Estimating scale-invariant directed dependence of bivariate distributions

Robert R. Junker[a,b], Florian Griessenberger[c], Wolfgang Trutschnig[c,*]

[a]*Evolutionary Ecology of Plants, Philipps-University Marburg, Karl-von-Frisch-Strasse 8, 35043 Marburg, Germany*
[b]*Department of Biosciences, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria*
[c]*Department of Mathematics, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria*

Abstract

Asymmetry of dependence is an inherent property of bivariate probability distributions. Being symmetric, commonly used dependence measures such as Pearsons r or Spearmans ρ mask asymmetry and implicitly assume that a random variable Y is equally dependent on a random variable X as vice versa. A copula-based, hence scale-invariant dependence measure called ζ_1 overcoming the just mentioned problem was introduced in 2011. ζ_1 attains values in $[0, 1]$, it is 0 if, and only if X and Y are independent, and 1 if, and only if Y is a measurable function of X. Working with so-called empirical checkerboard copulas allows to construct an estimator ζ_1^n for ζ_1 which is strongly consistent in full generality, i.e., without any smoothness assumptions on the underlying copula. The R-package qad (short for quantification of asymmetric dependence) containing the estimator ζ_1^n is used both, to perform a simulation study illustrating the small sample performance of the estimator as well as to estimate the directed dependence between some global climate variables as well as between world development indicators.

Keywords: asymmetry, copula, correlation, dependence, direction, invariance

1. Introduction

Despite the omnipresence of asymmetry in bivariate distributions all standard methods (applied throughout all fields of science) quantifying the statistical dependence between random variables X and Y largely ignore this key aspect. In fact, classical 'dependence' measures like Pearson's r,

*Corresponding author
Email addresses: `robert.junker@uni-marburg.de` (Robert R. Junker), `florian.griessenberger@sbg.ac.at` (Florian Griessenberger), `wolfgang.trutschnig@sbg.ac.at` (Wolfgang Trutschnig)

Spearman's rank correlation ρ or Kendall's τ are applicable only in specific situations (linearity or monotonicity/concordance) and are by construction symmetric and hence undirected, insinuating that the random variable X is equally dependent on the random variable Y as vice versa.

As a matter of fact, also in many real-life situations dependence between random variables is asymmetric (see, e.g., [1, 2, 3, 5]). Considering, for instance, only symmetric dependence measures may lead to wrong risk assessments in stock market trades (see [30]), or to inaccurate gene network reconstructions (see [43] and the references therein). Moreover, analyzing networks via Pearson's r or Spearman's ρ inevitably leads to symmetric community structures and functional linkages between organisms or individuals and may complicate the understanding of microbial, biochemical and genetic causes for human diseases (see [27]).

In the course of an extensive literature research we found additional, more sophisticated dependence measures, including, e.g., distance correlation $\mathcal{R}$ (see [41]), the maximal information coefficient (MIC) (see [34, 36, 35]), Linfoot correlation (see [26]), robust copula dependence (RCD) (see [4]), and Schweizer and Wolff's famous σ (see [38]). These methods detect independence in the sense that the measure m fulfills $m(X, Y) = 0$ if, and only if X and Y are independent and are, to a certain extend, capable of quantifying symmetric non-linear dependence. Nevertheless, none of these approaches assigns maximum dependence exclusively to pairs (X, Y) of random variables in which Y is a measurable function of X (but not necessarily vice versa). The mutual-information based method MIC assigns maximum dependence to (X, Y) if Y is a measurable function of X or vice versa, in which case we have $MIC(X, Y) = MIC(Y, X) = 1$. The reverse implication, however, does not hold, i.e., we can have $MIC(X, Y) = 1$ although neither Y is a measurable function of X nor vice versa (see Example Appendix A.1). According to Rényi's axioms (see [33]) a bivariate dependence measure must be symmetric. Considering, however, that bivariate distributions may be highly asymmetric we are convinced that a good and applicable dependence measure should be capable of detecting this asymmetry and assign the pair (X, Y) a different value than the pair (Y, X). Figure 1, e.g., depicts a U-shaped sample of size $n = 500$. In this case knowing X strongly improves the predictability of Y - we gain a lot of information about Y by observing X - whereas the information gain in the other direction is significantly smaller.

Ignoring asymmetry in dependence means ignoring valuable information. To the best of the authors' knowledge up to now there exist only two measures m quantifying the dependence of all pairs of continuous random variables (X, Y) which fulfill the following five natural requirements:

(R1) $m(X,Y) \in [0,1]$ (*normalization*)

(R2) $m(X,Y) \neq m(Y,X)$ for at least a pair (X,Y) (*asymmetry*)

(R3) $m(X,Y) = 0$ if, and only if X and Y are independent (*independence*)

(R4) $m(X,Y) = 1$ if, and only if Y is completely dependent on X, i.e., if there exists a measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $Y = f(X)$ almost surely (*complete dependence*)

(R5) If $T, S : \mathbb{R} \to \mathbb{R}$ are strictly monotone then we have $m(T \circ X, S \circ Y) = m(X,Y)$ (*scale invariance*)

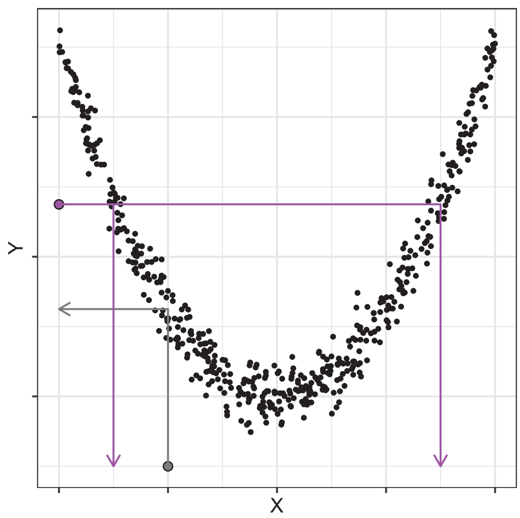

Figure 1: Sample of size $n = 500$ from an asymmetric random vector (X,Y). Knowing the value of X allows to predict the value of Y much better than vice versa (magenta). Standard methods for quantifying dependence are not capable of detecting the asymmetry - we get the following values: Pearson's r: $r(X,Y) = r(Y,X) = 0.04$, Spearman's ρ: $\rho(X,Y) = \rho(Y,X) = 0.02$, Kendall's τ: $\tau(X,Y) = \tau(Y,X) = 0.01$; Székely's distance correlation $\mathcal{R}$: $\mathcal{R}(X,Y) = \mathcal{R}(Y,X) = 0.50$. On the other hand, calculating our measure ζ_1^n in both directions yields $\zeta_1^n(X,Y) = 0.89$ and $\zeta_1^n(Y,X) = 0.47$.

These measures are ζ_1 introduced by Trutschnig in 2011 (see [42]) and the so-called measure of regression dependence *mrd* introduced by Dette et al. in 2013 in [8], which can also be regarded as the L^2-version of ζ_1. In this paper we focus on ζ_1, the results, however, can directly be translated to *mrd*. ζ_1 is constructed via the Markov kernel (conditional distributions) of the (unique) copula

C underlying the pair (X, Y) and is defined as the so-called D_1-distance of C to the product copula Π describing independence. In fact, letting K_A, K_B denote Markov kernels of the bivariate copulas A, B, the D_1-distance is given by

$$D_1(A, B) := \int_{[0,1]} \int_{[0,1]} |K_A(x, [0, y]) - K_B(x, [0, y])| \, d\lambda(x) d\lambda(y)$$

and $\zeta_1(C) \in [0, 1]$ is defined as $\zeta_1(C) = 3\, D_1(C, \Pi)$ whereby 3 is just a normalizing constant. In other words, we have

$$\zeta_1(C) = 3 \int_{[0,1]} \int_{[0,1]} |K_A(x, [0, y]) - y| \, d\lambda(x) d\lambda(y).$$

For proofs of the fact that ζ_1 fulfills the requirements (R1)-(R5) and for additional information on D_1 and ζ_1 we refer to [42], in this paper we focus on how to estimate ζ_1.

Considering that ζ_1 is constructed via conditional distributions it is apriori unclear if good estimators can be found at all, and, in the positive case, how restrictive the required smoothness assumptions are. In this paper we construct a simple estimator ζ_1^n for ζ_1 which is based on so-called empirical checkerboard copulas and show the surprising fact that this estimator is strongly consistent in full generality, i.e., without any smoothness assumptions on the underlying copula. Due to its good performance we implemented the estimator in the R-package *qad* (short for quantification of asymmetric dependence). Using qad we then estimate the directed dependence between some global climate variables as well as between world development indicators, and perform a simulation study illustrating the small sample performance of our estimator.

Remark 1.1. The estimator q returned in the output of the functions contained in the R-package *qad* (see [14]) is exactly the estimator ζ_1^n studied here. We opted for the name q in the R-package solely for the sake of simplicity and convenience.

The rest of the paper is organized as follows: Notation and preliminaries are reviewed in Section 2. Section 3 recalls the definition of an empirical checkerboard copula, sketches the idea underlying the estimator ζ_1^n and provides the main theoretical result of the paper, i.e., strong consistency of ζ_1^n. Section 4 presents a simulation study analyzing the small sample performance of the estimator for three different families of dependence structures. Again using the R-package *qad* the aforementioned dependence between global climate variables and between world development indicators is estimated and illustrated in Section 5. Finally, Section 6 provides an outlook and completes the paper.

2. Notation and preliminaries

Throughout the paper $\mathbb{R}$ will denote the real numbers, $\mathbb{N}$ the natural numbers, $\mathcal{C}$ will denote the family of all two-dimensional copulas (for background on copulas we refer to [10, 12, 28] and the references therein). For every copula $A \in \mathcal{C}$ the corresponding doubly stochastic measure will be denoted by μ_A. As usual, $d_\infty(A, B)$ will denote the uniform metric on $\mathcal{C}$, i.e. $d_\infty(A, B) := \max_{(x,y)\in[0,1]^2} |A(x, y) - B(x, y)|$. It is well known that $(\mathcal{C}, d_\infty)$ is a compact metric space. For every metric space (Ω, d) the Borel σ-field will be denoted by $\mathcal{B}(\Omega)$ and λ will denote the Lebesgue measure on $\mathcal{B}([0, 1])$.

A mapping $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ is called a Markov kernel from $\mathbb{R}$ to $\mathcal{B}(\mathbb{R})$ if $x \mapsto K(x, B)$ is measurable for every fixed $B \in \mathcal{B}(\mathbb{R})$ and $B \mapsto K(x, B)$ is a probability measure for every fixed $x \in \mathbb{R}$. A Markov kernel $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ is called regular conditional distribution of a (real-valued) random variable Y given (another random variable) X if for every $B \in \mathcal{B}(\mathbb{R})$

$$K(X(\omega), B) = \mathbb{E}(\mathbb{1}_B \circ Y|X)(\omega)$$

holds $\mathbb{P}$-almost surely, whereby $\mathbb{1}_B(x)$ denotes the indicator function. It is well known that a regular conditional distribution of Y given X exists and is unique $\mathbb{P}^X$-almost sure. For every $A \in \mathcal{C}$ the corresponding regular conditional distribution (i.e. the regular conditional distribution of Y given X in the case that $(X, Y) \sim A$) will be denoted by $K_A(\cdot, \cdot)$. Note that for every $A \in \mathcal{C}$ and Borel sets $E, F \in \mathcal{B}([0, 1])$ we have

$$\int_E K_A(x, F)d\lambda(x) = \mu_A(E \times F).$$

For more details and properties of regular conditional distributions and disintegration see [18, 22].

In the sequel we will mainly work with the metrics D_1 and D_∞ introduced in [42]. These metrics are defined by

$$D_1(A, B) := \int_{[0,1]} \underbrace{\int_{[0,1]} |K_A(x, [0, y]) - K_B(x, [0, y])| \, d\lambda(x)}_{=:\Phi_{A,B}(y)} \, d\lambda(y)$$

and

$$D_\infty(A, B) := \sup_{y\in[0,1]} \Phi_{A,B}(y),$$

respectively. According to [42] $\Phi_{A,B} : [0,1] \to [0,1]$ is Lipschitz-continuous with Lipschitz constant 2 and both metrics generate the same topology (without being equivalent). The resulting metric space $(\mathcal{C}, D_1)$ is complete and separable and it can be shown that, firstly, $D_1(A,\Pi)$ attains only values in $\left[0, \frac{1}{3}\right]$ and that, secondly, $D_1(A,\Pi)$ is maximal if and only if A is completely dependent, i.e. if a λ-preserving transformation $h : [0,1] \to [0,1]$ exists such that $K_A(x, \{h(x)\}) = 1$ for λ-a.e. $x \in [0,1]$. In the sequel we will let $\mathcal{C}_d$ denote the family of all completely dependent copulas, and write A_h and $K_h(\cdot,\cdot)$ for the completely dependent copula and the corresponding Markov kernel induced by the λ-preserving transformation h, respectively. For equivalent definitions and properties of completely dependent copulas we refer to [42] and the references therein.

We recall the definition of the dependence measure ζ_1 fulfilling (R1)-(R5) in the Introduction and introduce the notion of (dependence) asymmetry:

Definition 2.1. Suppose that X, Y are continuous random variables with joint distribution function H and copula A. Then the dependence measure ζ_1 is defined as

$$\zeta_1(X,Y) = \zeta_1(A) := 3D_1(A,\Pi). \tag{1}$$

The induced measure of (dependence) asymmetry α is given by

$$\alpha(X,Y) = \alpha(A) := \zeta_1(A) - \zeta_1(A^t) \in (-1,1), \tag{2}$$

where A^t denotes the transpose of A, i.e. $A^t(x,y) = A(y,x)$ for all $x, y \in [0,1]$.

Remark 2.2. The measure of (dependence) asymmetry α quantifies if knowing X on average increases predictability of Y more than vice versa but says nothing about the distances $d_\infty(A, A^t)$ or $D_1(A, A^t)$. In fact, it is even possible to construct a copula A which has maximal d_∞- (or D_1-) distance to its transposed copula A^t and nevertheless fulfills $\alpha(A) = 0$ (see, for instance, the mutually completely dependent copulas with maximal d_∞- (or D_1-) asymmetry as studied in [19] or [29].

3. Estimating $\zeta_1(X,Y)$

We first recall the notion of an empirical copula (E_n), sketch why it is not possible to estimate $\zeta_1(A)$ by simply plugging in the empirical copula, i.e., by $\zeta_1(E_n)$, and then introduce the so-called empirical checkerboard copula/aggregation and prove strong consistency.

3.1. Empirical (checkerboard) copula

Let (X, Y) be a random vector with continuous joint distribution function H, margin distributions F, G and copula $A \in \mathcal{C}$. Furthermore let $(x_1, y_1), \ldots, (x_n, y_n)$ denote a sample of (X, Y), H_n the bivariate empirical distribution function and F_n, G_n the univariate empirical marginal distribution functions. Sklar's Theorem (see [40]) implies that there exists a unique subcopula $E_n' : Range(F_n) \times Range(G_n) \to Range(H_n)$ (see [28]) fulfilling

$$H_n(x, y) = E_n'(F_n(x), G_n(y)),$$

for all $(x, y) \in Range(F_n) \times Range(G_n)$. Since X, Y are assumed to be continuous random variables ties only occur with probability 0 so $Range(F_n) = Range(G_n) = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$ holds almost surely. There are infinitely many ways to extend E_n' to a copula, in the sequel we will only work with the bilinear interpolation (see [28]) and refer to this extension E_n simply as empirical copula of the sample $(x_1, y_1), \ldots, (x_n, y_n)$ (also see [7, 13]).

Remark 3.1. The R-package *qad* (available on CRAN [32]) calculates the empirical copula also in the case of ties (the construction method via bilinear interpolation is exactly the same). If there are no ties the package calls the function `C.n` from the *copula*-package (see [16, 44]), otherwise it uses an internal function to calculate the empirical (checkerboard), i.e. in both cases E_n is really a copula. In what follows we only consider the case of continuous random variables since in this case the underlying copula is unique.

The following result for empirical copulas is the key for proving consistency of our empirical checkerboard estimator ζ_1^n.

Theorem 3.2 (Janssen, et al. [17]). *Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a random sample from (X, Y) and assume that (X, Y) has continuous joint distribution function H and copula A. Then the following asymptotic result holds for the empirical copula E_n with probability* 1*:*

$$d_\infty(E_n, A) = O\left(\sqrt{\frac{\log(\log(n))}{n}}\right) \tag{3}$$

Estimating ζ_1 by $\zeta_1(E_n)$ only produces a reasonable estimator if A is completely dependent - the following result holds:

Proposition 3.3. *Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a random sample from (X, Y), whereby (X, Y) has continuous joint distribution H and copula A. Then with probability 1 we have*

$$\lim_{n\to\infty} D_1(E_n, \Pi) = \frac{1}{3}.$$

Proof. For every $x \in [0,1]$ there exists exactly one $i(x) \in \{0, 1, \ldots, n-1\}$ such that

$$K_{E_n}\left(x, \left[\frac{i(x)}{n}, \frac{i(x)+1}{n}\right]\right) = 1$$

holds. As direct consequence we get

$$\begin{aligned}
D_1(E_n, \Pi) &= \int_{[0,1]} \int_{[0,1]} |K_{E_n}(x, [0,y]) - y| \, d\lambda(x) d\lambda(y) \\
&\geq \int_{[0,1]} \int_{\left[0, \frac{i(x)}{n}\right] \cup \left[\frac{i(x)+1}{n}, 1\right]} |K_{E_n}(x, [0,y]) - y| \, d\lambda(y) d\lambda(x) \\
&= \int_{[0,1]} \left(\int_{\left[0, \frac{i(x)}{n}\right]} y \, d\lambda(y) + \int_{\left[\frac{i(x)+1}{n}, 1\right]} (1-y) \, d\lambda(y) \right) d\lambda(x) \\
&= \int_{[0,1]} \left(\frac{i(x)^2}{2n^2} + 1 - \frac{i(x)+1}{n} - \left(\frac{1}{2} - \frac{(i(x)+1)^2}{2n^2} \right) \right) d\lambda(x) \\
&= \frac{1}{2} + \frac{1}{n} + \frac{1}{2n^2} + \underbrace{\int_{[0,1]} \frac{i(x)^2}{n^2} \, dx}_{=:I_1^n} - \underbrace{\int_{[0,1]} \frac{i(x)}{n} \, dx}_{=:I_2^n} + \underbrace{\int_{[0,1]} \frac{i(x)}{n^2} \, dx}_{=:I_3^n} .
\end{aligned}$$

A straightforward calculation shows $\lim_{n\to\infty} I_1^n = \frac{1}{3}$, $\lim_{n\to\infty} I_2^n = \frac{1}{2}$ and $\lim_{n\to\infty} I_3^n = 0$, hence, considering $D_1(E_n, \Pi) \leq \frac{1}{3}$ the desired result follows. □

Loosely speaking, the main problem with the empirical copula is that it is very close to complete dependence (the corresponding kernels concentrate mass on intervals of length $\frac{1}{n}$). One possibility to overcome the problem is to simply smooth/aggregate the empirical copula and use Bernstein-, Beta- or so-called Checkerboard-approximations of the empirical copula (see [6, 9, 23, 31, 37, 39]). Considering the simplicity of checkerboard copulas (they are merely two-dimensional histograms), the generality of the obtained results (no smoothness assumptions) and the low computational time in comparison to Bernstein- or Beta-copulas the rest of the paper focuses on checkerboards. Following [24] we proceed in this way: Fix $N \in \mathbb{N}$ and define the squares R_{ij}^N for $i, j \in \{1, \ldots, N\}$ by

$$R_{ij}^N = \left[\frac{i-1}{N}, \frac{i}{N}\right] \times \left[\frac{j-1}{N}, \frac{j}{N}\right].$$

Definition 3.4. A copula $A_N \in \mathcal{C}$ is called N-checkerboard copula, if A_N is absolutely continuous and (a version of) its density k_{A_N} is constant on the interior of each square R_{ij}^N. We refer to N as the resolution of A_N, denote the set of all $N-$checkerboard copulas by $\mathcal{CB}_N$, and set $\mathcal{CB} = \bigcup_{N=1}^{\infty} \mathcal{CB}_N$.

Definition 3.5. For $A \in \mathcal{C}$ and $N \in \mathbb{N}$ the (absolutely continuous) copula $CB_N(A) \in \mathcal{CB}_N$, defined by

$$CB_N(A)(x,y) := \int_0^x \int_0^y N^2 \sum_{i,j=1}^{N} \mu_A(R_{ij}^N) \mathbb{1}_{R_{ij}^N}(s,t)\, d\lambda(t) d\lambda(s) \tag{4}$$

is called N-checkerboard approximation of A or simply N-checkerboard of A.

Example 3.6. Suppose that X and Y are independent (so their underlying copula is Π) and let $(x_1, y_1), \ldots, (x_n, y_n)$ be a sample of size $n = 100$ from Π. Figure 2 depicts a scatterplot of the sample as well as the density of the corresponding empirical 5-checkerboard of E_n.

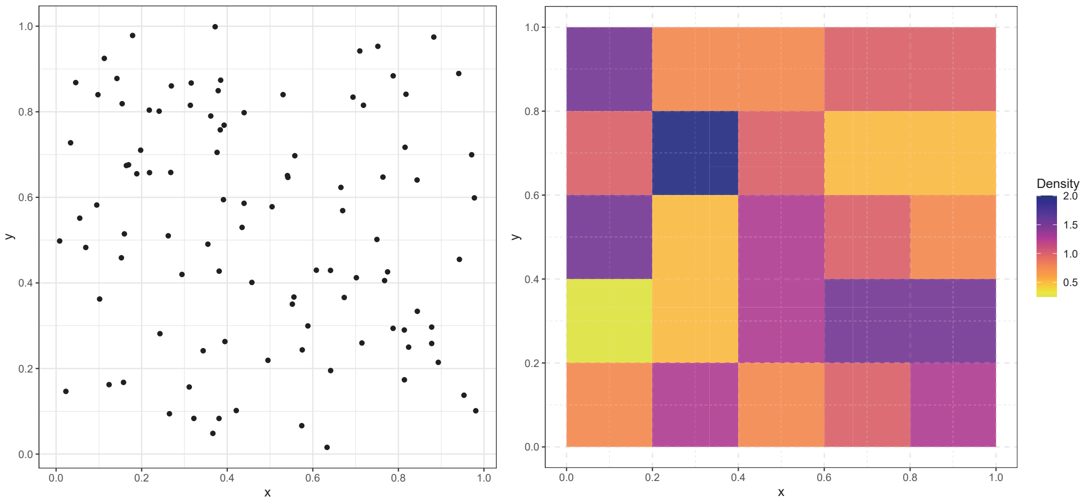

Figure 2: Scatterplot of the sample $(x_1, y_1), \ldots, (x_n, y_n)$ of size $n = 100$ from Π (left panel) and image plot of the density of the empirical 5-checkerboard copula of E_n (right panel).

According to [42] $\mathcal{CB}$ is dense in $(\mathcal{C}, D_1)$. Furthermore, translating the results in [24] to the metric D_1 yields the following:

Theorem 3.7 (Li, et al.)**.** *For every copula $A \in \mathcal{C}$ we have*

$$\lim_{N\to\infty} D_1(CB_N(A), A) = 0 = \lim_{N\to\infty} D_1(CB_N(A)^t, A^t).$$

Plugging in the checkerboard approximation $CB_N(E_n)$ the empirical copula E_n with resolution N yields what we will refer to as *empirical N checkerboard* in the sequel. We will see that ζ_1^n, defined by

$$\zeta_1^n := \zeta_1(CB_N(E_n))$$

is a strongly consistent estimator if the resolution N is chosen as a suitable function $N(n)$ of the sample size n. Before proving this main result in several steps we summarize and illustrate the chosen procedure (also implemented in the R-package *qad*):

1. Given a sample $(x_1, y_1), \dots, (x_n, y_n)$ from a (continuous) random vector (X, Y) (see Figure 3a and 4a) with copula A we first calculate the so-called pseudo-observations (normalized ranks) as well as the empirical copula E_n (Figure 3b and 4b depict the corresponding empirical copulas E_n as well as the pseudo-observations).
2. As second step we smooth/aggregate the empirical copula to the so-called empirical checkerboard copula $CB_N(E_n)$, i.e., the unit square $[0,1]^2$ is partitioned into N^2 squares of equal size and the total mass per square is calculated. In the package *qad* the resolution $N \in \mathbb{N}$ is chosen as the floor function of the square root of the sample size n. Figure 3c and 4c depict the resulting N checkerboard copula $CB_N(E_n)$.
3. We calculate the estimator $\zeta_1^n := \zeta_1(CB_N(E_n))$ of $\zeta_1(A)$ and interpret it as the dependence of Y on X, or, equivalently, as the information gain of Y by knowing X (see Figure 3d and 4d).
4. We proceed analogously with the transposed sample $(y_1, x_1), \dots, (y_n, x_n)$ and calculate $\zeta_1(CB_N(E_n^t))$ as estimator of $\zeta_1^n(A^t)$ (see Figure 3e and 4e).
5. Based on $\zeta_1(CB_N(E_n))$ and $\zeta_1(CB_N(E_n^t))$ we calculate

$$\alpha_n := \zeta_1(CB_N(E_n)) - \zeta_1(CB_N(E_n^t))$$

as estimator of the (dependence) asymmetry $\alpha(A) = \zeta_1(A) - \zeta_1(A^t)$.

3.2. Strong consistency of ζ_1^n

We start with the following lemma linking D_∞ and d_∞ for checkerboards.

Lemma 3.8. *For all $A_N, B_N \in \mathcal{CB}_N$ the inequality*

$$D_\infty(A_N, B_N) \leq 2(N-1)d_\infty(A_N, B_N) \tag{5}$$

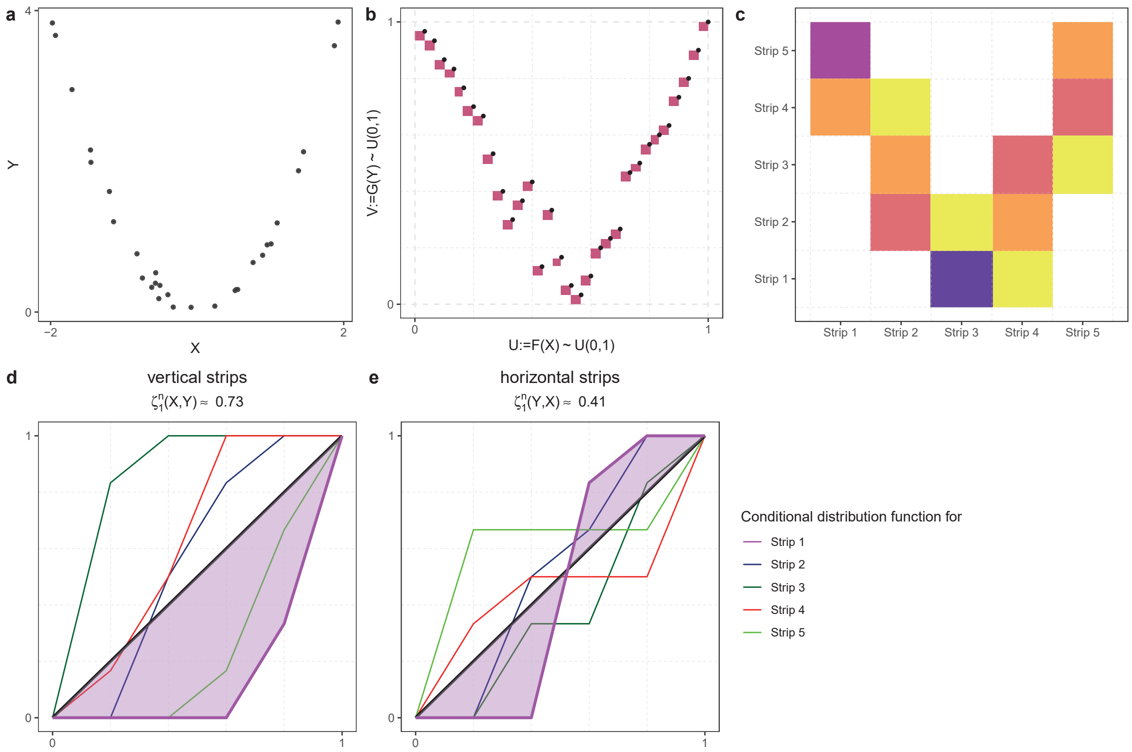

Figure 3: The bivariate sample $(x_1, y_1), \dots (x_n, y_n)$ of size $n = 30$ in panel (a) is drawn from the quadratic, slightly noisy relationship $Y = X^2 + \varepsilon$ with ε being normally distributed with mean 0 and standard deviation 0.1. The density of the empirical copula E_{30} (together with the pseudo-observations) and the density of the empirical 5 checkerboard are calculated (b,c). $\zeta_1(CB_N(E_n))$ and $\zeta_1(CB_N(E_n^t))$ are calculated as the average L^1-distances of the conditional distribution functions and the identity function $id_{[0,1]}$ modelling independence, i.e., the area between the colored lines and the black line is calculated (d,e). The area for the first vertical and horizontal strip is highlighted in magenta. The obtained values are $\zeta_1(CB_N(E_n)) = 0.73$, $\zeta_1(CB_N(E_n^t)) = 0.41$ and $\alpha_n = 0.32$.

holds. Ineq. (5) is sharp, i.e., for every $N \in \mathbb{N}$ we can find copulas $A_N, B_N \in \mathcal{CB}_N$ for which equality holds.

Proof. Fix an arbitrary $i \in \{1, \dots, N\}$. Then for every $x \in \left[\frac{i-1}{N}, \frac{i}{N}\right]$ and $y \in [0, 1]$, continuity of A_N, disintegration, and the fact that $K_{A_N}(x, \cdot)$ is constant on each interval of the form $\left(\frac{i-1}{N}, \frac{i}{N}\right)$

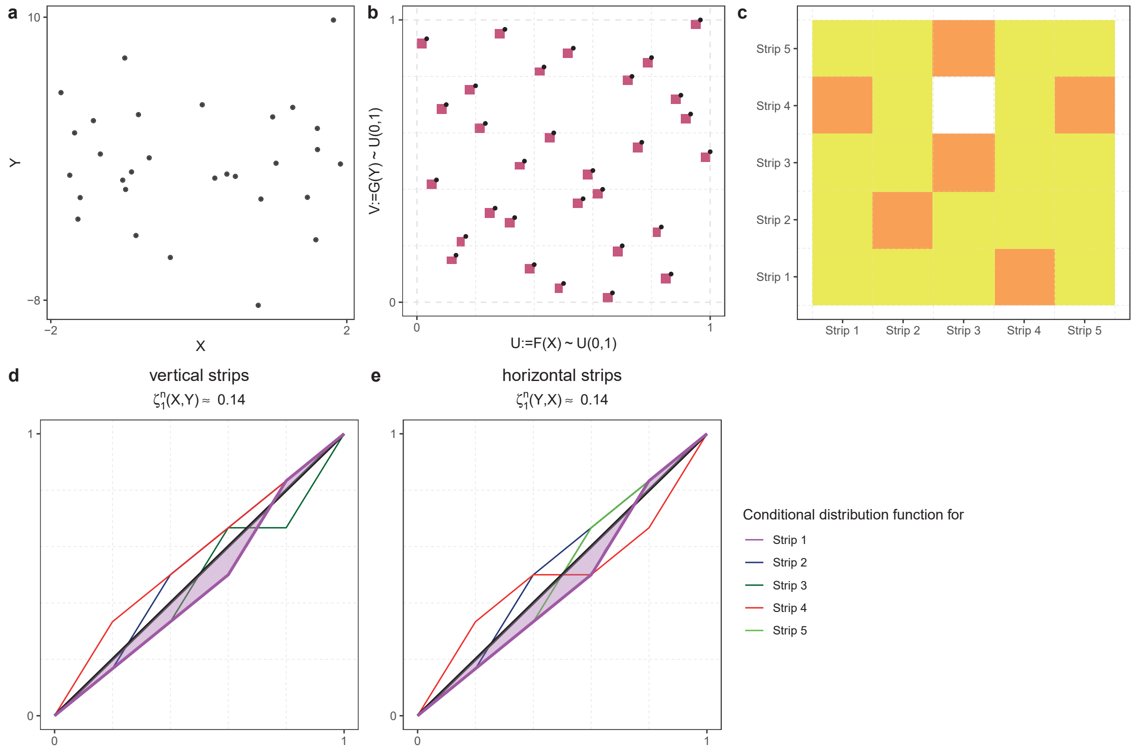

Figure 4: The bivariate sample $(x_1, y_1), \dots (x_n, y_n)$ of size $n = 30$ in panel (a) is drawn from the quadratic, very noisy relationship $Y = X^2 + \varepsilon$ with ε being normally distributed with mean 0 and standard deviation 4. The density of the empirical copula E_{30} (together with the pseudo-observations) and the density of the empirical 5 checkerboard are calculated (b,c). $\zeta_1(CB_N(E_n))$ and $\zeta_1(CB_N(E_n^t))$ are calculated as the average L^1-distances of the conditional distribution functions and the identity function $id_{[0,1]}$ modelling independence, i.e., the area between the colored lines and the black line is calculated (d,e). The area for the first vertical and horizontal strip is highlighted in magenta. The obtained values are $\zeta_1(CB_N(E_n)) = 0.14$, $\zeta_1(CB_N(E_n^t)) = 0.14$ and $\alpha_n(X,Y) = 0.00$.

yield

$$
\begin{aligned}
A_N\left(\tfrac{i}{N}, y\right) - A_N\left(\tfrac{i-1}{N}, y\right) &= \mu_{A_N}\left(\left[\tfrac{i-1}{N}, \tfrac{i}{N}\right] \times [0,y]\right) \\
&= \int_{\left[\frac{i-1}{N}, \frac{i}{N}\right]} K_{A_N}(s,[0,y])\, d\lambda(s) = \tfrac{1}{N} K_{A_N}(x,[0,y]).
\end{aligned}
$$

Altogether we therefore get

$$
D_\infty(A_N, B_N) = \sup_{y \in [0,1]} \int_{[0,1]} \left| K_{A_N}(x,[0,y]) - K_{B_N}(x,[0,y]) \right| d\lambda(x)
$$

$$
\begin{aligned}
&= \sup_{y\in[0,1]} \sum_{i=1}^{N} \int_{\left[\frac{i-1}{N},\frac{i}{N}\right]} |K_{A_N}(x,[0,y]) - K_{B_N}(x,[0,y])| \, d\lambda(x) \\
&= \sup_{y\in[0,1]} \sum_{i=1}^{N} \left| A_N\left(\tfrac{i}{N},y\right) - A_N\left(\tfrac{i-1}{N},y\right) - B_N\left(\tfrac{i}{N},y\right) + B_N\left(\tfrac{i-1}{N},y\right) \right| \\
&= \sup_{y\in[0,1]} \Bigg\{ \sum_{i=2}^{N-1} \left| A_N\left(\tfrac{i}{N},y\right) - A_N\left(\tfrac{i-1}{N},y\right) - B_N\left(\tfrac{i}{N},y\right) + B_N\left(\tfrac{i-1}{N},y\right) \right| + \\
&\qquad + \left| A_N(\tfrac{1}{N},y) - B_N(\tfrac{1}{N},y) \right| + \left| A_N(\tfrac{N-1}{N},y) - B_N(\tfrac{N-1}{N},y) \right| \Bigg\} \\
&\leq \left(\sum_{i=2}^{N-1} 2d_\infty(A_N,B_N) \right) + 2d_\infty(A_N,B_N) \\
&= 2(N-1)d_\infty(A_N,B_N),
\end{aligned}
$$

which completes the proof of ineq. (5). Figure A.23 and Figure A.24 in the Appendix depict copulas A_N, B_N for which equality holds. Since the construction idea easily extends to arbitrary $N \in \mathbb{N}$ the proof is complete. □

As next step we derive a slightly improved inequality linking D_1 and D_∞ (compare with Theorem 6 in [42] and Lemma 3 in [11]):

Lemma 3.9. *Let $N \in \mathbb{N}$ and $A_N, B_N \in \mathcal{CB}_N$. Then the following inequality holds:*

$$
D_1(A_N,B_N) \leq \left(\frac{N-1}{N}\right) D_\infty(A_N,B_N). \tag{6}
$$

Inequality (6) is best possible and we have equality if there exists some $c \in [0, \frac{2}{N}]$ such that $\Phi_{A_N,B_N}(y) = c$ holds on $\left[\frac{1}{N}, \frac{N-1}{N}\right]$.

Proof. Considering the facts that $\Phi_{A_N,B_N}(0) = \Phi_{A_N,B_N}(1) = 0$ and that $y \mapsto \Phi_{A_N,B_N}(y)$ is piecewise linear and Lipschitz-continuous with Lipschitz-constant 2 we get

$$
\max_{y\in[0,1]} \Phi_{A_N,B_N}(y) = \max_{y\in\left[\frac{1}{N},\frac{N-1}{N}\right]} \Phi_{A_N,B_N}(y),
$$

from which the assertions follow immediately. □

Considering N-checkerboard approximations of arbitrary copulas can not increase their d_∞-distance - the following lemma holds:

Lemma 3.10. *For arbitrary* $A, B \in \mathcal{C}$ *and their* N*-checkerboard approximations* $CB_N(A)$, $CB_N(B) \in \mathcal{CB}_N$ *we have*

$$d_\infty(CB_N(A), CB_N(B)) \leq d_\infty(A, B). \tag{7}$$

The inequality is sharp for every $N \in \mathbb{N}$.

Proof. The inequality directly follows from

$$\begin{aligned} d_\infty(CB_N(A), CB_N(B)) &= \sup_{(x,y)\in[0,1]^2} |CB_N(A)(x,y) - CB_N(B)(x,y)| \\ &= \sup_{(x,y)\in\{\frac{1}{N},\ldots,\frac{N-1}{N}\}^2} |CB_N(A)(x,y) - CB_N(B)(x,y)| \\ &= \sup_{(x,y)\in\{\frac{1}{N},\ldots,\frac{N-1}{N}\}^2} |A(x,y) - B(x,y)| \\ &\leq \sup_{(x,y)\in[0,1]^2} |A(x,y) - B(x,y)| \\ &= d_\infty(A, B). \end{aligned}$$

Considering the fact that for $C \in \mathcal{CB}_N$ we obviously have $CB_N(C) = C$ the inequality is best possible. □

Combining the previous inequality yields the following corollary:

Corollary 3.11. *For arbitrary copulas* $A, B \in \mathcal{C}$ *and their* N*-checkerboard approximations* $CB_N(A)$, $CB_N(B) \in \mathcal{CB}$ *the following inequality holds:*

$$D_1(CB_N(A), CB_N(B)) \leq 2\frac{(N-1)^2}{N}\, d_\infty(A, B). \tag{8}$$

Based on Corollary 3.11 we can now prove the desired strong consistency result for our estimator $\zeta_1^n = \zeta_1(CB_N(E_n))$.

Theorem 3.12. *Let* $(X_1, Y_1), (X_2, Y_2), \ldots$ *be a random sample from* (X, Y) *and assume that* (X, Y) *defined on* $(\Omega, \mathcal{A}, \mathbb{P})$ *has continuous joint distribution function* H *and copula* A*. Setting* $N(n) := \lfloor n^s \rfloor$ *for some* $s \in \left(0, \frac{1}{2}\right)$ *the following identity holds with probability* 1*:*

$$\lim_{n\to\infty} D_1(CB_{N(n)}(E_n), A) = 0. \tag{9}$$

Proof. Applying Corollary 3.11 and the triangle inequality yields

$$\begin{aligned} D_1(CB_{N(n)}(E_n), A) &\le D_1(CB_{N(n)}(E_n), CB_{N(n)}(A)) + D_1(CB_{N(n)}(A), A) \\ &\le 2\frac{(N(n)-1)^2}{N(n)}\, d_\infty(E_n, A) + D_1(CB_{N(n)}(A), A) \\ &\le 2N(n)\, d_\infty(E_n, A) + D_1(CB_{N(n)}(A), A) \end{aligned}$$

According to Theorem 3.2 there exists a set $\Lambda \in \mathcal{A}$ with $\mathbb{P}(\Lambda) = 1$ such that for every $\omega \in \Lambda$ we can find a constant $C(\omega) > 0$ and an index $n_0 = n_0(\omega) \in \mathbb{N}$ such that for all $n \ge n_0$

$$d_\infty(E_n(\omega), A) \le C(\omega)\sqrt{\frac{\log(\log(n))}{n}}$$

holds. Let $\varepsilon > 0$ be fixed. Theorem 3.7 implies the existence of an index $n_1 \in \mathbb{N}$ fulfilling

$$D_1(CB_{N(n)}(A), A) < \frac{\varepsilon}{2}$$

for all $n \ge n_1$. For every $\omega \in \Lambda$ and $n \ge \max\{n_0, n_1\}$ we therefore obtain

$$\begin{aligned} D_1(CB_{N(n)}(E_n(\omega)), A) &\le 2N(n)C(\omega) \cdot \sqrt{\frac{\log(\log(n))}{n}} + D_1(CB_{N(n)}(A), A) \\ &< 2\, C(\omega) \lfloor n^s \rfloor n^{-\frac{1}{2}} \sqrt{\log(\log(n))} + \frac{\varepsilon}{2} \end{aligned}$$

from which the result follows since $0 < s < \frac{1}{2}$ by assumption. □

We conclude this section with proving strong consistency of the estimator

$$\zeta_1^n = \zeta_1(CB_{N(n)}(E_n))$$

of $\zeta_1(A)$ - notice that no smoothness assumptions are needed.

Theorem 3.13. *Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a random sample from (X, Y) and assume that (X, Y) has continuous joint distribution function H and copula A. Fix $s \in \left(0, \frac{1}{2}\right)$ and set $N(n) := \lfloor n^s \rfloor$ for every $n \in \mathbb{N}$. Then with probability 1 we have*

$$\lim_{n\to\infty} \zeta_1(CB_{N(n)}(E_n)) = \zeta_1(A). \tag{10}$$

Proof. Direct consequence of Theorem 3.12 and the fact that

$$|D_1(CB_{N(n)}(E_n), \Pi) - D_1(A, \Pi)| \le D_1(CB_{N(n)}(E_n), A).$$

□

Remark 3.14. Simulations insinuate that $CB_{N(n)}(E_n)$ might also be a strongly consistent estimator for more flexible choices of $N(n)$, in particular for the case $N(n) := \lfloor n^s \rfloor$ and some $s \geq \frac{1}{2}$. Analogous to the selection of the optimal grid size function $B(n)$ within the calculation of MIC (see [34]) the choice of the parameter s is essential for the performance of the estimator ζ_1^n. Not surprisingly, the optimal value of s depends on the underlying copula: if s increases then the performance of the estimator ζ_1^n improves in the situation of complete dependence, in the independence setting, however, it decreases. Example Appendix A.3 and Appendix A.4 illustrate this behavior. Since in all simulation studies covering a broad spectrum of different dependence structures choosing $s = \frac{1}{2}$ (or, more precisely, choosing s slightly smaller than $\frac{1}{2}$ depending on the number of ties in the data) resulted in good results this value is used in the R-package *qad*.

Remark 3.15. We conjecture that a strong consistency result analogous to Theorem 3.13 also holds for Bernstein- and Beta-copulas. Nevertheless, as mentioned in the Introduction, empirical checkerboard copulas are much easier and faster to compute (even for large sample sizes) and proving strong consistency of the Bernstein- or Beta estimator in full generality seems even more tedious than it was in the current setup.

4. Simulation study

If not specified differently throughout this section we consider $s_0 = \frac{1}{2}$ and set the resolution to $N(n) := \lfloor n^{s_0} \rfloor$. In order to illustrate the small sample performance of our estimator $\zeta_1(CB_{N(n)}(E_n))$ we consider Marshall-Olkin and FGM copulas as well as completely dependent copulas A_h such that $\zeta_1(A_h)$ is much greater than $\zeta_1(A_h^t)$.

4.1. Marshall Olkin family

The Marshall Olkin (MO) family of copulas $(M_{\alpha,\beta})_{\alpha,\beta\in[0,1]}$ is defined by

$$M_{\alpha,\beta}(x,y) := \begin{cases} x^{1-\alpha}y & x^\alpha \geq y^\beta \\ xy^{1-\beta} & x^\alpha < y^\beta \end{cases}$$

and contains Π ($\alpha = 0$ or $\beta = 0$) as well as M ($\alpha = \beta = 1$). It was shown in [42] that in case of $\alpha, \beta > 0$

$$\zeta_1(M_{\alpha,\beta}) = 3\alpha(1-\alpha)^z + \frac{6}{\beta}\frac{1-(1-\alpha)^z}{z} - \frac{6}{\beta}\frac{1-(1-\alpha)^{z+1}}{z+1}$$

holds whereby z is defined as $z = \frac{1}{\alpha} + \frac{2}{\beta} - 1$.

To test the performance of ζ_1^n for different MO-copulas we considered

$$(\alpha, \beta) \in \{(1, 0), (1, 1), (0.2, 0.7), (0.3, 1), (1, 0.7), (0.5, 0.5)\}$$

and generated samples of size $n \in \{10, 50, 100, 500, 1.000, 5.000, 10.000\}$ (see Figures 5, 6, 7, 8, 9, 10 (left panel)) and calculated the empirical N checkerboard copula $CB_{N(n)}(E_n)$ as well as $\zeta_1(CB_{N(n)}(E_n))$. These steps were repeated $R = 1.000$ times, the obtained results are depicted as boxplots in Figures 5, 6, 7, 8, 9, 10 (right panel).

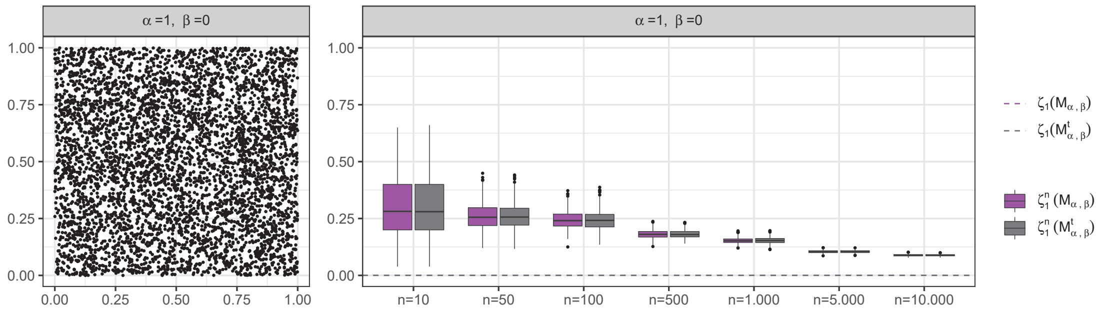

Figure 5: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M_{\alpha,\beta}^t)$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M_{\alpha,\beta}^t)$.

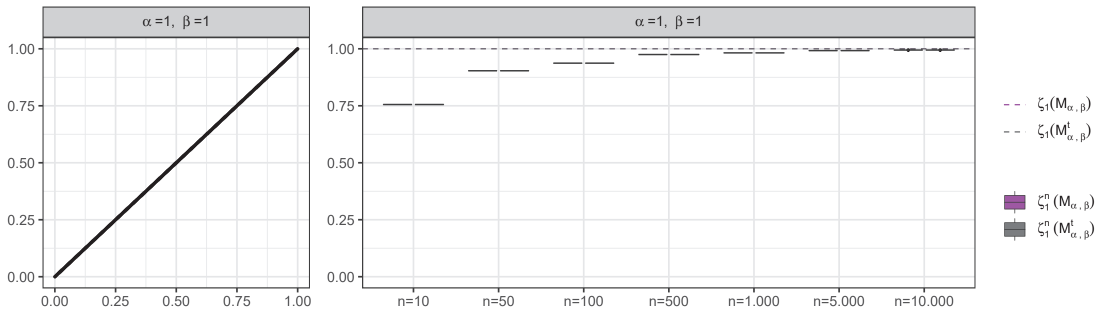

Figure 6: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M_{\alpha,\beta}^t)$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M_{\alpha,\beta}^t)$.

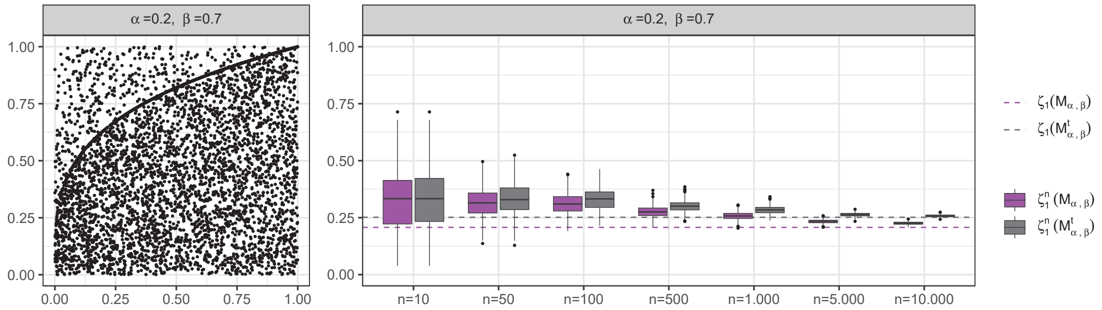

Figure 7: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M_{\alpha,\beta}^t)$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M_{\alpha,\beta}^t)$.

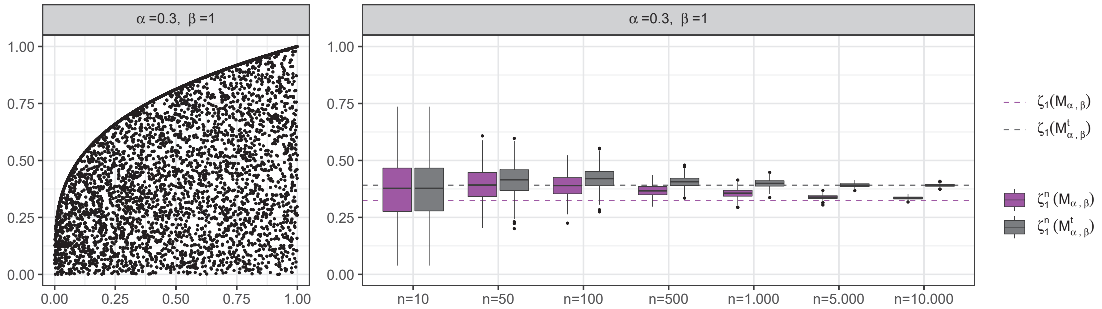

Figure 8: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M_{\alpha,\beta}^t)$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M_{\alpha,\beta}^t)$.

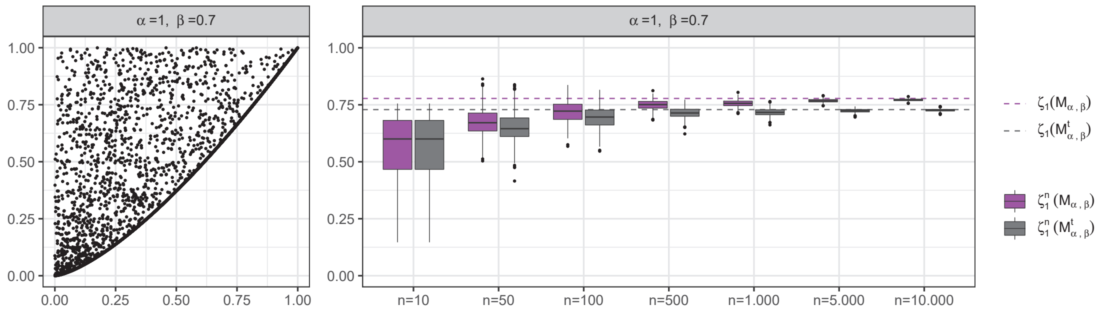

Figure 9: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M_{\alpha,\beta}^t)$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M_{\alpha,\beta}^t)$.

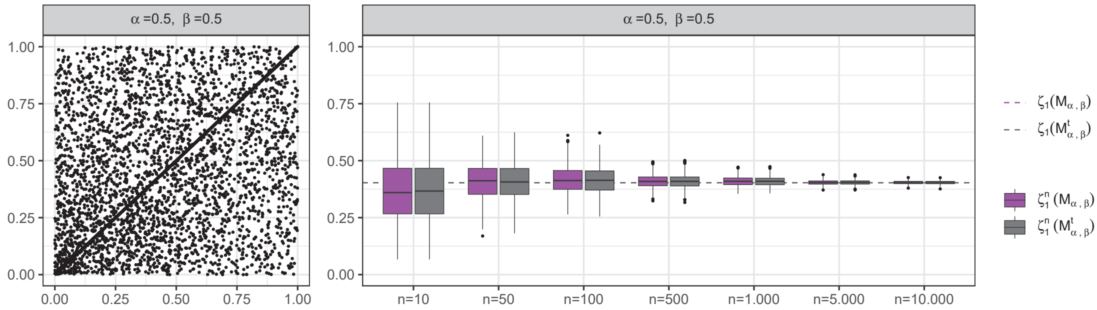

Figure 10: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(M_{\alpha,\beta})$ (magenta) and $\zeta_1^n(M^t_{\alpha,\beta})$ (gray). The dashed lines depict $\zeta_1(M_{\alpha,\beta})$ and $\zeta_1(M^t_{\alpha,\beta})$.

4.2. Farlie-Gumbel-Morgenstern family

The Farlie-Gumbel-Morgenstern family $(G_\theta)_{\theta\in[-1,1]}$ is defined by

$$G_\theta(x,y) := xy + \theta xy(1-x)(1-y).$$

According to [42] $\zeta_1(G_\theta)$ is given by

$$\zeta_1(G_\theta) = \frac{|\theta|}{4}.$$

Considering $\theta \in \{-1,-0.5\}$ and proceeding analogously as with MO-copulas before yields the results depicted in Figures 11, 12 (samples, left panel) and Figures 11, 12 (boxplots, right panel).

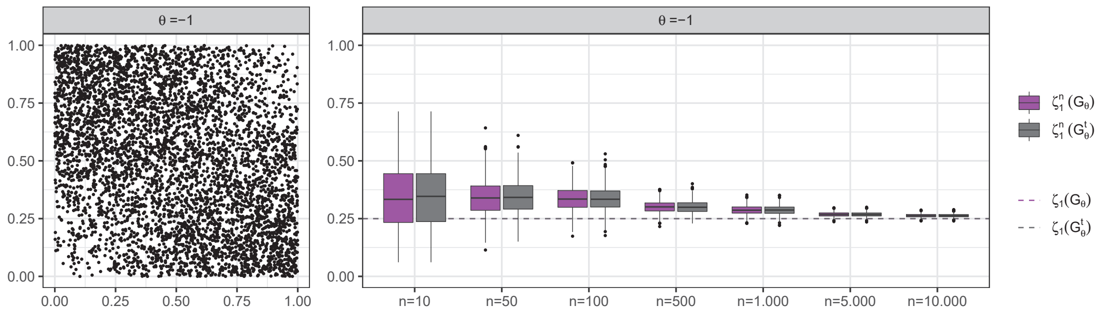

Figure 11: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(G_\theta)$ (magenta) and $\zeta_1^n(G^t_\theta)$ (gray). The dashed lines depict the $\zeta_1(G_\theta) = \zeta_1(G^t_\theta)$.

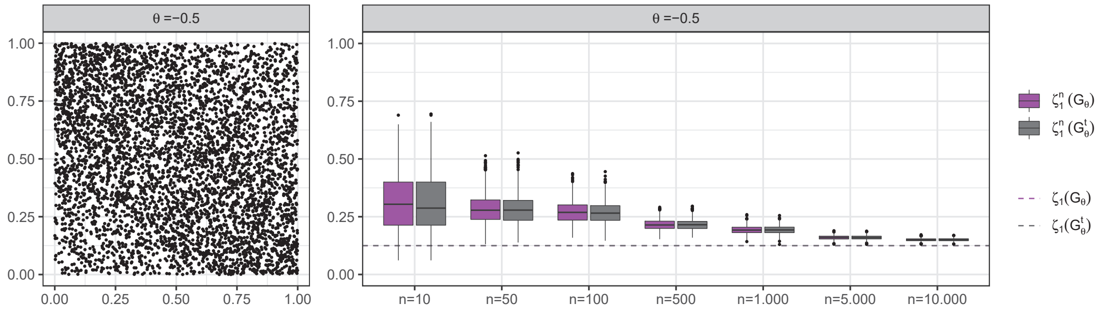

Figure 12: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(G_\theta)$ (magenta) and $\zeta_1^n(G_\theta^t)$ (gray). The dashed lines depict $\zeta_1(G_\theta) = \zeta_1(G_\theta^t)$.

4.3. Completely dependent copulas with high dependence asymmetry

To test the performance of our estimator in a situation where all classical measures fail we consider the completely dependent copula A_{h_a} for $h(x) = ax(mod1)$ and $a \in \{5, 10, 50\}$. Samples of size 5000 are depicted in Figures 13, 14, 15 (left panel). Notice that in these cases we have $\zeta_1(A_{h_a}) = 1$ whereas according to [42] we have

$$\lim_{a\to\infty} \zeta_1(A_{h_a}^t) = 0.$$

For every $a \in \{5, 10, 50\}$ we again generated samples of size

$$n \in \{10, 50, 100, 500, 1.000, 5.000, 10.000\},$$

calculated the empirical N checkerboard $CB_{N(n)}(E_n)$ as well as the values $\zeta_1(CB_{N(n)}(E_n))$ and $\zeta_1(CB_{N(n)}(E_n^t))$. These steps were repeated $R = 1.000$ times. The obtained results are depicted as boxplots in Figures 13, 14, 15 (right panel). Not surprisingly, the parameter a has a big influence on the precision of the estimator, the greater a the longer it takes our estimator to pick up the existing dependence asymmetry. Notice that, however, classical dependence measures like Schweizer and Wolff's σ [38] are not capable of detecting this asymmetry as in this case both $d_\infty(A_{h_a}, \Pi)$ and $d_\infty(A_{h_a}^t, \Pi)$ are very small since

$$\lim_{a\to\infty} d_\infty(A_{h_a}, \Pi) = 0 = \lim_{a\to\infty} d_\infty(A_{h_a}^t, \Pi).$$

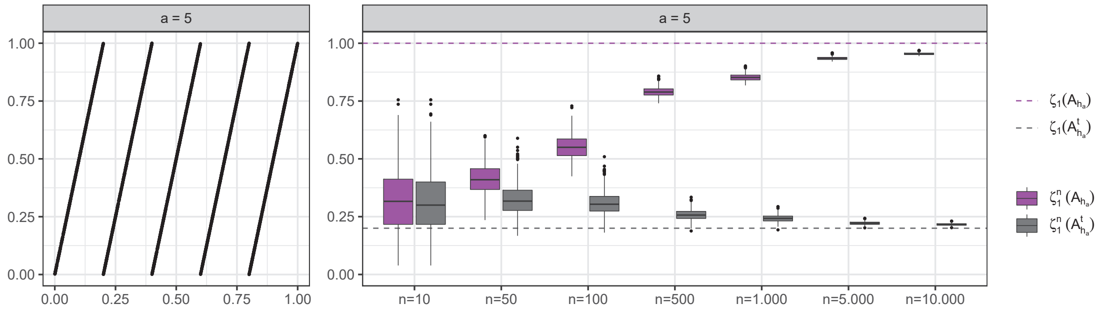

Figure 13: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(A_{h_a})$ (magenta) and $\zeta_1^n(A^t_{h_a})$ (gray). The dashed lines depict $\zeta_1(A_{h_a}) = \zeta_1(A^t_{h_a})$.

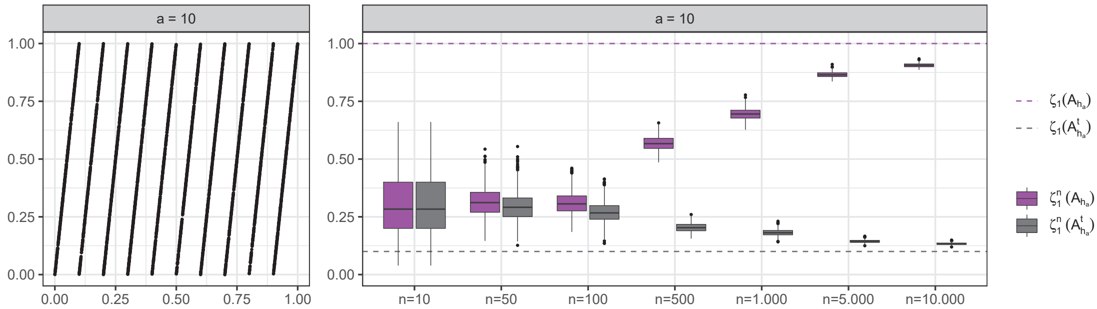

Figure 14: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(A_{h_a})$ (magenta) and $\zeta_1^n(A^t_{h_a})$ (gray). The dashed lines depict $\zeta_1(A_{h_a}) = \zeta_1(A^t_{h_a})$.

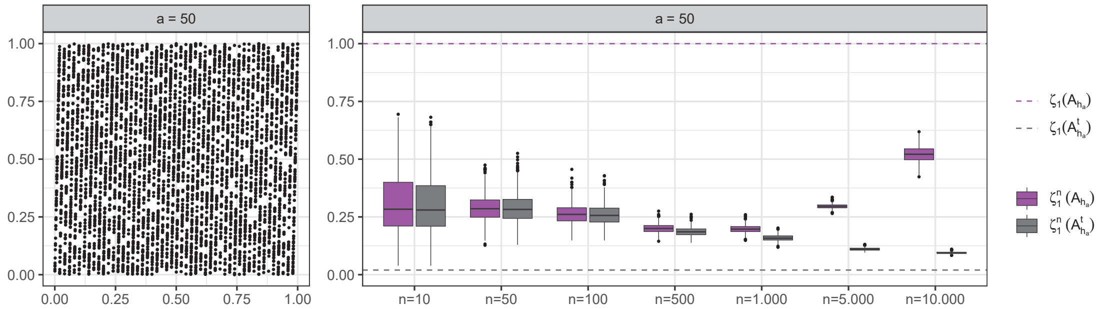

Figure 15: Sample of size 5000 (left panel) and boxplots (right panel) summarizing the 1.000 obtained estimates for $\zeta_1^n(A_{h_a})$ (magenta) and $\zeta_1^n(A^t_{h_a})$ (gray). The dashed lines depict $\zeta_1(A_{h_a}) = \zeta_1(A^t_{h_a})$.

5. Real Data Examples

5.1. Global climate

In the context of global warming and changes in precipitation regimes, information regarding past, present, and future climate at a local scale is required to assess the effects of climate on the environment. Data can be retrieved from databases that provide climate estimates in high geographic resolution. Commonly, estimates of temperature and precipitation are used to derive 19 bioclimatic variables that inform about climate features relevant to biological processes. Since these variables are based on a limited set of data, there is a strong underlying dependence structure between them, which, in particular, prevents their independent use in statistical models. We retrieved bioclimatic variables for $n = 1862$ locations homogenously distributed over the global landmass from CHELSA [20, 21] to quantify directional dependence and (dependence) asymmetry between pairs of variables. As expected, knowing one of the bioclimatic variables reduces the variability of the other ones, a fact that is reflected by the high dependence scores depicted in the boxplots in Figure 16. We can see

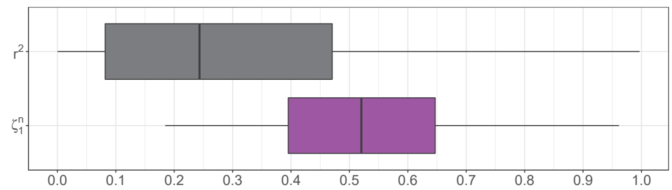

Figure 16: Values of ζ_1^n (magenta) vs. values of the squared Pearson correlation r^2 (gray) of all $342 = 19 \cdot 18$ pair-wise associations. Please note that ζ_1 is called q in the R-package *qad*.

that ζ_1^n shows generally higher values than the squared Pearson correlation r^2, which, considering that, firstly, ζ_1 detects asymmetries, that, secondly, ζ_1 is not restricted to any dependence structures like Pearson's r and that, thirdly, $\zeta_1^n \in [0, 1]$ and tends to 0 from above in the case of independence. Values of less than 0.25 for sample size $n = 300$ can therefore be interpreted as 'no dependence in this direction'. Table 1 summarizes the values of ζ_1^n and r^2 for all $19 \cdot 18$ pairs considered. Figure 17 depicts all obtained values for ζ_1^n and r^2 in terms of a heatmap. The heatmap for r^2 is symmetric, the one for ζ_1^n, however, is not.

Many of the pairs of bioclimatic variables exhibited dependence asymmetry (range of α_n : $[0.00, 0.23]$), including annual mean temperature (AMT) vs. annual precipitation (AP) (see Figure 18a) that would go unnoticed if only Pearson or Spearman correlation would be considered.

	$[0, \frac{1}{6}]$	$(\frac{1}{6}, \frac{2}{6}]$	$(\frac{2}{6}, \frac{3}{6}]$	$(\frac{3}{6}, \frac{4}{6}]$	$(\frac{4}{6}, \frac{5}{6}]$	$(\frac{5}{6}, 1]$
ζ_1^n :	0	45	109	109	55	24
r^2 :	138	74	48	26	24	32

Table 1: Frequency table for the values of ζ_1^n and the values of r^2 for the 19 bioclimatic variables.

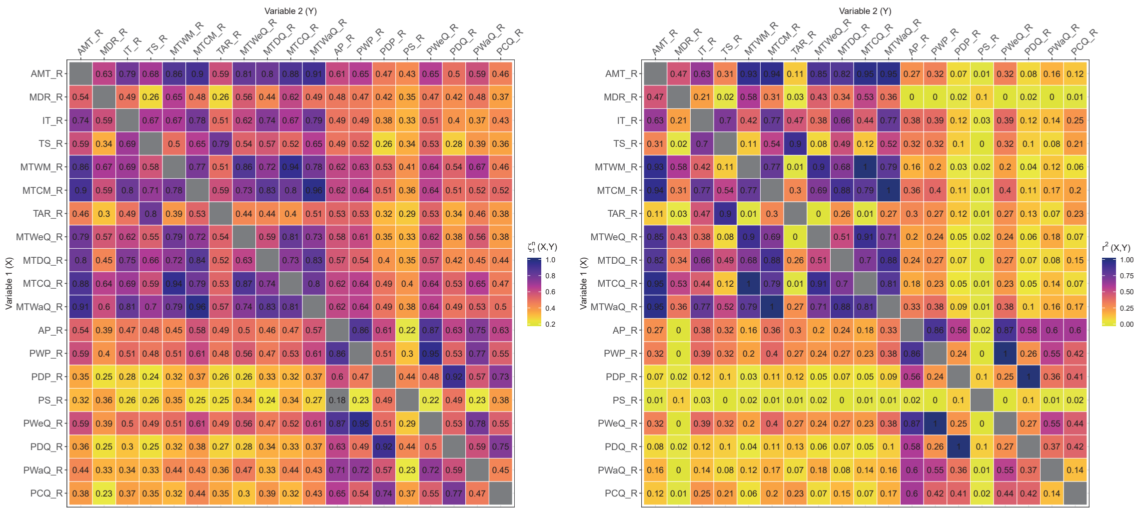

Figure 17: Pairwise values for ζ_1^n (left panel) and r^2 (right panel) and the 19 bioclimatic variables measured at $n = 1862$ locations homogeneously distributed over the global landmass. Please note that ζ_1 is called q in the R-package *qad*. Abbreviations: AMT: Annual Mean Temperature; MDR: Mean Diurnal Range; IT: Isothermality; TS: Temperature Seasonality; MTWM: Max Temperature of Warmest Month; MTCM: Min Temperature of Coldest Month; TAR: Temperature Annual Range; MTWeQ: Mean Temperature of Wettest Quarter; MTDQ: Mean Temperature of Driest Quarter; MTCQ: Mean Temperature of Coldest Quarter; MTWaQ: Mean Temperature of Warmest Quarter; AP: Annual Precipitation; PWP: Precipitation of Wettest Month; PDP: Precipitation of Driest Month; PS: Precipitation Seasonality; PWeQ: Precipitation of Wettest Quarter; PDQ: Precipitation of Driest Quarter; PWaQ: Precipitation of Warmest Quarter; PCQ: Precipitation of Coldest Quarter.

Sticking to AMT vs. AP, Figure 18 illustrates Steps 1-3 in the calculation of the estimator ζ_1 as described in the previous section. In this case the resolution was $N = \lfloor n^{0.5} \rfloor = 26$. For each vertical or horizontal strip, the conditional mass distribution of $CB_N(E_n)$ sums up to 1 and can also be used to predict the Y-value given the value of X and vice versa. In this concrete example annual precipitation can be better predicted by annual mean temperature ($\zeta_1^n(AMT, AP) = 0.61$) than vice versa ($\zeta_1^n(AP, AMT) = 0.54, \alpha_n = 0.08$).

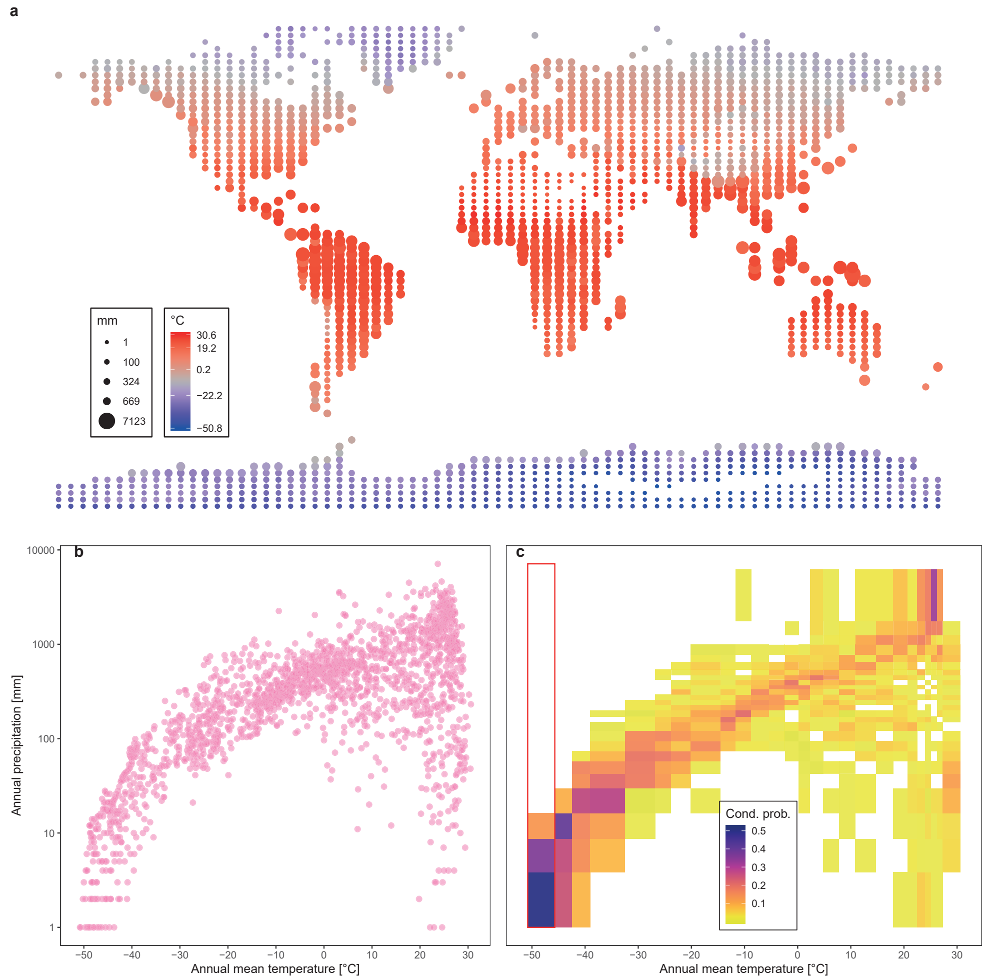

Figure 18: Association between the annual mean temperature AMT $[^{\circ}C]$ and annual precipitation AP $[mm]$. (a) Grid of $n = 1862$ locations homogenously distributed over the global landmass for which bioclimatic variables were retrieved. AMT is color coded and AP is proportional to the point size. (b) The knowledge of AMT strongly improves the predictability of AP ($\zeta_1^n = 0.61$), whereas the predictability is weaker in the opposite direction ($\zeta_1^n = 0.54$; $\alpha_n = 0.08$). (c) The association between AMT and AP visualized as a two-dimensional empirical checkerboard distribution (retransformation of $CB_N(E_n)$ via the corresponding marginal distributions). $CB_N(E_n)$ can also be used for predicting the value of Y given the value of X. For instance, in an area with an AMT of $-50.8^{\circ}C$ (first column), the estimated probability of AP lying between 1 mm and 6 mm is 0.516, between 6 mm and 14 mm is 0.358, and between 14 mm and 26 mm is 0.126 (red frame). Please note that ζ_1 is called q in the R-package *qad*.

5.2. World development indicators

The World Bank provides data sets on national development indicators related to, e.g., the economy, education, health, and infrastructure of states (World Development Indicators WDI, The World Bank). We explored the WDI data set published for 2015 to identify indicator pairs that feature strong dependence asymmetry and thus stand out from other indicator pairs. The indicators were restricted to those that were available for at least 100 countries and to those WDIs not having too many ties ($\geq 75\%$ unique entries) which resulted in a total of 412 WDIs of $n = 179$ countries to be included in the analysis. Notice that this approach differs from previously applied strategies for identifying interesting pairs of variables in large data sets, where pairs with a strong dependence (linear or non-linear) were selected [34]. Figure 19 depicts the distribution of the obtained asymmetry values α_n.

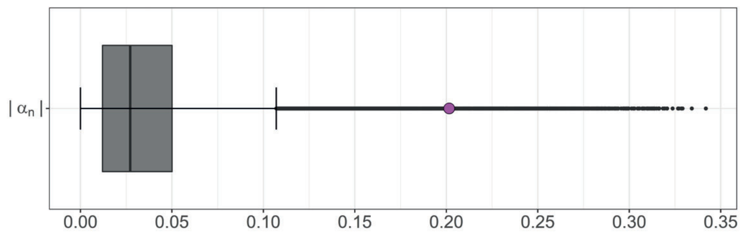

Figure 19: Boxplot, depicting the distribution of $|\alpha_n(X, Y)|$ for the 412 selected WDIs. The magenta colored point depicts the asymmetry value for the variables birth rate and death rate.

As concrete example we selected birth rate as X and death rate as Y (see Figure 20) and obtained a strong dependence asymmetry: $\zeta_1^n(X, Y) = 0.53, \zeta_1^n(Y, X) = 0.32$, implying $\alpha_n(X, Y) = 0.2$. Further data exploration revealed that the gross domestic product per capita (GDP) was a potential underlying factor for this dependence asymmetry. The GDP seems strongly and symmetrically associated with the birth rate ($\zeta_1^n(X, GDP) = 0.68, \zeta_1^n(GDP, X) = 0.65$) and weakly associated to the death rate ($\zeta_1^n(Y, GDP) = 0.29, \zeta_1^n(GDP, Y) = 0.27$). Generally, in countries with a GDP below the median GDP of all countries, the birth rate and death rate are moderately concordant (Spearman's $\rho = 0.38$), in countries with a GDP above the median GDP, we found a high discordance (Spearman's $\rho = -0.53$). The relationship between the economy and population growth has previously been discussed in [15, 25] and seems to have important implications for societies. We conclude that ζ_1^n strongly facilitates the detection of interesting associations in large data sets

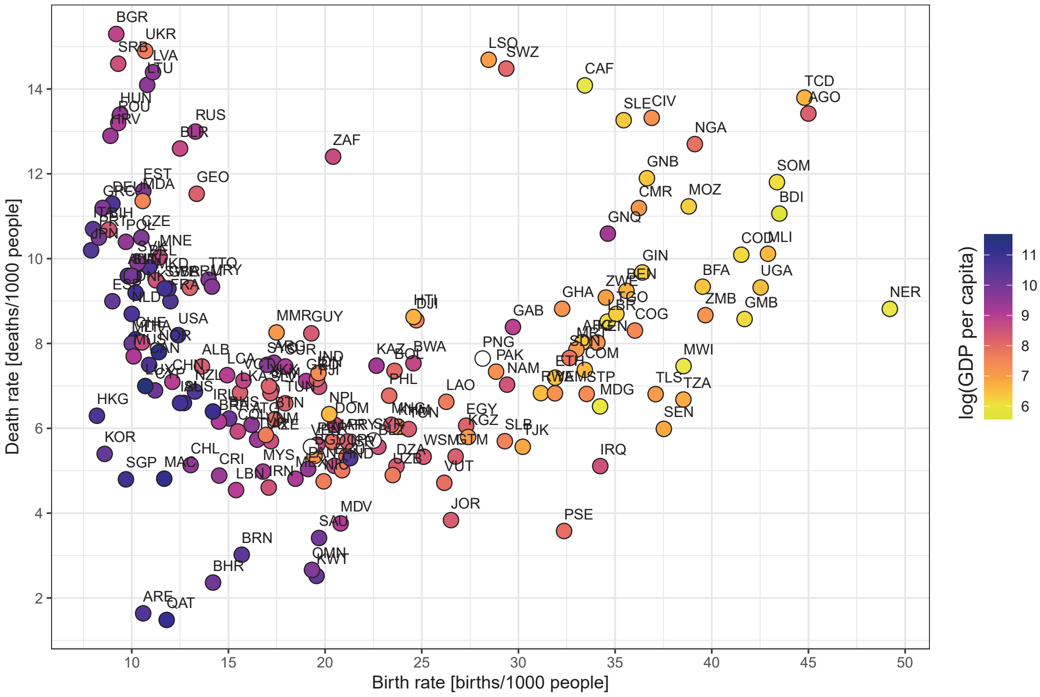

Figure 20: Association between birth rate (X) and death rate (Y). The gross domestic product per capita (GDP) of each country is color coded. The death rate is well predictable by the birth rate ($\zeta_1^n(X,Y) = 0.53$) whereas the information gain is smaller in the other direction ($\zeta_1^n(Y,X) = 0.32$). Please note that ζ_1 is called q in the R-package *qad*.

and thus contributes to a more thorough exploration of data sets and the improved detection of meaningful patterns compared to traditional methods.

6. Conclusion and future work

In this paper we constructed an estimator for the copula-based (hence scale-free), non-parametric dependence measure ζ_1 and showed that this estimator is strongly consistent in full generality and, additionally, exhibits a good small sample performance. The estimator was implemented in the R-package *qad* to allow for a straightforward application throughout all disciplines. As one of the next steps we will apply the developed estimator in order to quantify the influence of species on other species.

Moreover, first simulation studies showed that the estimator ζ_1^n can also be used successfully for constructing alternative tests for independence - a power study of the resulting hypothesis test will be executed in the near future. Figure 21 depicts probability histograms of 10.000 values of $\sqrt{n}\,\zeta_1^n$ for samples $(x_1, y_1), \ldots, (x_n, y_n)$ of size $n \in \{10, 100, 1.000\}$ drawn from the product copula Π (describing independence). Since for other underlying copulas A the resulting histograms insinuate normality too we will tackle a formal proof for asymptotic normality of $\sqrt{n}(\zeta_1^n - \zeta_1)$.

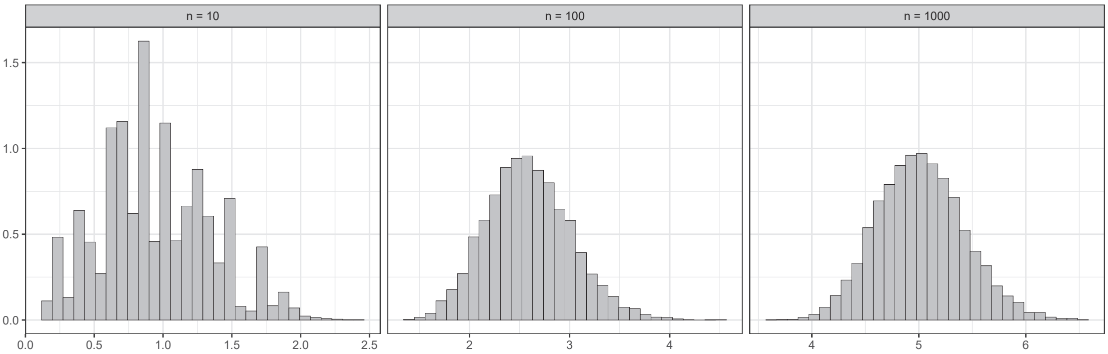

Figure 21: Histograms summarizing 10.000 values of $\sqrt{n}\zeta_1^n$ for samples $(x_1, y_1), \ldots, (x_n, y_n)$ of size $n \in \{10, 100, 1.000\}$ drawn from the product copula Π.

Additionally, we want to prove or disprove the conjecture that ζ_1^n is also strongly consistent for $s \in \left[\frac{1}{2}, 1\right)$. After answering the afore-mentioned questions the natural next step is to extend ζ_1^n to the general multivariate setting estimating the influence of several variables on a target variable.

Acknowledgement

The authors gratefully acknowledge the support of the Austrian FWF START project Y1102 'Successional Generation of Functional Multidiversity'. Moreover, the third author gratefully acknowledges the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

Appendix A.

The following example shows that the maximal information coefficient (MIC) can be maximal also outside the situation of complete dependence.

Example Appendix A.1. Recall that MIC is defined as (see [34, 36, 35])

$$MIC(X,Y) = \sup_G \frac{MI((X,Y)|_G)}{\log\|G\|}, \tag{A.1}$$

where MI denotes the mutual information, G defines an arbitrary grid and $\|G\|$ denotes the minimum of the number of rows and columns of G. If the random vector (X,Y) has as joint distribution function the ordinal sum Σ_Π of Π (see Figure A.22) then we can obviously find a 2×2 grid for which $MI(X,Y)|_G = log(2)$. In other words $MIC(X,Y)$ attains its maximum value 1 although Y is not completely dependent on X.

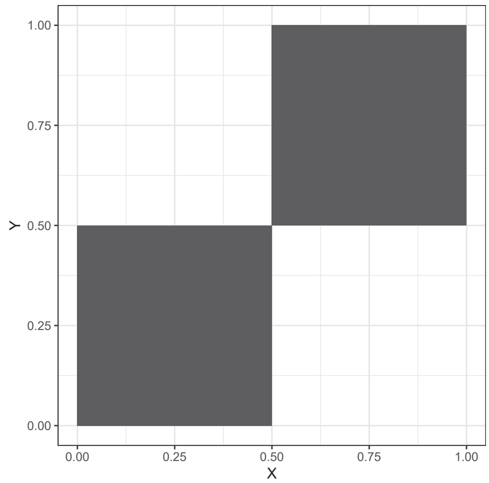

Figure A.22: Density of the copula Σ_Π considered in Example Appendix A.1. For $(X,Y)\sim\Sigma_\Pi$ we have $MIC(X,Y) = MIC(Y,X) = 1$ although Σ_Π is not completely dependent. ζ_1 assigns this dependence structure the values $\zeta_1(\Sigma_\Pi) = \zeta_1(\Sigma_\Pi^t) = \frac{3}{4}$.

The next example shows the existence of copulas A_N and B_N for which inequality (5) in Lemma 3.8 is sharp.

Example Appendix A.2. Considering the checkerboard copulas depicted in Figure (A.23) for $N = 8$ it is straightforward to verify that in this case we have equality in ineq. (5). Examples for other values of N can be constructed in the same manner. Figure A.24 depicts the map $y \mapsto \Phi_{A_N,B_N}$ (left panel) and the map $(x,y)\mapsto |A_N(x,y) - B_N(x,y)|$ (right panel) for the case $N = 8$.

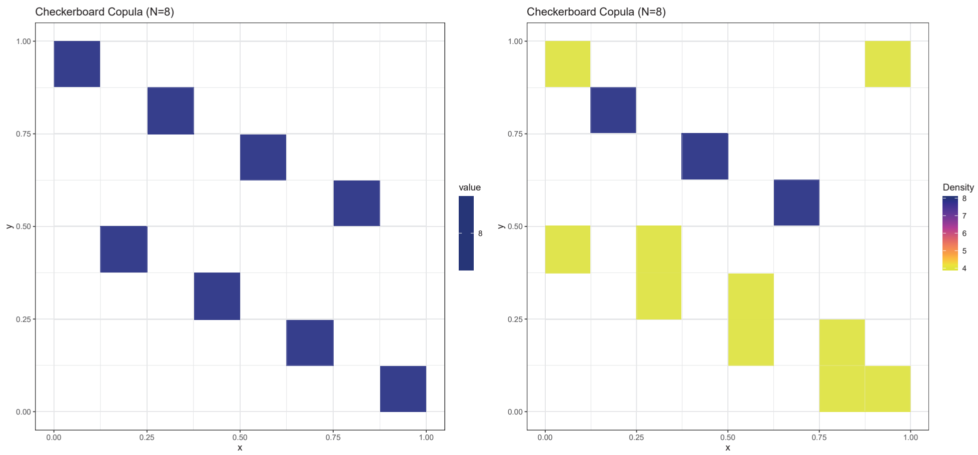

Figure A.23: Checkerboard copulas A_N (left) and B_N (right) for $N = 8$ as considered in Example Appendix A.2.

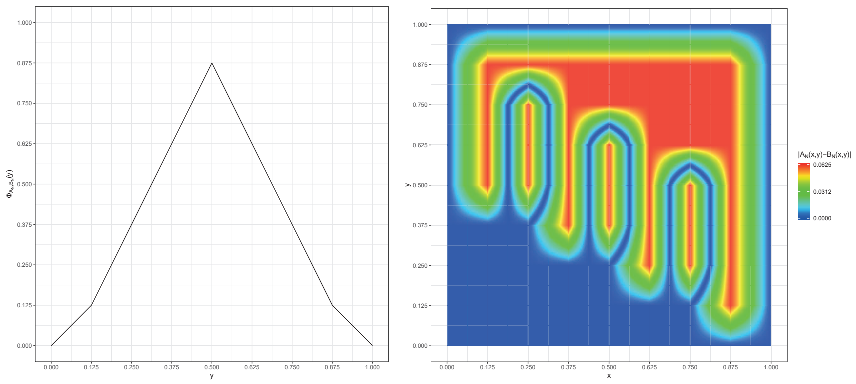

Figure A.24: The maps $y \mapsto \Phi_{A_N,B_N}(y)$ (left panel) and $(x, y) \mapsto |A_N(x, y) - B_N(x, y)|$ (right panel) for A_N and B_N from Figure A.23 (top). In this case we have $D_\infty(A_N, B_N) = 7/8$ and $d_\infty(A_N, B_N) = 1/16$.

Since the choice of the parameter s in Theorem 3.13 is key for the performance of the estimator ζ_1^n, we conducted a small simulation study illustrating the influence of s and justifying the choice $s = 1/2$ in the R-package *qad*. Doing so we calculate ζ_1^n in two extreme settings - that of complete dependence and that of independence.

Example Appendix A.3. Suppose that $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a random sample of size n from $(X, Y) \sim A_h \in \mathcal{C}$, whereby A_h is the completely dependent copula induced by the

λ-preserving transformation $h : [0,1] \to [0,1]$ given by

$$h(x) = (1-2x)\mathbb{1}_{\left(0,\frac{1}{2}\right]}(x) + (2x-1)\mathbb{1}_{\left(\frac{1}{2},1\right]}(x).$$

The support of μ_{A_h} coincides with the graph of V (see Figure A.25). To illustrate the performance

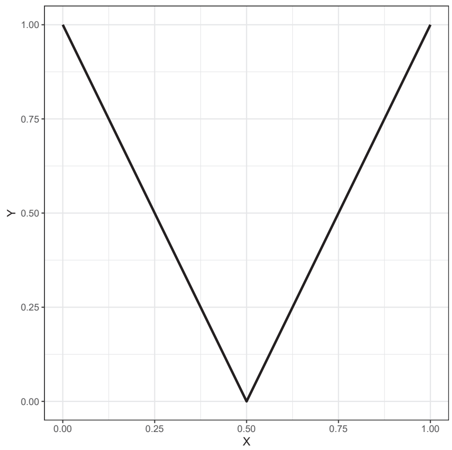

Figure A.25: Support of μ_{A_h} as considered in Example Appendix A.3.

of ζ_1^n for different choices of the parameter s we considered

$$s \in \left\{ \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \ldots, \frac{7}{8} \right\},$$

generated samples of size $n \in \{50, 100, 500, 1.000\}$, considered $N(n) = \lfloor n^s \rfloor$, and calculated

$$\left|\zeta_1(CB_{N(n)}(E_n)) - \zeta_1(A_h)\right| \text{ and } \left|\zeta_1(CB_{N(n)}(E_n^t)) - \zeta_1(A_h^t)\right|.$$

These steps were repeated $R = 100$ times, the obtained results are depicted as boxplots in Figures A.26 and A.27. As mentioned in Remark 3.14, the larger s the better the performance of ζ_1^n in the case of A_h. In the case of A_h^t considering $s = \frac{1}{2}$ yielded the best results.

Example Appendix A.4. Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a sample of size n from $(X, Y) \sim \Pi \in \mathcal{C}$. Proceeding analogously as in Example Appendix A.3 yields the results depicted in Figure A.28. Again in accordance with Remark 3.14, the smaller s the better the performance of ζ_1^n.

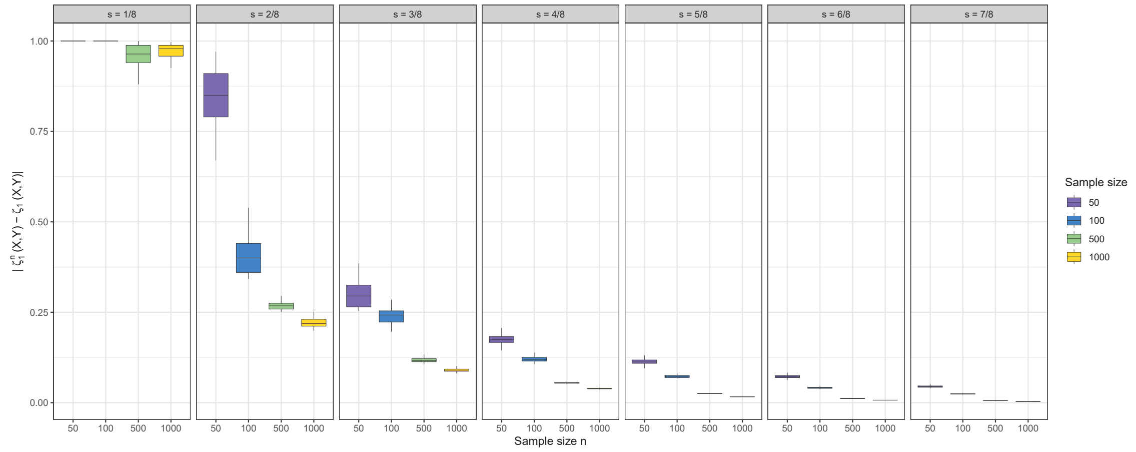

Figure A.26: Boxplots summarizing the 100 obtained values for $\left|\zeta_1(CB_{N(n)}(E_n)) - \zeta_1(A_h)\right|$ according to Example Appendix A.3.

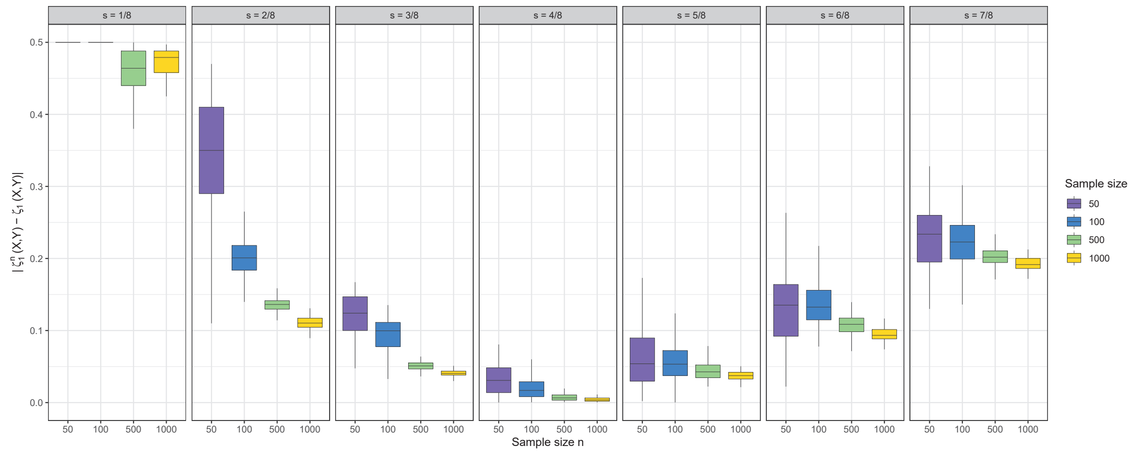

Figure A.27: Boxplots summarizing the 100 obtained values for $\left|\zeta_1(CB_{N(n)}(E_n^t)) - \zeta_1(A_h^t)\right|$ according to Example Appendix A.3.

References

[1] M. M. Al-Sadoon. Testing subspace granger causality. *Econometrics and Statistics*, 9:42 – 61, 2019.

[2] P. Amblard and O. Michel. On directed information theory and granger causality graphs. *Journal of Computational Neuroscience*, 30:7–16, 2010.

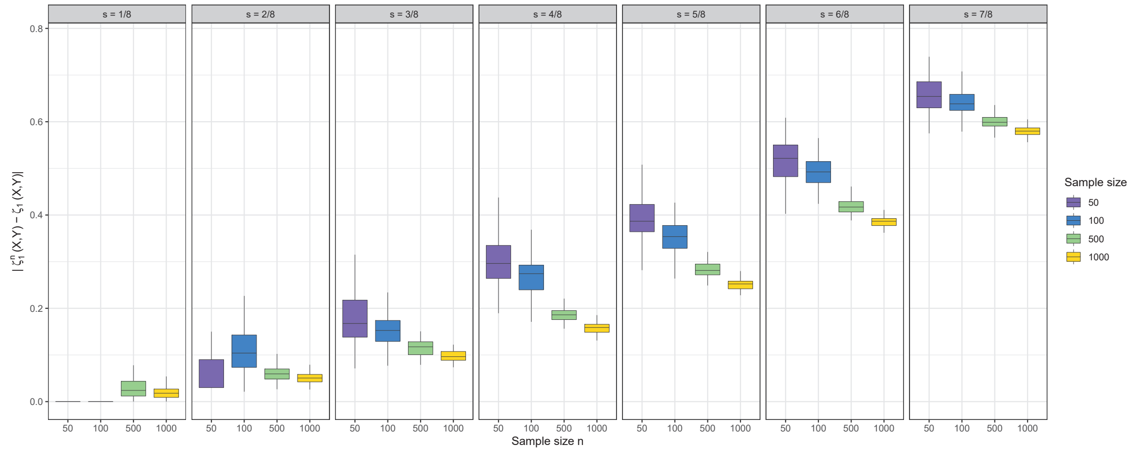

Figure A.28: Boxplots summarizing the 100 obtained values for $\left|\zeta_1(CB_{N(n)}(E_n)) - \zeta_1(\Pi)\right|$ according to Example Appendix A.4.

[3] A. Bárdossy and S. Hörning. Process-driven direction-dependent asymmetry: Identification and quantification of directional dependence in spatial fields. *Mathematical Geosciences*, 49:871–891, 2017.

[4] Y. Chang, Y. Li, A. Ding, and J. Dy. A robust-equitable copula dependence measure for feature selection. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 84–92, Cadiz, Spain, 09–11 May 2016. PMLR.

[5] A. R. Coenen and J. S. Weitz. Limitations of correlation-based inference in complex virus-microbe communities. *mSystems*, 3(4), 2018.

[6] C. Cottin and D. Pfeifer. From Bernstein polynomials to Bernstein copulas. *Journal of Applied Functional Analysis*, 9:277–288, 05 2014.

[7] P. Deheuvels. La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Académie Royale de Belgique, Bulletin de la Classe des Sciences, 5e série*, 65(6):274–292, 1979.

[8] H. Dette, K. F. Siburg, and P. A. Stoimenov. A copula-based non-parametric measure of regression dependence. *Scandinavian Journal of Statistics*, 40(1):21–41, 2013.

[9] F. Durante, J. Fernández-Sánchez, J. J. Quesada-Molina, and M. Úbeda-Flores. Convergence results for patchwork copulas. *European Journal of Operational Research*, 247(2):525 – 531, 2015.

[10] F. Durante and C. Sempi. *Principles of copula theory.* CRC Press, Hoboken, NJ, 2015.

[11] J. Fernández Sánchez and W. Trutschnig. Conditioning-based metrics on the space of multivariate copulas and their interrelation with uniform and levelwise convergence and iterated function systems. *Journal of Theoretical Probability*, 28(4):1311–1336, 2015.

[12] C. Genest, I. Kojadinovic, and F. Durante. Introduction to the special topic on copula modeling. *Econometrics and Statistics*, 12:146 – 147, 2019.

[13] C. Genest, J. G. Nešlehová, and B. Rémillard. On the empirical multilinear copula process for count data. *Bernoulli*, 20(3):1344–1371, 2014.

[14] F. Griessenberger, R. R. Junker, and W. Trutschnig. *qad: Quantification of Asymmetric Dependence*, 2020. R package version 0.1.2.

[15] S. Harper. Economic and social implications of aging societies. *Science*, 346(6209):587–591, 2014.

[16] M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas*, 2020. R package version 0.999-20.

[17] P. Janssen, J. Swanepoel, and N. Veraverbeke. Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, 142:1189 – 1197, 05 2012.

[18] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, 2002.

[19] N. Kamnitui, J. Fernández-Sánchez, and W. Trutschnig. Maximum asymmetry of copulas revisited. *Dependence Modeling*, 6:47–62, 03 2018.

[20] D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. Zimmermann, H. Linder, and M. Kessler. Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 2017.

[21] D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. Zimmermann, H. Linder, and M. Kessler. Data from: Climatologies at high resolution for the earth's land surface areas, 2018. [Dataset].

[22] A. Klenke. *Wahrscheinlichkeitstheorie*. Springer-Lehrbuch Masterclass Series. Springer, Berlin Heidelberg, 2008.

[23] T. Kulpa. On approximation of copulas. *International Journal of Mathematics and Mathematical Sciences*, 22, 01 1999.

[24] X. Li, P. Mikusiński, and M. Taylor. Strong approximation of copulas. *Journal of Mathematical Analysis and Applications*, 225(2):608 – 623, 1998.

[25] J. M. Lindo. Aggregation and the estimated effects of economic conditions on health. *Journal of Health Economics*, 40:83 – 96, 2015.

[26] E. Linfoot. An informational measure of correlation. *Information and Control*, 1(1):85 – 89, 1957.

[27] J. Lloyd-Price, C. Arze, A. N. Ananthakrishnan, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569(7758):655–662, 2019.

[28] R. B. Nelsen. *An Introduction to Copulas (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[29] R. B. Nelson. Extremes of nonexchangeability. *Statistical Papers*, 48(2):329–336, Apr 2007.

[30] T. Okimoto. New evidence of asymmetric dependence structures in international equity markets. *The Journal of Financial and Quantitative Analysis*, 43(3):787–815, 2008.

[31] D. Pfeifer, D. Straßburger, and J. Philipps. Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas. *International ASTIN-Colloquium*, Helsinki, June 1-4, 2009.

[32] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2019.

[33] A. Rényi. On measures of dependence. *Acta Mathematica Hungarica*, 10(3-4):441–451, 1959.

[34] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524, 2011.

[35] D. N. Reshef, Y. A. Reshef, P. C. Sabeti, and M. Mitzenmacher. An empirical study of the maximal and total information coefficients and leading measures of dependence. *Ann. Appl. Stat.*, 12(1):123–155, 2018.

[36] Y. A. Reshef, D. N. Reshef, H. K. Finucane, P. C. Sabeti, and M. Mitzenmacher. Measuring dependence powerfully and equitably. *Journal of Machine Learning Research*, 17(211):1–63, 2016.

[37] A. Sancetta and S. Satchell. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(3):535–562, 2004.

[38] B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *Ann. Statist.*, 9(4):879–885, 1981.

[39] J. Segers, M. Sibuya, and H. Tsukahara. The empirical beta copula. *Journal of Multivariate Analysis*, 155:35 – 51, 2017.

[40] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8:229 – 231, 1959.

[41] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794, 2007.

[42] W. Trutschnig. On a strong metric on the space of copulas and its induced dependence measure. *Journal of Mathematical Analysis and Applications*, 384(2):690 – 705, 2011.

[43] Y. R. Wang and H. Huang. Review on statistical methods for gene network reconstruction using expression data. *Journal of Theoretical Biology*, 362:53 – 61, 2014.

[44] J. Yan. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21, 2007.