in Chatterjee (2020), stating that the approach discussed here is powerful in situations where the relation between X and Y is 'oscillatory' in nature. Note that for small and medium sample sizes, $\widehat{R}_n$ seems to be strongly biased—observe the high value of $\widehat{R}_n = 0.194$ for independent data. This bias has mostly vanished at the large sample size $n = 1000$. Comparing the Setups (a) and (b), we notice that there does not seem to be a big difference in how the dependence measures behave, indicating that a fast (in fact exponential) decay of the eigenvalues, as stated in Theorem 4 is not necessary for the methods discussed to perform well.

4.2 Finite sample distribution and running times

Next, we investigate the distribution of the test statistic $\mathcal{I}_n$ under independence (setting `ind`). In Figure 3, we created 5000 samples of X – under Setup (b) – and Y – uniformly distributed on $[0, 1]$ and independent of X – and plot the histogram of the $\mathcal{I}_n$'s along



(a) $r^2 = 100\%$ (b) $r^2 = 50\%$ (c) $r^2 = 10\%$



(a) $r^2 = 100\%$ (b) $r^2 = 50\%$ (c) $r^2 = 10\%$

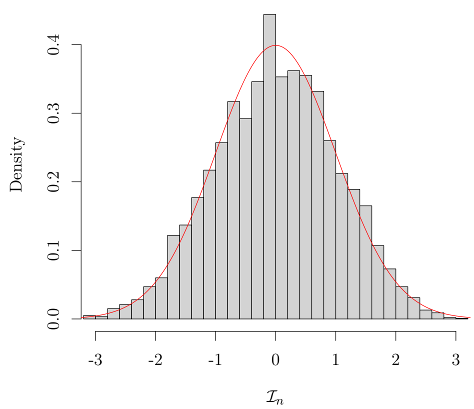
Figure 2: Visualization of the `sqnorm` (top) and `sin` (bottom) relationships under Setup (a). The $n = 1000$ curves X are colored according to Y. The bar plots compare the values of $\widehat{T}_n$ and $\widehat{R}_n$.

with the density of a standard normal distribution. Even for sample size $n = 20$, the quality of the normal approximation is quite reasonable. The p-values of the Shapiro-Wilk test for $n = 20, 100, 1000$ are 0.00001, 0.34 and 0.89, respectively, confirming the good approximation by the limiting law.

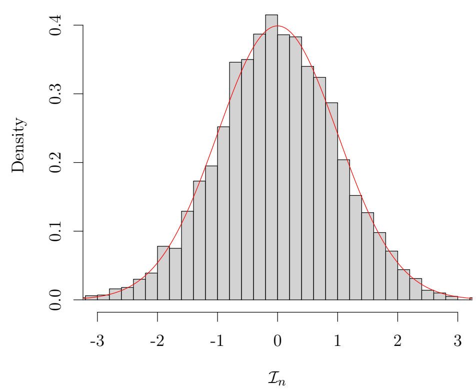
One of the advantages of $\mathcal{I}_n$ compared to the competing methods is that it comes with significantly shorter running time. Since $\mathcal{I}_n^{\mathrm{DC}}$ and $\mathcal{I}_n^{\mathrm{CvM}}$ are permutation and bootstrap tests, respectively, high computational costs are expected. With regard to the sample size, the number of operations for calculation of $\mathcal{I}_n^{\mathrm{DC}}$ is of order n^2, while it is only of order $n \log n$ for computing $\mathcal{I}_n$. For the simulation, we chose 200 resampling steps for the permutation and bootstrap tests. This is usually not considered enough (García-Portugués et al. (2014) recommend at least 5000 steps for their test). This small number was chosen to keep the computation times reasonable for the simulation experiment. Table 2 shows

	$\widehat{T}_n$				$\widehat{R}_n$			
`ind`	0 (0.12)				0.19 (0.02)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
`int`	0.86 (0.02)	0.66 (0.05)	0.29 (0.1)	0.05 (0.12)	0.94 (0.01)	0.88 (0.02)	0.65 (0.05)	0.34 (0.07)
`sqnorm`	0.64 (0.05)	0.4 (0.08)	0.14 (0.11)	0.01 (0.11)	0.5 (0.03)	0.47 (0.04)	0.36 (0.04)	0.25 (0.03)
`weight`	0.86 (0.02)	0.67 (0.05)	0.29 (0.1)	0.05 (0.11)	0.94 (0.01)	0.88 (0.02)	0.65 (0.05)	0.34 (0.07)
`sin`	0.57 (0.06)	0.49 (0.07)	0.27 (0.1)	0.04 (0.11)	0.27 (0.04)	0.27 (0.03)	0.25 (0.03)	0.22 (0.03)
`max`	0.69 (0.05)	0.52 (0.07)	0.22 (0.1)	0.03 (0.12)	0.84 (0.03)	0.79 (0.03)	0.59 (0.06)	0.32 (0.06)
`range`	0.44 (0.07)	0.34 (0.09)	0.12 (0.11)	-0.01 (0.11)	0.41 (0.03)	0.39 (0.03)	0.32 (0.04)	0.24 (0.03)
`eval`	0.78 (0.03)	0.63 (0.03)	0.28 (0.03)	0.05 (0.03)	0.83 (0.03)	0.79 (0.03)	0.58 (0.03)	0.32 (0.03)

Table 1: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 100$. Standard deviations are given in brackets.



(a) $n = 20$



(b) $n = 100$

Figure 3: Histogram of $\mathcal{I}_n$ under independence and comparison to the standard normal density.

the results. As expected, the computation time of $\mathcal{I}_n$ increases considerably slower than the time required to perform the other tests. For $n = 10000$, the $\mathcal{I}_n^{\text{CvM}}$-based test was aborted after 90 minutes.

n	$\mathcal{I}_n$	$\mathcal{I}_n^{\mathrm{DC}}$	$\mathcal{I}_n^{\mathrm{CvM}}$
100	<0.01	<0.01	0.98
500	0.01	0.07	4.15
1000	0.06	0.38	15.59
2000	0.25	1.75	100.03
10000	5.52	52.84	>90 mins

Table 2: Running times (in seconds) of the three independence tests for different sample sizes.

4.3 Power study

In this section, we compare the independence tests $\mathcal{I}_n$, $\mathcal{I}_n^{\mathrm{DC}}$ and $\mathcal{I}_n^{\mathrm{CvM}}$ in terms of power. We estimated the powers of the respective tests for the models `int`, `sqnorm`, `weight`, `sin`, `max`, `range` and `eval`. We create $B = 500$ samples of medium size $n = 100$ for each type of relationship and $r_i^2 = i/10$, $0 \leq \ i \leq \ 10$. The results can be seen in Figure 4. Similar results for Setup (b) are available in Figure 6 in Appendix A. We can observe that in most settings $\mathcal{I}_n^{\mathrm{DC}}$ has the highest power. In absence of noise $\mathcal{I}_n$ and $\mathcal{I}_n^{\mathrm{DC}}$ have power close to one in most considered settings. It seems that $\mathcal{I}_n^{\mathrm{DC}}$ is less sensitive to noise, in the sense that the power of the test decreases slower with increasing noise level. The power of $\mathcal{I}_n^{\mathrm{CvM}}$ can be rather low if the relationship between X and Y is non-linear, even in the absence of noise. For `sin` and `range` the power is close to the nominal level of the test in the absence of noise. For `sin`, the test based on $\mathcal{I}_n$ performs remarkably well, even in comparison to the one based on $\mathcal{I}_n^{\mathrm{DC}}$. *Moreover, it is the only test that achieves (near) perfect power in the absence of noise in all considered settings.* Comparing Setups (a) and (b), there does not seem to be a notable difference. At the chosen sample size, the tests are in tendency only slightly more powerful when the eigenvalues decay rapidly.

4.4 Real data illustration

This data set was previously studied in Ofner (2021) in the context of functional quantile regression. We apply the procedures discussed to COVID-19 vaccination data from 2117 Austrian municipalities. Specifically, we consider for each municipality:

- a curve X representing the population's age distribution on 01.01.2021. This data is provided by Statistik Austria (2021).
- the proportion Y of the population who had received at least two COVID-19 vaccinations by 13.10.2021. This data is provided by the Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (BMSGPK) (2021).

Figure 5 shows the entire sample of age curves. From visual inspection, there is no obvious connection between age structure and vaccination rate. However, the values of $\widehat{T}_n = 0.31$ and $\widehat{R}_n = 0.43$ indicate a clear relationship. As the absolute values of these coefficients cannot be directly compared, we perform the corresponding independence tests, including also the $\mathcal{I}_n^{\mathrm{CvM}}$ test. The p-values returned by $\mathcal{I}_n$, $\mathcal{I}_n^{\mathrm{DC}}$ and $\mathcal{I}_n^{\mathrm{CvM}}$ are $3 \times 10^{-4}\%, 0.02\%$ and $2 \times 10^{-14}\%$, respectively. For the permutation and the bootstrap test, a number of 5000 repetitions was used. Hence, all three tests indicate a highly significant connection between the age structure and the vaccination rate. An interesting observation is that $\mathcal{I}_n^{\mathrm{CvM}}$ is very sensitive to changes in the dataset: for example, when only the 598 municipalities

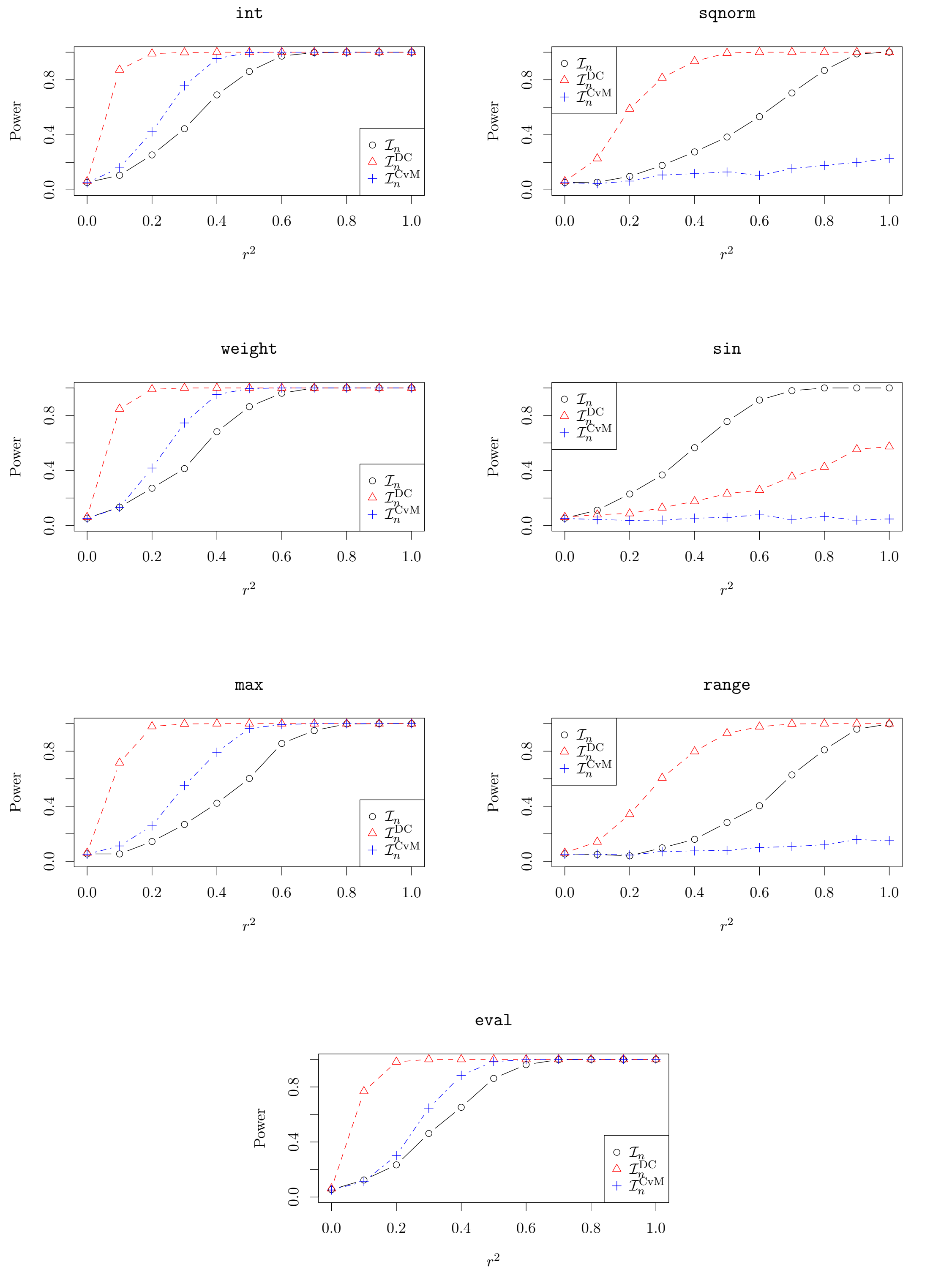


Figure 4: Estimated powers of the three tests for independence at different levels of noise.

with at least 3000 inhabitants are considered, $\mathcal{I}_n$ and $\mathcal{I}_n^{\mathrm{DC}}$ give similar p-values as before, but the p-value of $\mathcal{I}_n^{\mathrm{CvM}}$ rises to values between 5% and 11% (the variation arises from

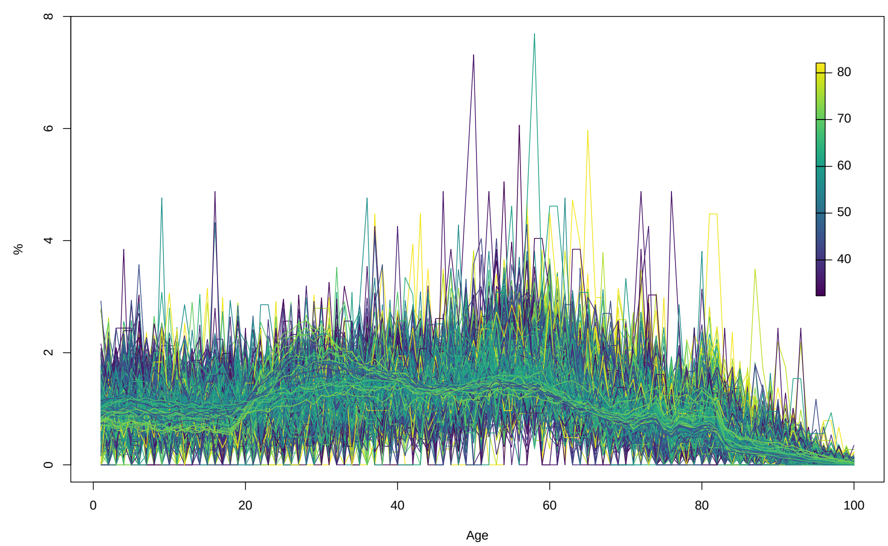


Figure 5: Age curves are colored according to the vaccination rates of the corresponding municipalities.

the test being randomized). Also note that for the full dataset, the computation of $\mathcal{I}_n$ takes 0.36 seconds, the computation of $\mathcal{I}_n^{\mathrm{DC}}$ takes 33.14 seconds, and the computation of $\mathcal{I}_n^{\mathrm{CvM}}$ takes 2:29 minutes.

5 Proofs

5.1 Proof of Theorem 1

We need some preparatory lemmas. Variants of Lemmas 1 and 2 are given by Azadkia and Chatterjee (2021), but we present them here for the sake of completeness.

Lemma 1. *Let X_i be random elements with values in a separable metric space. Let $X_{N(1)}$ be the nearest neighbor of X_1 among $X_2, \ldots, X_n$. Then $X_{N(1)} \xrightarrow{a.s.} X_1$ as $n \to \infty$.*

Proof. Let $\varepsilon > 0$. Due to separability, the space H can be covered by countably many balls of diameter ε. By σ-subadditivity of P some of those balls must have positive probability. Therefore, with probability one, X_1 lies in a ball B which has positive probability. By the triangle inequality,

$$P(d(X_1, X_{N(1)}) > \varepsilon) \leq \; P(X_2, \ldots, X_n \notin B) = (1 - P(X_1 \in B))^{n-1},$$

which tends to zero, since $P(X_1 \in B) > 0$. $\square$

Lemma 2. *Let $\widehat{Q}_n$ be defined as in* (3) *and set $\widetilde{Q}_n := \frac{1}{n}\sum_{i=1}^n \min\left\{F(Y_i), F\left(Y_{N(i)}\right)\right\}$. Then $|\widehat{Q}_n - \widetilde{Q}_n| \xrightarrow{a.s.} 0$ and $|\mathbb{E}\,\widehat{Q}_n - \mathbb{E}\,\widetilde{Q}_n| \to 0$.*

Proof. It is easily seen that $|\widetilde{Q}_n - \widehat{Q}_n| \leq \; \sup_{t\in\mathbb{R}} |F(t) - F_n(t)|$ and hence by the Glivenko-Cantelli theorem $|\widetilde{Q}_n - \widehat{Q}_n| \xrightarrow{a.s.} 0$. Since $\sup_{t\in\mathbb{R}} |F(t) - F_n(t)| \leq \; 1$, dominated convergence yields $\lim_{n\to\infty} \mathbb{E}\,|\widetilde{Q}_n - \widehat{Q}_n| = 0$, which implies the second statement. $\square$

Lemma 3. *It holds that* $P(N(1) = N(2)) = o(1)$.

Proof. Since $|X_1 - X_{N(1)}| \xrightarrow{a.s.} 0$ by Lemma 1, we have

$$\begin{aligned} P(N(1) = N(2)) \leq & P\left(|X_1 - X_{N(1)}| \geq \frac{|X_1 - X_2|}{2} \vee |X_2 - X_{N(2)}| \geq \frac{|X_1 - X_2|}{2}\right) \\ \leq & P\left(|X_1 - X_{N(1)}| \geq \frac{|X_1 - X_2|}{2}\right) + P\left(|X_2 - X_{N(2)}| \geq \frac{|X_1 - X_2|}{2}\right) \\ = & o(1). \end{aligned}$$

□

Lemma 4. *It holds that* $\mathrm{Var}(\widetilde{Q}_n) \to 0$.

Proof. We have

$$\begin{aligned} \mathrm{Var}(\widetilde{Q}_n) = & \frac{1}{n^2} \sum_{i \neq j} \mathrm{Cov}(\min\{F(Y_i), F(Y_{N(i)})\}, \min\{F(Y_j), F(Y_{N(j)})\}) \\ & + \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(\min\{F(Y_i), F(Y_{N(i)})\}) =: \psi_{1,n} + \psi_{2,n}. \end{aligned}$$

Clearly, $\psi_{2,n} \to 0$ for $n \to \infty$. For reasons of symmetry, the summands of $\psi_{1,n}$ are all equal and hence

$$\psi_{1,n} \leq \mathrm{Cov}(F(Y_1) \wedge F(Y_{N(1)}), F(Y_2) \wedge F(Y_{N(2)})).$$

Let $F_X(t) = P(Y \leq t|X)$ and note that we may represent $Y_i = F_{X_i}^{-1}(Z_i)$, with i.i.d. random variables Z_i, which are uniformly distributed on $(0, 1)$ and such that $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $\mathbf{X} = (X_1, \ldots, X_n)$ are independent. Then set $F(Y_i) =: f(X_i, Z_i)$. We want to prove that $F(Y_1) \wedge F(Y_{N(1)})$ and $F(Y_2) \wedge F(Y_{N(2)})$ are asymptotically independent, which then implicates that $\psi_{1,n} \to 0$. Hence, consider the joint distribution function

$$\begin{aligned} & P\left(F(Y_1) \wedge F(Y_{N(1)}) \leq s, F(Y_2) \wedge F(Y_{N(2)}) \leq t\right) \\ & = \mathbb{E}\, P\left(f(X_1, Z_1) \wedge f(X_{N(1)}, Z_{N(1)}) \leq s, f(X_2, Z_2) \wedge f(X_{N(2)}, Z_{N(2)}) \leq t \middle| \mathbf{X}\right). \end{aligned}$$

Since $\mathbf{X}$ and $\mathbf{Z}$ are independent and the nearest neighbor graph is a function of $\mathbf{X}$, the above expectation can be written as (see Durrett (2019), Example 4.1.7.)

$$\int P\big(f(x_1, Z_1) \wedge f(x_{n(1)}, Z_{n(1)}) \leq s, f(x_2, Z_2) \wedge f(x_{n(2)}, Z_{n(2)}) \leq t\big) dP_{\mathbf{X}}(\mathbf{x}), \tag{6}$$

where we integrate over the product space of the X_i's and where $x_{n(i)}$ is the nearest neighbor of x_i within $\mathbf{x} = (x_1, \ldots, x_n)$. We now split the integral over regions

$$R_1 := \{\mathbf{x} \colon |\{1, 2, n(1), n(2)\}| = 4\} \quad \text{and} \quad R_2 := R_1^c.$$

We have

$$P(\mathbf{X} \in R_1^c) \leq P(N(1) = 2) + P(N(2) = 1) + P(N(1) = N(2)).$$

Notice that $P(N(1)=2)=P(N(2)=1)=1/n$ and $P(N(1)=N(2))=o(1)$ by Lemma 3. We conclude, that $P(\mathbf{X}\in R_1^c)\to 0$. Hence

$$\begin{aligned}
&P\left(F(Y_1)\wedge F(Y_{N(1)})\leq\ s, F(Y_2)\wedge F(Y_{N(2)})\leq\ t\right)\\
&=\int_{R_1} P\big(f(x_1,Z_1)\wedge f(x_{n(1)},Z_{n(1)})\leq\ s, f(x_2,Z_2)\wedge f(x_{n(2)},Z_{n(2)})\leq\ t\big)dP_{\mathbf{X}}(\mathbf{x})\\
&\qquad + o(1)\\
&=\int_{R_1} P\big(f(x_1,Z_1)\wedge f(x_{n(1)},Z_{n(1)})\leq\ s\big)P\big(f(x_2,Z_2)\wedge f(x_{n(2)},Z_{n(2)}\big)\leq\ t\big)dP_{\mathbf{X}}(\mathbf{x})\\
&\qquad + o(1)\\
&=\int P\big(f(x_1,Z_1)\wedge f(x_{n(1)},Z_{n(1)})\leq\ s\big)P\big(f(x_2,Z_2)\wedge f(x_{n(2)},Z_{n(2)}\big)\leq\ t\big)dP_{\mathbf{X}}(\mathbf{x})\\
&\qquad + o(1)\\
&=P\left(F(Y_1)\wedge F(Y_{N(1)})\leq\ s\right)P\left(F(Y_2)\wedge F(Y_{N(2)})\leq\ t\right)+o(1).
\end{aligned}$$

□

Proof of Theorem 1. Due to Lemmas 2 and 4 above, the proof follows if we can show that $\mathbb{E}\,\widetilde{Q}_n\to Q$. Noting that the variables $\min\{F(Y_i),F(Y_{N(i)})\}$, $1\leq\ i\leq\ n$, are identically distributed and using our derivation in Section 2 we get

$$\mathbb{E}\,\widetilde{Q}_n=\mathbb{E}\,\min\{F(Y_1),F(Y_{N(1)})\}=\int\mathbb{E}\,G_{X_1}(t)G_{X_{N(1)}}(t)dF(t).$$

By Lemma 1, Assumption 1 (c), and dominated convergence we get the desired result. □

5.2 Proof of Theorem 2

For the remainder of this section, assume X_i and Y_i are independent.

In a first step, we show that the quantity $\widetilde{Q}_n$ is asymptotically normal. To this end, we will use the following result by Rinott (1994).

Lemma 5. *[Rinott (1994), Thm. 2.2] Let $V_1,\ldots,V_n$ be random variables having a dependence graph whose maximal degree is strictly less than k, satisfying*

- $|V_i-\mathbb{E}\,(V_i)|\leq\ B$ *almost surely for* $i=1,\ldots,n$*,*
- $\mathbb{E}\,(\sum_{i=1}^n V_i)=0$ *and*
- $\mathrm{Var}(\sum_{i=1}^n V_i)=1$.

Then

$$\left|P\left(\sum_{i=1}V_i\leq\ z\right)-\Phi(z)\right|\leq\ \sqrt{\frac{1}{2\pi}}kB+16\sqrt{n}k^{3/2}B^2+10nk^2B^3. \tag{7}$$

Lemma 6. *Let $\mathcal{G}_n$ be the nearest neighbor graph generated by vertices $X_1,\ldots,X_n$. Then*

$$\widetilde{W}_n:=\mathrm{Var}(\sqrt{n}\,\widetilde{Q}_n|\mathcal{G}_n)=\frac{1}{n}\left(\frac{W_{n,1}}{45}+\frac{W_{n,2}}{18}\right),$$

where

$$W_{n,1}=n+\sum_{i=1}^n L_{i,n}^2-2f_n\quad\textit{and}\quad W_{n,2}=n+f_n,$$

and where f_n is the number of indices $i\in\{1,\ldots,n\}$, with $N(N(i))=i$. We have $\widetilde{W}_n>\frac{4}{45}$.

Proof. Define $U_i = F(Y_i)$ and notice that the U_i are i.i.d. uniform on $[0,1]$. We have

$$\begin{aligned}\widetilde{W}_n &= \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \min\{U_i, U_{N(i)}\}\middle|\mathcal{G}_n\right)\\ &= \frac{1}{n}\sum_{i,j=1}^n \operatorname{Cov}\left(\min\{U_i, U_{N(i)}\}, \min\{U_j, U_{N(j)}\}|\mathcal{G}_n\right). \qquad (8)\end{aligned}$$

We partition $I_n = \{1,\dots,n\}^2$ into the three $\mathcal{G}_n$-measurable sets

$$I_n^{(k)} := \{(i,j) \in I_n \colon |\{j, N(j)\} \cap \{i, N(i)\}| = k\}, \quad k = 0,1,2.$$

Independence of X and Y implies that distinct variables among $U_i, U_{N(i)}, U_j, U_{N(j)}$ in (8) are i.i.d. Hence, we get

$$\begin{aligned}\widetilde{W}_n &= \frac{|I_n^{(1)}|}{n}\operatorname{Cov}\left(\min\{U_1,U_2\}), \min\{U_1,U_3\}\right) + \frac{|I_n^{(2)}|}{n}\operatorname{Var}\left(\min\{U_1,U_2\}\right)\\ &= \frac{1}{n}\left(\frac{|I_n^{(1)}|}{45} + \frac{|I_n^{(2)}|}{18}\right).\end{aligned}$$

We need to show that $I_n^{(k)} = W_{n,k}$, $k = 1,2$. For (i,j) to be in $I_n^{(2)}$ we have two options: either $j = i$ (n cases) or $j \neq i$ and $N(i) = j, N(j) = i$. The latter is equivalent to $N(N(i)) = i$ (f_n cases) and hence $I_n^{(2)} = W_{n,2}$ follows.

If $(i,j) \in I_n^{(1)}$ then $j \neq i$ and we have two different possibilities. Option (a) is to have $N(i) = j$ and $N(j) \neq i$ or the other way around. Option (b) is that $N(i) = N(j)$. The number of cases in (a) is equal to

$$W_{n,11} = 2 \times \sum_{i=1}^n \sum_{j=1}^n 1\{N(i) = j\}1\{N(j) \neq i\}1\{i \neq j\}.$$

Observe that for given i,

$$\sum_{j=1}^n 1\{N(i) = j\}1\{N(j) \neq i\}1\{i \neq j\} = 1 - 1\{N(N(i)) = i\},$$

and hence $W_{n,11} = 2(n - f_n)$. The number of cases in (b) is equal to

$$\begin{aligned}W_{n,12} &= \sum_{i=1}^n \sum_{j=1}^n 1\{N(i) = N(j)\}1\{i \neq j\}\\ &= \sum_{i=1}^n \sum_{j=1}^n (1\{N(i) = N(j)\} - 1) = \sum_{i=1}^n |\{j \colon N(j) = N(i)\}| - n = \sum_{i=1}^n L_{i,n}^2 - n.\end{aligned}$$

As for the lower bound for $\widetilde{W}_n$ note first that $-\frac{2}{45}f_n + \frac{1}{18}f_n$ is increasing in f_n. Hence, the lowest thinkable contribution coming from f_n is when $f_n = 0$. In addition, note that we have $\sum_{i=1}^n L_{i,n}^2 \geq n$. Hence, $\widetilde{W}_n \geq \frac{2}{45} + \frac{1}{18} = \frac{1}{10}$. $\square$

In the sequel, we have the following data generating mechanism in mind:

1. We sample the X_i's and generate $\mathcal{G}_n$.

2. Independent of $\mathcal{G}_n$, we generate a random sample $\widetilde{Y}_i$ with distribution function F.

3. We choose a random permutation π_n of $(1, \dots, n)$ which is independent of everything and set $(Y_1, \dots, Y_n) = \pi_n(\widetilde{Y}_{(1)}, \dots, \widetilde{Y}_{(n)})$, where $\widetilde{Y}_{(1)} \leq \cdots \leq \widetilde{Y}_{(n)}$ denotes the ordered sample.

Below, we shall condition on $\mathcal{G}_n = g$ and sometimes express dependence on g for different variables explicitly, e.g. by writing $\widetilde{W}_n(g)$ or $\widetilde{Q}_n(g)$. We let $\deg(g)$ be the maximal degree of g.

Proposition 1. *Let $\mathcal{K}_n$, $n \geq 1$, be a collection of nearest neighbor graphs g with n vertices, such that $\sup_{g \in \mathcal{K}_n} \deg(g) = o(n^{1/4})$. Then there is a null-sequence ψ_n such that*

$$\sup_{g \in \mathcal{K}_n} \sup_{z \in \mathbb{R}} \left| P\left(\sqrt{n} \frac{\widetilde{Q}_n - \frac{1}{3}}{\sqrt{\widetilde{W}_n}} \leq z \middle| \mathcal{G}_n = g \right) - \Phi(z) \right| \leq \psi_n.$$

Proof. We may write

$$\sqrt{\frac{n}{\widetilde{W}_n}} \left(\widetilde{Q}_n - \frac{1}{3} \right) = \sum_{i=1}^{n} V_i,$$

with $V_i = \frac{1}{\sqrt{n\widetilde{W}_n}} \left(\min\{F(Y_i), F(Y_{N(i)})\} - \frac{1}{3} \right)$. With this definition, we infer from Lemma 6 that $|V_i| \leq \frac{2}{3} \frac{1}{\sqrt{n\widetilde{W}_n}} < \sqrt{5/n}$ and $\mathrm{Var}(\sum_{i=1}^{n} V_i | \mathcal{G}_n = g) = 1$, while $E(V_i|\mathcal{G}_n = g) = 0$. Hence, the variables satisfy the conditions of Lemma 5 with $B = \sqrt{5/n}$. The maximum degree of the dependence graph of the Y_i is bounded by $2L_n$. To see this, fix i. There is an edge between Y_i and Y_j with $j \neq i$ if either $j = N(i)$ (one case) or if $N(j) = i$ ($\leq L_n$ cases) or if $N(j) = N(i)$ (at most $L_n - 1$ cases). Conditional on $\mathcal{G}_n = g$ with $g \in \mathcal{K}_n$ we have $k = 2L_n(g) = o(n^{1/4})$. □

We now want to replace the theoretically simpler but practically infeasible variable $\widetilde{Q}_n$ by the statistic $\widehat{Q}_n$. Deb et al. (2020) and Shi et al. (2024) have considered corresponding centering terms $\widehat{C}_n := \frac{1}{n(n-1)} \sum_{i \neq j} F_n(Y_i) \wedge F_n(Y_j)$ and $\widetilde{C}_n := \frac{1}{n(n-1)} \sum_{i \neq j} F(Y_i) \wedge F(Y_j)$, respectively, in their derivations. Note that both quantities are permutation invariant with respect to the Y_i, and thus it is easily seen that $\widehat{C}_n = \frac{n+1}{3n} = \frac{1}{3} + O(n^{-1})$. It follows that $\sqrt{n}(\widehat{Q}_n - \widehat{C}_n)$ is asymptotically equivalent to $\sqrt{n}(\widehat{Q}_n - \frac{1}{3})$. Using a Hàjek representation theorem, the arguments by Deb et al. (2020) show that $\sqrt{n}(\widehat{Q}_n - \widehat{C}_n - (\widetilde{Q}_n - \widetilde{C}_n)) \xrightarrow{\mathcal{P}} 0$ and hence the limiting law of $\sqrt{n}(\widehat{Q}_n - 1/3)$ can be derived from $\sqrt{n}(\widetilde{Q}_n - \widetilde{C}_n)$. Unfortunately, $\widetilde{C}_n$ cannot be replaced by $1/3$ in the asymptotic expansion, and therefore we cannot replace $\widetilde{Q}_n$ by $\widehat{Q}_n$ in Proposition 1. Instead of directly deriving the limiting law of $\sqrt{n}(\widetilde{Q}_n - \widetilde{C}_n)$ (and hence of our statistic $\sqrt{n}(\widehat{Q}_n - 1/3)$) as in Deb et al. (2020), we rather show that the difference $\widehat{Q}_n - \widetilde{Q}_n$ is also asymptotically normally distributed and independent of $\widehat{Q}_n$. With this and with Proposition 1, we will subsequently be able to obtain the limiting law of $\sqrt{n}(\widehat{Q}_n - 1/3)$. The advantages of this detour are discussed after Theorem 2.

Proposition 2. *Under the same setting as in Proposition 1, there is a sequence* $(A_n)_{n\in\mathbb{N}}$ *of* $N(0,4/45)$ *distributed random variables, such that for every* $\varepsilon > 0$ *there exists a null-sequence* $(\eta_n(\varepsilon))$ *such that,*

$$\sup_{g\in\mathcal{K}_n} P\left(|\sqrt{n}(\widehat{Q}_n - \widetilde{Q}_n) - A_n| > \varepsilon | \mathcal{G}_n = g\right) \leq \ \eta_n(\varepsilon).$$

Proof. As presented, for example, by van der Vaart (1998), we can define the variables Y_i on a common probability space $(\Omega, \mathcal{A}, P)$ on which there exists a sequence $(B_n)_{n\in\mathbb{N}}$ of Brownian bridges such that

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\log^2 n} \sup_{t\in\mathbb{R}} |\sqrt{n}(F_n(t) - F(t)) - B_n(F(t))| < \infty \quad \text{a.s.} \tag{9}$$

Now define $A_n := \int_0^1 (B_n \cdot h)(t)dt$, with $h(t) = (2-2t)1\{t \in [0,1]\}$ and for some positive integer sequence $m_n \to \infty$, $m_n = O(n^{3/4})$, and $j \in \{1,\ldots,m_n\}$ define the interval $I_j^{m_n} = ((j-1)/m_n, j/m_n]$. Further, we set $Z_i = \min\{F(Y_i), F(Y_{N(i)})\}$ and note that the Z_i are identically distributed with density function h. We then decompose

$$\sqrt{n}(\widehat{Q}_n - \widetilde{Q}_n) - A_n = D_n^1 + D_n^2 + D_n^3 + D_n^4 + D_n^5,$$

where

$$\begin{aligned}
D_n^1 &= \sqrt{n}(\widehat{Q}_n - \widetilde{Q}_n) - \frac{1}{n}\sum_{i=1}^n B_n(Z_i),\\
D_n^2 &= \frac{1}{n}\sum_{i=1}^n B_n(Z_i) - \sum_{j=1}^{m_n} B_n(j/m_n)\cdot \frac{|\{Z_1,\ldots,Z_n\}\cap I_j^{m_n}|}{n},\\
D_n^3 &= \sum_{j=1}^{m_n} B_n(j/m_n)\cdot \frac{|\{Z_1,\ldots,Z_n\}\cap I_j^{m_n}|}{n} - \sum_{j=1}^{m_n} B_n(j/m_n)\int_{I_j^{m_n}} h(t)dt,\\
D_n^4 &= \sum_{j=1}^{m_n} B_n(j/m_n)\int_{I_j^{m_n}} h(t)dt - \sum_{j=1}^{m_n} \frac{1}{m_n} B_n(j/m_n)h(j/m_n),\\
D_n^5 &= \sum_{j=1}^{m_n} \frac{1}{m_n} B_n(j/m_n)h(j/m_n) - \int_0^1 (B_n\cdot h)(t)dt.
\end{aligned}$$

We will find bounds for the quantities $|D_n^i|$ and use the triangular inequality. Denote by $\|\cdot\|_\infty$ the sup-norm on $L^2([0,1])$. Then (9) implies that

$$P\left(|D_n^1| > \frac{\log^3 n}{\sqrt{n}}\right) \ \leq \ P\left(\|\sqrt{n}(F_n - F) - B_n\circ F\|_\infty > \frac{\log^3 n}{\sqrt{n}}\right) \to 0.$$

Let $C > \sqrt{2}$ and $\alpha < 1/2$. Theorem 1.1.1 in Csörgő and Révész (1981), implies that there is a null-sequence (δ_k) such that

$$P\left(\sup_{0\leq\ t\leq\ 1-1/k}\ \sup_{0\leq\ s\leq\ 1/k} |B_n(t+s) - B_n(t)| > Ck^{-\alpha}\right) \leq \delta_k, \quad k \geq\ 1. \tag{10}$$

Since the B_n are identically distributed, this inequality holds uniformly in n. Therefore, with probability $1-\delta_{m_n}$ we get that

$$|D_n^2| \leq \frac{1}{n}\sum_{j=1}^{m_n}\sum_{i=1}^{n}|B_n(Z_i)-B_n(j/m_n)|1\{Z_i \in I_j^{m_n}\} \leq \ Cm_n^{-\alpha}.$$

We note that the bounds derived for $|D_n^1|$ and $|D_n^2|$ do not depend on the nearest neighbor graph $\mathcal{G}_n$. Quantities $|D_n^4|$ and $|D_n^5|$ do not involve the nearest neighbor graph at all. This will be only relevant for bounding $|D_n^3|$. Here we have

$$|D_n^3| \leq \sup_{t\in[0,1]}|B_n(t)| \times \sum_{j=1}^{m_n}\left|\frac{|\{Z_1,\dots,Z_n\}\cap I_j^{m_n}|}{n} - \int_{I_j^{m_n}} h(t)dt\right|.$$

We have that $\sup_{t\in[0,1]}|B_n(t)| = O_P(1)$ and thus it remains to show that the sum on the right above converges to zero in probability, uniformly for $g \in \mathcal{K}_n$. We remark that $\mathbb{E}\left(\frac{1}{n}|\{Z_1,\dots,Z_n\}\cap I_j^{m_n}|\right) = \int_{I_j^{m_n}} h(t)dt$. Therefore, by Chebyshev's inequality

$$\begin{aligned}
&P\left(\sum_{j=1}^{m_n}\left|\frac{|\{Z_1,\dots,Z_n\}\cap I_j^{m_n}|}{n} - \int_{I_j^{m_n}} h(t)dt\right| > \delta\right)\\
&\leq \ \frac{1}{\delta}\sum_{j=1}^{m_n}\mathbb{E}\left|\frac{|\{Z_1,\dots,Z_n\}\cap I_j^{m_n}|}{n} - \int_{I_j^{m_n}} h(t)dt\right|\\
&\leq \ \frac{1}{\delta}\sum_{j=1}^{m_n}\mathrm{Var}^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}1\{Z_i\in I_j^{m_n}\}\right)\\
&\leq \ \frac{1}{n\delta}\sum_{j=1}^{m_n}\left(\sum_{i,k=1}^{n}\mathrm{Cov}(1\{Z_i\in I_j^{m_n}\},1\{Z_k\in I_j^{m_n}\})\right)^{1/2}. \qquad (11)
\end{aligned}$$

In the proof of Proposition 1 we showed that conditional on $\mathcal{G}_n = g$, the dependence graph of the variables V_i (which is the same as the dependence graph of the Z_i) has maximal degree $2L_n(g)$. Hence, conditional on $\mathcal{G}_n = g$, we obtain by the Cauchy-Schwarz inequality and the fact the Z_i are identically distributed, that uniformly on $\mathcal{K}_n$

$$\sum_{i,k=1}^{n}\left|\mathrm{Cov}(1\{Z_i\in I_j^{m_n}\},1\{Z_k\in I_j^{m_n}\}\right| \leq \ n\times\deg(g)P(Z_1\in I_j^{m_n}) = o\left(n^{5/4}/m_n\right).$$

Inserting into (11) shows that $|D_n^3|$ converges to zero in probability, uniformly on $\mathcal{K}_n$.

By similar arguments we obtain $|D_n^4| \leq \ 2\sup_{t\in[0,1]}|B_n(t)|/m_n = O_P(m_n^{-1})$ and

$$|D_n^5| \leq \ 2\sup_{t\in[0,1]}|B_n(t)|/m_n + 2\sum_{j=1}^{m_n}\int_{I_j^{m_n}}|B_n(j/m_n)-B_n(t)|dt = O_P(m_n^{-1/2}).$$

In the last step, we used the basic fact that $\int_{I_j^{m_n}} E|B_n(j/m_n)-B_n(t)|dt = \sqrt{\frac{2}{\pi}}\frac{2}{3}m_n^{-3/2}$.

It remains to show that A_n follows an $N(0,4/45)$ distribution. Since B_n is a zero mean Gaussian process, it is clear that A_n is normally distributed with mean zero. To calculate the variance, recall that $h(t) = 2 - 2t$ on $[0,1]$ and consider

$$\begin{aligned}\operatorname{Var}\left(\int_0^1 B_n(t)h(t)dt\right) &= \mathbb{E}\left(\left(\int_0^1 B_n(t)h(t)dt\right)^2\right)\\ &= \mathbb{E}\left(\int_0^1 B_n(t)h(t)dt \int_0^1 B_n(s)h(s)ds\right)\\ &= \int_0^1\int_0^1 \mathbb{E}\left(B_n(t)B_n(s)\right)h(t)h(s)dsdt\\ &= \int_0^1\int_0^1 (\min\{s,t\} - st)h(s)h(t)dsdt = \frac{4}{45}.\end{aligned}$$

□

Lemma 7. *Define $W_{n,1}$ and $W_{n,2}$ as in Lemma 6. The conditional variance of $\widehat{Q}_n$ given $\mathcal{G}_n$ is*

$$W_n := \operatorname{Var}(\sqrt{n}\widehat{Q}_n|\mathcal{G}_n) = \frac{W_{n,1}v_1 + W_{n,2}v_2 + (n^2 - W_{n,1} - W_{n,2})v_0}{n}, \tag{12}$$

where

$$v_0 = -\frac{4(n+1)}{45n^2}, \quad v_1 = \frac{4n^4 - 25n^3 + 30n^2 + 25n^1 - 34}{180n^2(n-1)(n-2)}, \quad v_2 = \frac{n^2 - n - 2}{18n^2}.$$

Proof. As in the proof of Lemma 6 we partition I_n into $I_n^{(k)}$, $k = 0,1,2$. The covariance of pairs U_i', U_j', where $U_i' = \min\{F_n(Y_i), F_n(Y_{N(i)})\}$, for $(i,j) \in I_n^{(0)}$ is given by

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{\substack{i,j,k,l=1\\ \text{all distinct}}}^{n} \left(\min\left\{\frac{i}{n},\frac{k}{n}\right\} - \frac{n+1}{3n}\right)\left(\min\left\{\frac{j}{n},\frac{l}{n}\right\} - \frac{n+1}{3n}\right) = v_0,$$

for $(i,j) \in I_n^{(1)}$ by

$$\frac{1}{n(n-1)(n-2)} \sum_{\substack{i,j,k=1\\ \text{all distinct}}}^{n} \left(\min\left\{\frac{i}{n},\frac{k}{n}\right\} - \frac{n+1}{3n}\right)\left(\min\left\{\frac{j}{n},\frac{k}{n}\right\} - \frac{n+1}{3n}\right) = v_1,$$

and for $(i,j) \in I_n^{(2)}$ by

$$\frac{1}{n(n-1)} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} \left(\min\left\{\frac{i}{n},\frac{j}{n}\right\} - \frac{n+1}{3n}\right)^2 = v_2,$$

where we used that $\mathbb{E}\left(\min\{F_n(Y_i), F_n(Y_{N(i)})\}\right) = (n+1)/3n$. □

Note that under Assumption 2, $\frac{W_n}{\widetilde{W}_n - \frac{4}{45}} \to 1$ for $n \to \infty$, since $\frac{v_1}{n} \sim \frac{1}{45n}$, $\frac{v_2}{n} \sim \frac{1}{18n}$ and $n^2 - W_{n,1} - W_{n,2} \sim n^2$, if $n \to \infty$.

Proposition 3. *Suppose that Assumption 2 holds. Then, for any $z \in \mathbb{R}$,*

$$\left| P\left(\sqrt{n}\frac{\widehat{Q}_n - \frac{1}{3}}{W_n} \leq z\right) - \Phi(z)\right| \to 0,$$

as $n \to \infty$.

Proof. The construction of the Brownian bridges in (9) only involves F_n which is a function of the order statistics $Y_{(i)}$ and thus A_n can be assumed to be independent of $\mathcal{G}_n$ and π_n. In contrast, we may represent $\widehat{Q}_n = \frac{1}{n^2}\sum_{i=1}^n \pi_n(i) \wedge \pi_n(N(i))$ and hence it is measurable with respect to $\mathcal{G}_n$ and π_n. It follows that $\widehat{Q}_n$ and A_n are independent.

Due to Proposition 2 and the fact that $\widetilde{W}_n \geq \frac{1}{10}$

$$\sup_{g\in\mathcal{K}_n} P\left(\frac{1}{\sqrt{\widetilde{W}_n}}\left|\sqrt{n}(\widetilde{Q}_n - \frac{1}{3}) - \sqrt{n}\left(\widehat{Q}_n - \frac{1}{3}\right) - A_n\right| > \varepsilon \middle| \mathcal{G}_n = g\right) \leq \eta_n(\varepsilon/\sqrt{10}).$$

Thus, we have for any $z \in \mathbb{R}$

$$\begin{aligned}
&\sup_{g\in\mathcal{K}_n} P\left(\sqrt{\frac{1}{\widetilde{W}_n}}\left(\sqrt{n}\left(\widehat{Q}_n - \frac{1}{3}\right) + A_n\right) \leq z \middle| \mathcal{G}_n = g\right) \\
&\leq \sup_{g\in\mathcal{K}_n} P\left(\sqrt{\frac{n}{\widetilde{W}_n}}\left(\widetilde{Q}_n - \frac{1}{3}\right) \leq z + \varepsilon \middle| \mathcal{G}_n = g\right) + \eta_n(\varepsilon/\sqrt{10}) \\
&\leq \Phi(z+\varepsilon) + \psi_n + \eta_n(\varepsilon/\sqrt{10}) \leq \Phi(z) + \frac{\varepsilon}{\sqrt{2\pi}} + \psi_n + \eta_n(\varepsilon/\sqrt{10}).
\end{aligned}$$

An analogous lower bound can be established, yielding that

$$\sup_{g\in\mathcal{K}_n}\sup_{z\in\mathbb{R}}\left|P\left(\sqrt{\frac{1}{\widetilde{W}_n}}\left(\sqrt{n}\left(\widehat{Q}_n - \frac{1}{3}\right) + A_n\right) \leq z \middle| \mathcal{G}_n = g\right) - \Phi(z)\right| = o(1).$$

Note that conditionally on $\mathcal{G}_n = g$, $\widetilde{W}_n$ is deterministic, $S_{2,n} := \sqrt{\frac{1}{\widetilde{W}_n}}A_n \sim N\left(0, \frac{1}{\widetilde{W}_n}\frac{4}{45}\right)$ and $S_{2,n}$ is independent of $S_{1,n} := \sqrt{\frac{n}{\widetilde{W}_n}}(\widehat{Q}_n - \frac{1}{3})$. For the rest of the proof, we use conditional characteristic functions. For some $A \in \mathcal{A}$ we denote $\varphi_{X|A}(t) = E[e^{\mathbf{i}tX}|A]$, $t \in \mathbb{R}$ and $\mathbf{i} = \sqrt{-1}$.

It follows from our previous derivations that for any $t \in \mathbb{R}$

$$\sup_{g\in\mathcal{K}_n}\left|\varphi_{S_{1,n}|\mathcal{G}_n=g}(t)e^{-\frac{1}{\widetilde{W}_n(g)}\frac{2}{45}t^2} - e^{-\frac{1}{2}t^2}\right| \to 0. \tag{13}$$

In fact, it can be shown that the convergence in (13) holds uniformly on any compact interval (see e.g. Theorem 2.66 in Jeffreys (2003)). The lower bound $\widetilde{W}_n(g) \geq 1/10$ implicates that uniformly for t in any compact interval

$$\sup_{g\in\mathcal{K}_n}\left|\varphi_{S_{1,n}|\mathcal{G}_n=g}(t) - e^{-\frac{1}{2}\left(1-\frac{1}{\widetilde{W}_n(g)}\frac{4}{45}\right)t^2}\right| \to 0.$$

Define $R_n := \left(1 - \frac{1}{\widetilde{W}_n}\frac{4}{45}\right)$ and observe that $1/\sqrt{R_n} \in [1,3]$. Hence, if t is fixed, then $t/\sqrt{R_n}$ remains in a compact interval. Thus,

$$\sup_{g\in\mathcal{K}_n}\left|\varphi_{S_{1,n}|\mathcal{G}_n=g}(t/\sqrt{R_n(g)}) - e^{-\frac{1}{2}t^2}\right| = \sup_{g\in\mathcal{K}_n}\left|\varphi_{S_{1,n}/\sqrt{R_n(g)}|\mathcal{G}_n=g}(t) - e^{-\frac{1}{2}t^2}\right| \to 0.$$

Now note that $S_{1,n}/\sqrt{R_n} = \sqrt{\frac{n}{\widetilde{W}_n - 4/45}}\left(\widehat{Q}_n - \frac{1}{3}\right)$. We just proved that for any $z \in \mathbb{R}$

$$\left|P\left(\sqrt{\frac{n}{\widetilde{W}_n - 4/45}}\left(\widehat{Q}_n - \frac{1}{3}\right) \leq z \middle| \mathcal{G}_n \in \mathcal{K}_n\right) - \Phi(z)\right| \to 0.$$

The proof follows by observing that under Assumption 2 we have that $P(\mathcal{G}_n \in \mathcal{K}_n) \to 1$, and that $W_n \sim \widetilde{W}_n - \frac{4}{45}$, $n \to \infty$. $\square$

5.3 Proof of Theorem 3

In this section we construct an example of a functional random variable X as in Theorem 3. Set $\sigma^2 = \sum_{k=1}^{\infty} \lambda_k$ and $p_k = \lambda_k/\sigma^2$. Let A be a continuous random variable with mean 0 and variance σ^2 and K a discrete random variable independent of A with $P(K = k) = p_k$.

Lemma 8. *Set $X = A \times e_K$. Then X has mean 0 and covariance Σ.*

Proof. Obviously we have $EX = 0$. Then

$$\mathbb{E}\, X\langle v, X\rangle = \sigma^2 \sum_{k\geq 1} P(K = k) e_k \langle v, e_k\rangle,$$

which, by the spectral theorem for compact operators, equals to $\Sigma(v)$. $\square$

Now consider a random sample $X_1, \ldots, X_n$ with $X_i = A_i \times e_{K_i} \sim X$. It holds that

$$\|X_i - X_j\|^2 = A_i^2 + A_j^2 - 2A_iA_j 1\{K_i = K_j\}. \tag{14}$$

Lemma 9. *For $k \geq 1$, define $\mathcal{M}_{k,n} = \{i \leq n \colon K_i = k\}$, $f_{k,n} = |\mathcal{M}_{k,n}|$ and further $G_n = |\{k : f_{k,n} = 1\}|$. Then $L_n \geq G_n - 1$.*

Proof. Suppose $f_{k,n} = 1$. Then $\mathcal{M}_{k,n}$ contains a single element, say i_k. This means that X_{i_k} is the only curve in the sample with a shape that is proportional to the function e_k. By (14), for any $j \in \{1, \ldots, n\}$ with $j \neq i_k$ we have $\|X_{i_k} - X_j\|^2 = A_{i_k}^2 + A_j^2$. This implies that the nearest neighbor of each X_{i_k} is X_{i^*} with $i^* = \operatorname{argmin}_{j\neq i_k} A_j^2$. If $f_{i^*,n} = 1$, then $L_{i^*,n} \geq G_n - 1$, if $f_{i^*,n} > 1$, then even $L_{i^*,n} \geq G_n$ holds. $\square$

Proof of Theorem 3. Without loss of generality, assume that $\sigma^2 = 1$.

(i) For any sequence $x_n \to \infty$, it holds that

$$\theta_n := P(K \geq x_n) = \sum_{k\geq x_n} \lambda_k \quad \text{and} \quad P(K = k) \leq \lambda_{x_n} \quad \text{for all } k \geq x_n.$$

Having drawn a random sample $(A_1, K_1), \ldots, (A_n, K_n)$, let $\mathcal{R}_n = \{1 \leq i \leq n \colon K_i \geq x_n\}|$. It holds that $|\mathcal{R}_n|$ has a binomial distribution B_{n,θ_n} and that

$$|\{i : f_{K_i,n} = 1\} \cap \mathcal{R}_n| \leq G_n.$$

The target is to show now (a) that $|\mathcal{R}_n| \to \infty$ and (b) that $|\{i : f_{K_i,n} = 1\} \cap \mathcal{R}_n| = |\mathcal{R}_n|$ with high probability. For (a), we choose a sequence (x_n) which grows slowly enough for $n\theta_n \to \infty$. In this case, we can always find a diverging sequence (ψ_n) such that also $\frac{n\theta_n}{\psi_n} \to \infty$. By elementary properties of the binomial distribution and the Cauchy-Schwarz inequality, we have

$$P\left(B_{n,\theta_n} \leq n\theta_n - \sqrt{\psi_n n\theta_n}\right) = P\left(B_{n,1-\theta_n} - n(1-\theta_n) \geq \sqrt{\psi_n n\theta_n}\right) \leq \psi_n^{-1},$$

and likewise $P\left(B_{n,\theta_n} \geq n\theta_n + \sqrt{\psi_n n\theta_n}\right) \leq \psi_n^{-1}$. By the choice of ψ_n we have

$$n\theta_n/2 \leq n\theta_n - \sqrt{\psi_n n\theta_n} < n\theta_n + \sqrt{\psi_n n\theta_n} \leq 2n\theta_n$$

if n is large enough. It follows that

$$P(|\mathcal{R}_n| \geq \ n\theta_n/2) \to 1 \quad \text{and} \quad P(|\mathcal{R}_n| \leq \ 2n\theta_n) \to 1. \tag{15}$$

From the left relation in (15) we immediately deduce (a). Moreover, for big enough n,

$$\begin{aligned}
&P(|\{i : f_{K_i,n} = 1\} \cap \mathcal{R}_n| = |\mathcal{R}_n|)/P(|\mathcal{R}_n| \leq \ 2n\theta_n) \\
&\quad = P\left(\text{all } K_i \text{ with } i \in \mathcal{R}_n \text{ are distinct} \,\middle|\, |\mathcal{R}_n| \leq \ 2n\theta_n\right) \\
&\quad \geq \prod_{i=1}^{2n\theta_n} (1 - i\lambda_{x_n}/\theta_n) \\
&\quad = \exp\left(\sum_{i=1}^{2n\theta_n} \log(1 - i\lambda_{x_n}/\theta_n)\right) \\
&\quad \geq \ \exp\left(-2\frac{\lambda_{x_n}}{\theta_n}\sum_{i=1}^{2n\theta_n} i\right) \\
&\quad \geq \ \exp\left(-4\frac{\lambda_{x_n}}{\theta_n}\left(n\theta_n + 1\right)^2\right).
\end{aligned}$$

Hence, using the left relation in (15), (b) follows if

$$n^2\lambda_{x_n}\theta_n \to 0. \tag{16}$$

Since by assumption $\lambda_{x_n} = o\left(\theta_n\right)$, it is possible to choose a sequence x_n such that $n\theta_n \to \infty$ and (16) holds. Combining these results with Lemma 9 we infer $P(L_n \geq \ n\theta_n/2) \to 1$, and hence that L_n diverges with a rate at least as fast as $n\theta_n$.

(ii) Set $x_n = n^{\frac{2}{2a-1}}g(n)$, where g is another slowly varying function. Using basic properties of slowly varying functions, we obtain $\theta_n \sim \frac{\ell(x_n)x_n^{-a+1}}{a-1}$ and thus

$$n\theta_n \sim \frac{1}{a-1}\ell(n^{\frac{2}{2a-1}}g(n))g^{1-a}(n)n^{\frac{1}{2a-1}}, \quad n \to \infty,$$

and

$$n^2\lambda_{x_n}\theta_n \sim \frac{1}{a-1}\ell^2(n^{\frac{2}{2a-1}}g(n))g^{1-2a}(n), \quad n \to \infty.$$

where $h(x) := \frac{1}{a-1}\ell(x^{\frac{2}{2a-1}}g(x))g^{1-a}(x)$ is again slowly varying. By suitable choice of g we can ensure that 16 holds. The rest of the proof follows from the arguments in (i).

(iii) Note that, for $K_i = K_j$, it holds that

$$||X_i - X_j||^2 = (A_i - A_j)^2,$$

while for $K_i \neq K_j$, it holds that

$$||X_i - X_j||^2 \geq \ ||X_i||^2 = A_i^2.$$

This implies that the only $i \in \mathcal{M}_{k,n}$, for which $N(i) \notin \mathcal{M}_{k,n}$ is possible, are

$$i = \arg\min\{A_j \colon j \in \mathcal{M}_{k,n}, A_j \geq \ 0\} \quad \text{or} \quad i = \arg\max\{A_j \in \mathcal{M}_{k,n}, A_j \leq \ 0\}.$$

Moreover, there can be at most two indices $j \in M_{K_i,n}$ such that $N(j) = i$, since this implies that A_i is the nearest neighbor of A_j among $\{A_l : l \in \mathcal{M}_{K_i,n}\}$ and the scalar A_i can be the nearest neighbor to at most two other scalars in any set. Combining these observations, we obtain that

$$L_n \leq \ 2 + 2x_n + |\mathcal{R}_n|. \tag{17}$$

Now choosing x_n as in (ii) and observing that $P(|\mathcal{R}_n| \leq \ 2n\theta_n) \to 1$, we use (17) to obtain $L_n = \ o_P(n^{1/4})$.

□

5.4 Proof of Theorem 4

We define X_i^q as the projection of X_i onto the space spanned the first q eigenfunctions, i.e. $X_1^q = \sum_{k=1}^q \langle X_1, e_k\rangle e_k$. In analogy to $L_{1,n}$ define $L_{i,n}^q$ as the corresponding numbers we get with $X_1, \ldots, X_n$ replaced by $X_1^q, \ldots, X_n^q$. It is elementary that

$$P(L_n > k) \leq \ P(L_n^q > k) + P(L_n > L_n^q) \leq \ P(L_n^q > k) + nP(L_{1,n} > L_{1,n}^q). \tag{18}$$

For bounding the first term in (18) we will use the multivariate bound of Kabatjanski and Levenstein (1978) and for the second the following result:

Proposition 4. *Under Assumption 3 we have*

$$P(L_{1,n} > L_{1,n}^q) \leq \mathrm{const} \times n^2 \left(\sum_{j>q} \lambda_j\right)^{1/2},$$

where the constant is independent of n and q.

For the proof of Proposition 4 we need a few preparatory lemmas. In the sequel, we consider independent random variables Z_1, Z_2 and Z having bounded densities $f_1(s)$, $f_2(s)$ and $f(s)$.

Lemma 10. *Let $f_{1+2}(s)$ be the density function of $Z_1 + Z_2$. Then*

$$\sup_s f_{1+2}(s) \leq \ \min\{\sup_s f_1(s), \sup_s f_2(s)\}.$$

Proof. It holds that $f_{1+2}(s) = \int_{-\infty}^{\infty} f_1(s-t) f_2(t) dt$. Now factor out either $\sup_t f_1(s-t)$ or $\sup_t f_2(t)$. □

The next lemma is also elementary.

Lemma 11. *Let $c \in \mathbb{R}$ be a constant. Then $(Z-c)^2$ has density*

$$\frac{1}{2\sqrt{s}}\big(f(c+\sqrt{s}) + f(c-\sqrt{s})\big) I\{s > 0\}.$$

Lemma 12. *Let $c_1, c_2 \in \mathbb{R}$ be constants. The density of $(Z_1 - c_1)^2 + (Z_2 - c_2)^2$ is bounded from above by $\pi \times \sup_s f_1(s) \times \sup_s f_2(s)$.*

Proof. Let g_1 and g_2 be the densities of $(Z_1 - c_1)^2$ and $(Z_2 - c_2)^2$, respectively. Then the density of $(Z_1 - c_1)^2 + (Z_2 - c_2)^2$ is given as

$$\begin{aligned}
g(s) &= \int_{-\infty}^{\infty} g_1(s-t) g_2(t) dt \\
&= \int_0^s \frac{1}{4\sqrt{s-t}\sqrt{t}} \big(f_1(c_1 + \sqrt{s-t}) + f_1(c_1 - \sqrt{s-t})\big)\big(f_2(c_2 + \sqrt{t}) + f_2(c_2 - \sqrt{t})\big) dt \\
&\leq \ \int_0^s \frac{1}{\sqrt{s-t}\sqrt{t}} dt \times \sup_s f_1(s) \sup_s f_2(s) = \pi \times \sup_s f_1(s) \times \sup_s f_2(s).
\end{aligned}$$

□

Let $X^q_{N(1|q)}$ and $X^q_{M(1|q)}$ be the nearest and second-nearest neighbor of X^q_1, respectively, among the observations $X^q_2, \ldots, X^q_n$.

Lemma 13. *Under Assumption 3 we have*

$$P\left(\|X_1^q - X^q_{M(1|q)}\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2 < \epsilon\right) \leq \text{ const} \times n^2\epsilon,$$

where the constant is independent of n, q and ϵ.

Proof. Let us first obtain a bound conditional on $X_1^q = x$. Conditional on this event, we have that $\|X_1^q - X_k^q\|^2 \overset{iid}{\sim} \|x - X_k^q\|^2$, $k = 2, \ldots, n$. Letting $V_k = V_k(x,q) = \|x - X_k^q\|^2$ and $V_{(k)}$ being the corresponding order statistics with $V_{(1)} \leq \cdots \leq V_{(n-1)}$, we have

$$P\left(\|X_1^q - X^q_{M(1|q)}\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2 < \epsilon \big| X_1^q = x\right) = P(V_{(2)} - V_{(1)} < \epsilon).$$

We will now derive the density of $V_{(2)} - V_{(1)}$ and show that it is bounded by some constant depending only on n, f_1 and f_2. To this end, note that Assumption 3 assures that $V_1(x,q)$ has a density function. This is because $V_1(x,q) = \sum_{k=1}^q (c_k - Z_{1k})^2$, where $c_k = \langle x, e_k\rangle$. Let us denote this density by $g_{x,q}(s)$ and the corresponding distribution function by $G_{x,q}(s)$. Then by Pyke (1965) the density of the spacing $D := V_{(2)} - V_{(1)}$ between the two smallest variables is given as

$$\begin{aligned} f_D(s) &= (n-1)(n-2)\int_0^\infty (1 - G_{x,q}(t+s))^{n-3} g_{x,q}(t) g_{x,q}(t+s)dt \\ &\leq\ n^2 \sup_{t\geq\ 0} g_{x,q}(t) \leq\ n^2 \sup_{t\geq\ 0} g_{x,2}(t) \leq\ n^2 \pi \sup_t f_1(t) \sup_t f_2(t). \end{aligned}$$

The latter two inequalities follow from Lemma 10 and Lemma 12. It follows that

$$P\left(\|X_1^q - X^q_{M(1|q)}\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2 < \epsilon \big| X_1^q = x\right) \leq\ \text{const} \times n^2\epsilon,$$

where the constant is independent of n, q, ϵ and x. □

Proof of Proposition 4. We observe that

$$\{L_{1,n} > L^q_{1,n}\} \subset \{\exists i \in \{2, \ldots, n\} \colon \|X_i - X_{N(i|q)}\|^2 > \|X_1 - X_i\|^2\}.$$

Thus, by reasons of symmetry, we have

$$P(L_{1,n} > L^q_{1,n}) \leq\ nP(\|X_1 - X_{N(1|q)}\|^2 > \|X_2 - X_1\|^2).$$

Denote by Π^q the projection onto the space spanned by $\{e_{q+1}, e_{q+2}, \ldots\}$. Then by Pythagoras' theorem we have $\|X_1 - X_{N(1|q)}\|^2 = \|X_1^q - X^q_{N(1|q)}\|^2 + \|\Pi^q(X_1 - X_{N(1|q)})\|^2$. Thus for any $\epsilon > 0$ it holds that

$$\begin{aligned} &P(\|X_1 - X_{N(1|q)}\|^2 > \|X_1 - X_2\|^2) \\ &= P(\|\Pi^q(X_1 - X_{N(1|q)})\|^2 > \|X_1 - X_2\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2) \\ &\leq\ P(\|\Pi^q(X_1 - X_{N(1|q)})\|^2 > \epsilon) + P(\|X_1 - X_2\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2 \leq\ \epsilon) \\ &\leq\ \epsilon^{-1} \times E\|\Pi^q(X_1 - X_{N(1|q)})\|^2 + P(\|X_1^q - X_2^q\|^2 - \|X_1^q - X^q_{N(1|q)}\|^2 \leq\ \epsilon) \\ &\leq\ \epsilon^{-1} \times \sum_{k>q} E\langle X_1 - X_{N(1|q)}, e_k\rangle^2 + \text{const} \times n^2\epsilon. \end{aligned}$$

We have used the Markov inequality and Lemma 13. We remark that independence of the scores $\langle X_1, e_k\rangle$, $k \geq 1$, and the fact that $N(1|q)$ is measurable with respect to the σ-algebra generated by $\{\langle X_i, e_k\rangle : 1 \leq i \leq n, 1 \leq k \leq q\}$ imply that $\langle X_1, e_k\rangle$ and $\langle X_{N(1|q)}, e_k\rangle$ are independent and have the same distribution if $k > q$. Thus,

$$\sum_{k>q} E\langle X_1 - X_{N(1|q)}, e_k\rangle^2 = 2\sum_{k>q} E\langle X_1, e_k\rangle^2 = 2\sum_{k>q}\lambda_k.$$

We have shown that $P(L_{1,n} > L_{1,n}^q) \leq \text{const} \times n\left(\epsilon^{-1} \times \sum_{k>q}\lambda_k + n^2\epsilon\right)$. The proof is concluded by setting $\epsilon = n^{-1} \times \left(\sum_{k>q}\lambda_k\right)^{1/2}$. $\square$

Proposition 5. *Let Assumption 3 hold and assume that we can choose $q = q_n$ such that*

$$\sum_{j>q}\lambda_j = o(n^{-6}), \tag{19}$$

$$P(L_n^q > n^{1/4}) = o(1). \tag{20}$$

Then Assumption 2 is fulfilled.

Proof. The result is immediate from (18) and Proposition 4. $\square$

Proof of Theorem 4. It holds that $P(L_n^q > n^{1/4}) = 0$ for large enough n and $q_n < \frac{1}{4}\log_\gamma n$ with $\gamma > 2^{0.401}$, as shown by Kabatjanski and Levenstein (1978). This condition is met for our choice of q_n. Finally, condition (19) follows directly from our assumptions. $\square$

Remark 6. *We note that a much slower rate of decay for λ_j can be obtained if we could allow a bigger value for q in (20). While to the best of our knowledge very little is known about distributional properties of L_n^q (even for fixed q), it seems certain that the exponential growth obtained from Kabatjanski and Levenstein (1978) is way too conservative when applied to a random sample. This is confirmed in extensive simulations, where we found L_n to grow at a relatively slow rate, even in cases of very slowly decaying eigenvalues, as, for example, for the Brownian motion. Based on the fact that nearest neighbors constitute a "local" property, we conjecture that for fixed q and k, the probability $P(L_n^q > k)$ does not depend on the distribution of the score vectors $\Pi^q X$ as long as they have continuous densities.*

Acknowledgments

The authors would like to thank Maximilian Ofner for helpful discussions and the compilation of the dataset for the real-world application example. We would also like to thank Norbert Henze and Mathew Penrose for providing some insightful feedback.

Funding

This research was funded by the Austrian Science Fund (FWF) [P 35520]. For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

Arnab Auddy, Nabarun Deb, and Sagnik Nandy. Exact detection thresholds and minimax optimality of Chatterjee's correlation coefficient. *Bernoulli*, 30(2):1640–1668, 2024.

Mona Azadkia and Sourav Chatterjee. A simple measure of conditional dependence. *The Annals of Statistics*, 49(6):3070–3102, 2021.

Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When Is "Nearest Neighbor" Meaningful? *ICDT 1999. LNCS*, 1540:217–235, 1997.

Peter Bickel. Measures of independence and functional dependence. *arXiv preprint arXiv:2206.13663*, 2022.

Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (BMSGPK). Covid-19 Schutzimpfungen – Impfungen in Gemeinden, 2021. URL `https://www.data.gv.at`. Accessed: 23.10.2021.

Sky Cao and Peter Bickel. Correlations with tailored extremal properties. *arXiv preprint arXiv:2008.10177*, 2020.

Sourav Chatterjee. A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610, 2008.

Sourav Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):1–21, 2020.

Sourav Chatterjee. A survey of some recent developments in measures of association. *arXiv preprint arXiv:2211.04702*, 2023.

Miklós Csörgő and Pál Révész, editors. *Strong Approximations in Probability and Statistics*. Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press, New York, 1981.

Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Measuring association on topological spaces using kernels and geometric graphs. *arXiv preprint arXiv:2010.01768*, 2020.

Herold Dehling, Muneya Matsui, Thomas Mikosch, Gennady Samorodnitsky, and Laleh Tafakori. Distance covariance for discretized stochastic processes. *Bernoulli*, 26(4):2758 – 2789, 2020.

Holger Dette, Karl Siburg, and Pavel Stoimenov. A copula-based non-parametric measure of regression dependence. *Scandinavian Journal of Statistics*, 40:21–41, 2013.

Robert Durrant and Ata Kabán. When is 'nearest neighbour' meaningful: A converse theorem and implications. *Journal of Complexity*, 25(4):385–397, 2009.

Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 5th edition, 2019.

William Feller. The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22(3):427–432, 1951.

Eduardo García-Portugués, Wenceslao González-Manteiga, and Manuel Febrero-Bande. A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, 23(3):761–778, 2014.

N. Henze. On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction. *Advances in Applied Probability*, 19:873–895, 1987.

Zhen Huang, Nabarun Deb, and Bodhisattva Sen. Kernel partial correlation coefficient — a measure of conditional dependence. *Journal of Machine Learning Research*, 23(216): 1–58, 2022.

Harold Jeffreys. *The Theory of Probability*. Oxford University Press, Oxford, 3rd edition, 2003.

Grigori Kabatjanski and Vladimir Levenstein. Bounds for packings on the sphere and in space. *Problemy Peredaci Informacii*, 14(1):3–25, 1978.

Zhexiao Lin and Fang Han. Limit theorems of Chatterjee's rank correlation. *arXiv preprint arXiv:2204.08031*, 2022.

Zhexiao Lin and Fang Han. On boosting the power of Chatterjee's rank correlation. *Biometrika*, 110(2):283–299, 2023.

Maximilian Ofner. Quantile regression for weakly dependent functional data. Master's thesis, Graz University of Technology, 2021. Unpublished.

Ronald Pyke. Spacings. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(3):395–436, 1965.

Yosef Rinott. On normal approximation rates for certain sums of dependent random variables. *Journal of Computational and Applied Mathematics*, 55(2):135–143, 1994.

Maria Rizzo and Gabor Szekely. *energy: E-Statistics: Multivariate Inference via the Energy of Data*, 2022. URL `https://CRAN.R-project.org/package=energy`. R package version 1.7-11.

H Shi, M Drton, and F Han. On the power of Chatterjee's rank correlation. *Biometrika*, 109(2):317–333, 2022.

Hongjian Shi, Mathias Drton, and Fang Han. On Azadkia–Chatterjee's conditional dependence coefficient. *Bernoulli*, 30(2):851–877, 2024.

Statistik Austria. Bevölkerung zu Jahresbeginn 2021, 2021. URL `https://statcube.at`. Accessed: 23.10.2021.

Daniel Strenger. *FDEP: Investigating Dependence in Functional Data*, 2024. URL `https://github.com/danielstrenger/FDEP`. R package version 0.1.0.

Gábor Székely, Maria Rizzo, and Nail Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794, 2007.

Aad van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

Qingyang Zhang. On the asymptotic null distribution of the symmetrized Chatterjee's correlation coefficient. *Statistics & Probability Letters*, 194:109759, 2023.

A Additional simulation results and graphics

Tables 3 and 4 compare the means and standard deviations of $\widehat{T}_n$ and $\widehat{R}_n$ for sample sizes $n = 20$ and $n = 1000$ obtained by repeated sampling for the case of rapidly decaying eigenvalues of the covariance operator (Setup (a)). One can observe that for $n = 1000$, the variances of $\widehat{T}_n$ and $\widehat{R}_n$ are almost zero. In addition, the large bias of $\widehat{R}_n$ under independence for $n = 20$ is noteworthy. This bias has largely disappeared for $n = 1000$.

	$\widehat{T}_n$				$\widehat{R}_n$			
`ind`	0 (0.25)				0.43 (0.07)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
`int`	0.74 (0.09)	0.59 (0.13)	0.26 (0.22)	0.05 (0.24)	0.94 (0.03)	0.89 (0.05)	0.7 (0.09)	0.5 (0.08)
`sqnorm`	0.32 (0.19)	0.18 (0.23)	0.05 (0.23)	0 (0.25)	0.63 (0.08)	0.61 (0.08)	0.54 (0.08)	0.47 (0.07)
`weight`	0.74 (0.09)	0.6 (0.13)	0.27 (0.22)	0.04 (0.25)	0.94 (0.03)	0.89 (0.04)	0.71 (0.09)	0.5 (0.09)
`sin`	0.35 (0.2)	0.31 (0.21)	0.16 (0.24)	0.04 (0.24)	0.44 (0.07)	0.46 (0.07)	0.46 (0.06)	0.45 (0.06)
`max`	0.52 (0.14)	0.38 (0.19)	0.16 (0.24)	0.01 (0.25)	0.85 (0.06)	0.81 (0.07)	0.66 (0.09)	0.49 (0.09)
`range`	0.13 (0.23)	0.08 (0.23)	-0.03 (0.23)	-0.03 (0.26)	0.56 (0.07)	0.56 (0.07)	0.51 (0.07)	0.46 (0.06)
`eval`	0.64 (0.11)	0.52 (0.12)	0.26 (0.12)	0.03 (0.12)	0.85 (0.05)	0.81 (0.06)	0.66 (0.06)	0.49 (0.05)

Table 3: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 20$ under Setup (a). Standard deviations are given in brackets.

Tables 5-7 compare the means and standard deviations of $\widehat{T}_n$ and $\widehat{R}_n$ for sample sizes $n = 20$ and $n = 1000$ obtained by repeated sampling for the case of a Brownian motion X with slowly decaying eigenvalues of the covariance operator (Setup (b)). The results do not differ significantly from the ones for Setup (a), supporting our belief that the rapidly decaying eigenvalues as assumed for our theoretical results are in practice more restrictive than necessary.

Figure 6 shows the estimated power of the tests for independence based on $\mathcal{I}_n, \mathcal{I}_n^{\mathrm{DC}}$ and $\mathcal{I}_n^{\mathrm{CvM}}$. The general picture is the same as for Setup (a). At the chosen sample size, all three tests are in tendency — if at all — only slightly less powerful when the eigenvalues decay more slowly.

Figures 7 and 8 visualize the relationships `int`, `weight`, `max`, `range` and `eval`.

	$\widehat{T}_n$				$\widehat{R}_n$			
ind	0 (0.04)				0.06 (0.01)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
int	0.93 (0)	0.69 (0.01)	0.31 (0.03)	0.06 (0.03)	0.94 (0)	0.88 (0.01)	0.64 (0.02)	0.28 (0.03)
sqnorm	0.81 (0.01)	0.49 (0.03)	0.2 (0.04)	0.04 (0.04)	0.45 (0.01)	0.41 (0.01)	0.29 (0.01)	0.14 (0.01)
weight	0.93 (0)	0.69 (0.02)	0.31 (0.03)	0.05 (0.04)	0.94 (0)	0.88 (0.01)	0.63 (0.02)	0.28 (0.03)
sin	0.75 (0.01)	0.61 (0.02)	0.3 (0.03)	0.05 (0.04)	0.2 (0.01)	0.19 (0.01)	0.15 (0.01)	0.09 (0.01)
max	0.81 (0.01)	0.6 (0.02)	0.26 (0.03)	0.04 (0.04)	0.84 (0.01)	0.78 (0.01)	0.57 (0.02)	0.25 (0.02)
range	0.67 (0.01)	0.52 (0.02)	0.21 (0.03)	0.02 (0.04)	0.36 (0.01)	0.33 (0.01)	0.25 (0.01)	0.13 (0.01)
eval	0.87 (0.01)	0.67 (0.01)	0.30 (0.01)	0.05 (0.01)	0.83 (0.01)	0.78 (0.01)	0.57 (0.01)	0.25 (0.01)

Table 4: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 1000$ under Setup (a). Standard deviations are given in brackets.

	$\widehat{T}_n$				$\widehat{R}_n$			
ind	0 (0.25)				0.43 (0.07)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
int	0.78 (0.07)	0.61 (0.13)	0.26 (0.22)	0.05 (0.26)	0.97 (0.01)	0.92 (0.04)	0.72 (0.09)	0.49 (0.09)
sqnorm	0.48 (0.16)	0.29 (0.22)	0.07 (0.24)	-0.01 (0.25)	0.65 (0.08)	0.62 (0.09)	0.53 (0.09)	0.46 (0.07)
weight	0.78 (0.07)	0.61 (0.14)	0.26 (0.24)	0.04 (0.25)	0.97 (0.02)	0.92 (0.03)	0.72 (0.09)	0.5 (0.1)
sin	0.38 (0.2)	0.34 (0.2)	0.17 (0.23)	0.02 (0.25)	0.43 (0.07)	0.44 (0.07)	0.45 (0.06)	0.44 (0.07)
max	0.57 (0.13)	0.43 (0.17)	0.17 (0.25)	0.02 (0.24)	0.91 (0.04)	0.86 (0.05)	0.68 (0.1)	0.49 (0.09)
range	0.15 (0.22)	0.11 (0.24)	0.02 (0.26)	-0.02 (0.26)	0.57 (0.07)	0.56 (0.08)	0.5 (0.08)	0.45 (0.07)
eval	0.63 (0.13)	0.53 (0.15)	0.25 (0.22)	0.03 (0.25)	0.89 (0.04)	0.85 (0.06)	0.68 (0.09)	0.49 (0.09)

Table 5: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 20$ under Setup (b). Standard deviations are given in brackets.

	$\widehat{T}_n$				$\widehat{R}_n$			
ind	0 (0.11)				0.19 (0.03)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
int	0.88 (0.02)	0.67 (0.05)	0.3 (0.11)	0.06 (0.12)	0.98 (0.01)	0.92 (0.02)	0.67 (0.05)	0.34 (0.07)
sqnorm	0.68 (0.05)	0.42 (0.09)	0.17 (0.11)	0.02 (0.12)	0.52 (0.03)	0.48 (0.04)	0.37 (0.05)	0.25 (0.04)
weight	0.89 (0.02)	0.67 (0.05)	0.3 (0.1)	0.06 (0.12)	0.98 (0.01)	0.92 (0.02)	0.67 (0.05)	0.34 (0.07)
sin	0.58 (0.06)	0.5 (0.07)	0.27 (0.1)	0.05 (0.11)	0.24 (0.03)	0.24 (0.03)	0.23 (0.03)	0.22 (0.03)
max	0.67 (0.05)	0.52 (0.07)	0.22 (0.11)	0.03 (0.11)	0.91 (0.02)	0.85 (0.03)	0.63 (0.06)	0.32 (0.07)
range	0.35 (0.09)	0.28 (0.1)	0.1 (0.11)	-0.02 (0.12)	0.42 (0.04)	0.4 (0.04)	0.33 (0.04)	0.24 (0.03)
eval	0.73 (0.04)	0.6 (0.07)	0.27 (0.1)	0.05 (0.12)	0.88 (0.02)	0.83 (0.03)	0.61 (0.06)	0.32 (0.07)

Table 6: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 100$ under Setup (b). Standard deviations are given in brackets.

	$\widehat{T}_n$				$\widehat{R}_n$			
ind	0 (0.04)				0.06 (0.01)			
r^2	1	0.9	0.6	0.1	1	0.9	0.5	0.1
int	0.93 (0)	0.69 (0.02)	0.31 (0.03)	0.06 (0.04)	0.98 (0)	0.91 (0)	0.66 (0.02)	0.29 (0.03)
sqnorm	0.79 (0.01)	0.46 (0.03)	0.19 (0.04)	0.04 (0.04)	0.48 (0.01)	0.44 (0.01)	0.3 (0.01)	0.14 (0.01)
weight	0.93 (0)	0.69 (0.02)	0.31 (0.03)	0.06 (0.04)	0.98 (0)	0.92 (0.01)	0.66 (0.02)	0.29 (0.03)
sin	0.73 (0.01)	0.6 (0.02)	0.3 (0.03)	0.05 (0.04)	0.17 (0.01)	0.16 (0.01)	0.13 (0.01)	0.08 (0.01)
max	0.74 (0.01)	0.56 (0.02)	0.24 (0.03)	0.04 (0.04)	0.91 (0.01)	0.85 (0.01)	0.61 (0.02)	0.27 (0.03)
range	0.49 (0.02)	0.38 (0.03)	0.15 (0.04)	0 (0.04)	0.37 (0.01)	0.35 (0.01)	0.26 (0.01)	0.13 (0.01)
eval	0.79 (0.01)	0.64 (0.02)	0.29 (0.03)	0.05 (0.04)	0.88 (0.01)	0.83 (0.01)	0.6 (0.02)	0.27 (0.03)

Table 7: Comparison of $\widehat{T}_n$ and $\widehat{R}_n$ with a sample size of $n = 1000$ under Setup (b). Standard deviations are given in brackets.

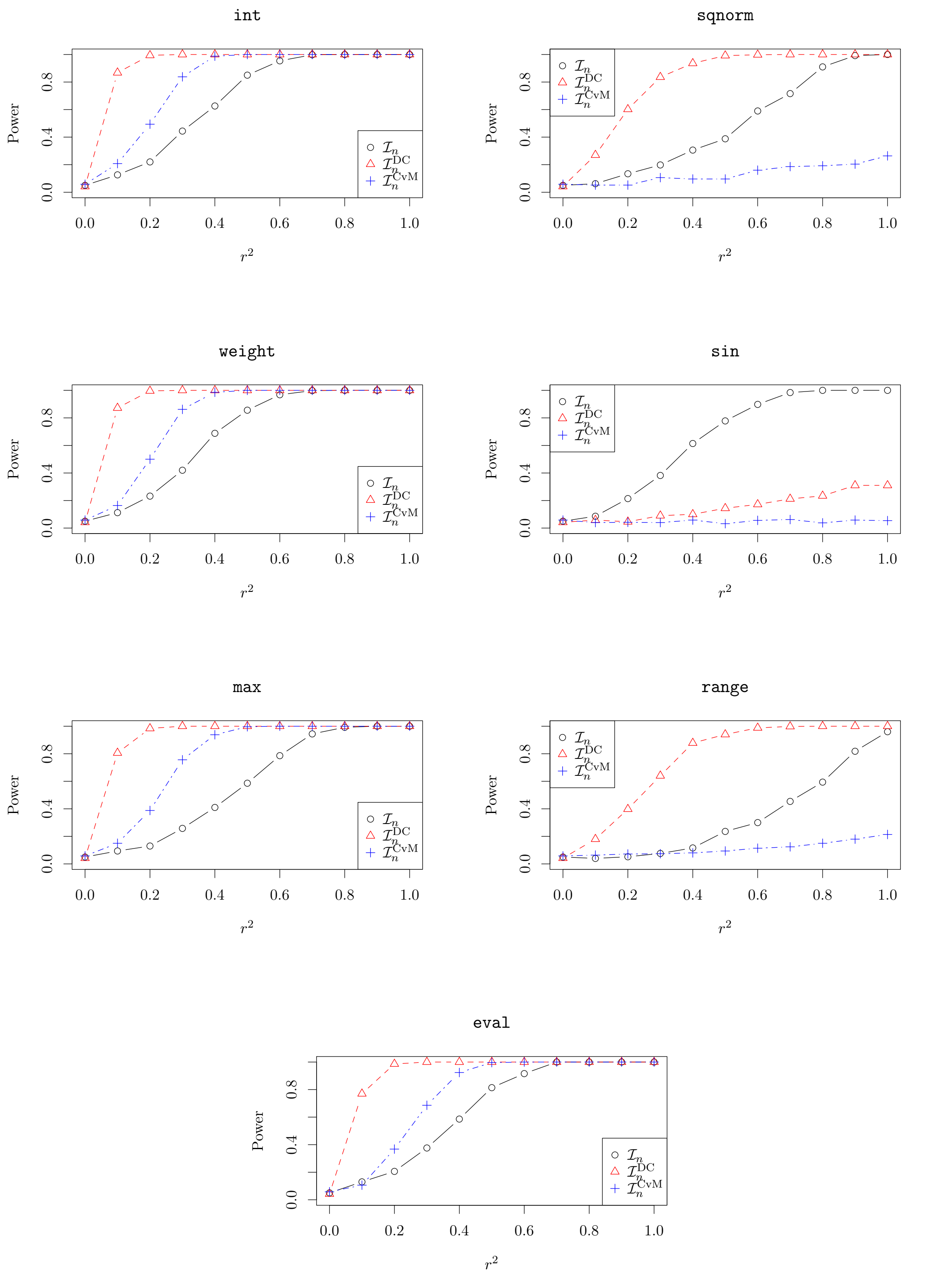
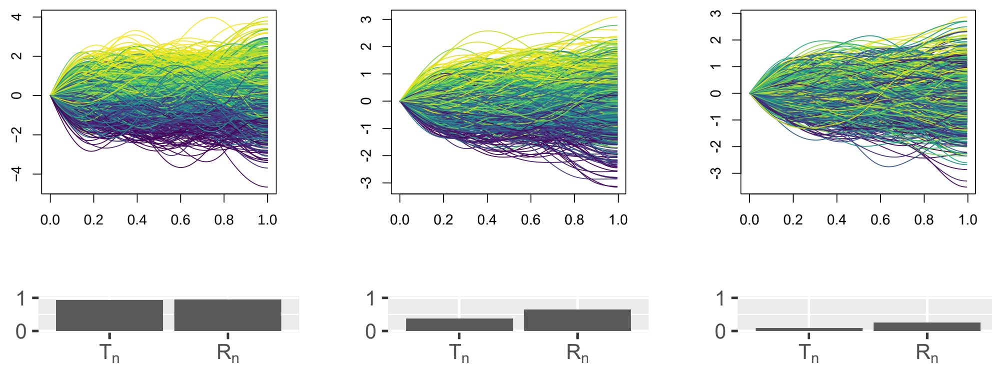
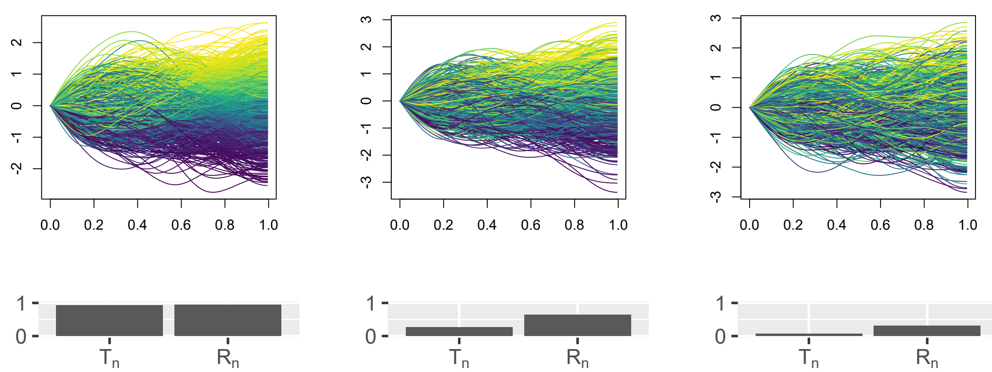


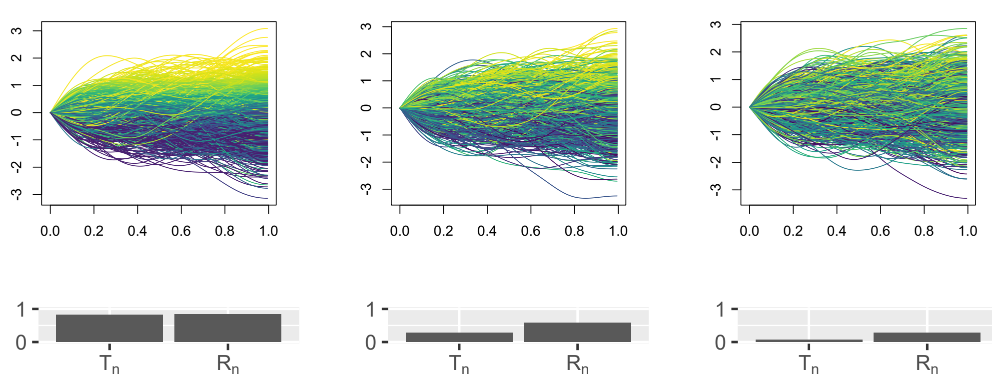
Figure 6: Estimated powers of the three tests for independence at different levels of noise under Setup (b).



(a) $r^2 = 100\%$ (b) $r^2 = 50\%$ (c) $r^2 = 10\%$

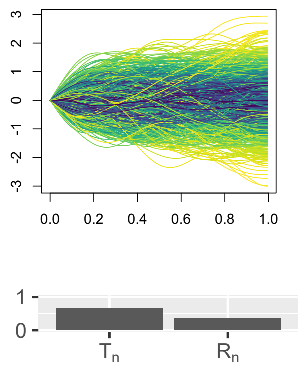


(a) $r^2 = 100\%$ (b) $r^2 = 50\%$ (c) $r^2 = 10\%$

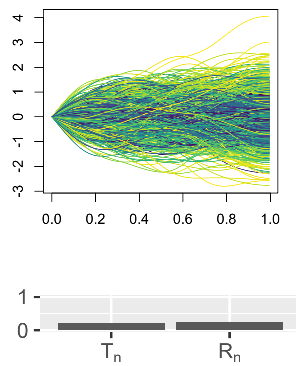


(a) $r^2 = 100\%$ (b) $r^2 = 50\%$ (c) $r^2 = 10\%$

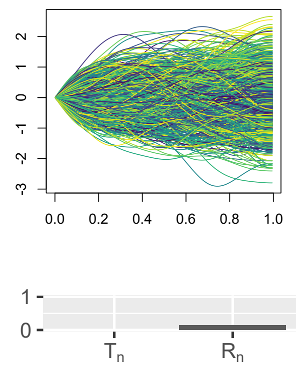
Figure 7: Visualization of the `int` (top), `weight` (middle) and `max` (bottom) relationships under Setup (a). The $n = 1000$ curves X are colored according to Y. The bar plots compare the values of $\widehat{T}_n$ and $\widehat{R}_n$.



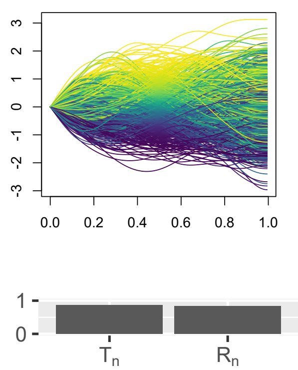
(a) (a) $r^2 = 100\%$



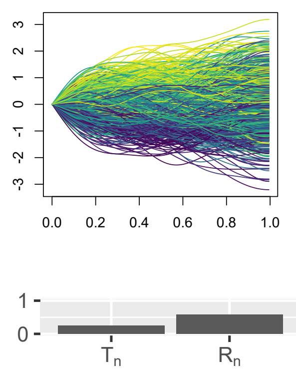
(b) (b) $r^2 = 50\%$



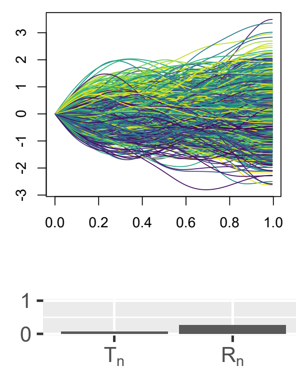
(c) (c) $r^2 = 10\%$



(d) (a) $r^2 = 100\%$



(e) (b) $r^2 = 50\%$



(f) (c) $r^2 = 10\%$

Figure 8: Visualization of the `range` (top) and `eval` (bottom) relationships under Setup (a). The $n = 1000$ curves X are colored according to Y. The bar plots compare the values of $\widehat{T}_n$ and $\widehat{R}_n$.