A bivariate copula capturing the dependence of a random variable and a random vector, its estimation and applications

Sebastian Fuchs
Universität Salzburg
Salzburg, Austria

Abstract
We define a bivariate copula that captures the scale-invariant extent of dependence of a single random variable $Y$ on a set of potential explanatory random variables $X_1, \ldots, X_d$. The copula itself contains the information whether $Y$ is completely dependent on $X_1, \ldots, X_d$, and whether $Y$ and $X_1, \ldots, X_d$ are independent. Evaluating this copula uniformly along the diagonal, i.e. calculating Spearman’s footrule, leads to the so-called 'simple measure of conditional dependence' recently introduced by Azadkia and Chatterjee [1]. On the other hand, evaluating this copula uniformly over the unit square, i.e. calculating Spearman’s rho, leads to a distribution-free coefficient of determination. Applying the techniques introduced in [1], we construct an estimate for this copula and show that this copula estimator is strongly consistent. Since, for $d = 1$, the copula under consideration coincides with the well-known Markov product of copulas, as by-product, we also obtain a strongly consistent copula estimator for the Markov product. A simulation study illustrates the small sample performance of the proposed estimator.

1 Introduction
Quantifying the degree of predictability or explainability of a single (continuous) random variable $Y$ using the information contained in a set of potential explanatory (continuous) random variables $X_1, \ldots, X_d$ is a long-studied problem. An index measuring such a degree should be capable of detecting perfect dependence also known as complete dependence (see [12, 20, 27]): $Y$ is said to be completely dependent on $X = (X_1, \ldots, X_d)$ if there exists a measurable function $f$ such that $Y = f(X)$ almost surely.
Regardless of whether one is interested in predictability or explainability with respect to a single or a set of explanatory variables, the concept of complete dependence is indispensable:

1. Quite recently, Azadkia and Chatterjee [1] introduced their so-called 'simple measure of conditional dependence' $T$ which is based on [3, 5] and quantifies the scale-invariant extent of dependence of $Y$ on $X$. As $T$ equals 1 if and only if $Y$ is completely dependent on $X$, and $T$ equals 0 if and only if $Y$ and $X$ are independent, it belongs to a class of indices capable of detecting complete dependence and independence which also include [7, 12, 14, 27].

2. Similar to the coefficient of determination in regression analysis, the Sobol index measures the proportion of the variance of $Y$ that is explained by the regression function $r(x) := E(Y|X = x)$. It is well-known that the proportion of the explained variance equals 1 if and only if $Y$ is completely dependent on $X$. 

\*E-mail Sebastian Fuchs: sebastian.fuchs@sbg.ac.at (Corresponding author)
If the random vector \((X, Y)\) has continuous marginal distribution functions \(F_i\) of \(X_i\), \(i \in \{1, \ldots, d\}\), and \(G\) of \(Y\), and connecting \((d + 1)\)-dimensional copula \(A\)

1. the index \(T\) can be represented in terms of the well-known measure of dependence Spearman’s footrule, and

2. the distribution-free version (in the following referred to as distribution-free \(R^2\)) of the variance of \(Y\) that is explained by the regression function \(r\) possesses a representation in terms of the quite popular bivariate measure of dependence Spearman’s rank correlation (Spearman’s rho) both evaluating a certain bivariate copula \(\psi(A)\) derived from \(A\). This bivariate copula \(\psi(A)\) turns out to be a generalization of the well-known Markov product of copulas (see [3, 6]) and can be interpreted as the distribution function of two random variables \(V := G(Y)\) and \(V' := G(V')\) such that \((X, Y)\) and \((X, Y')\) share the same distribution and are conditionally independent given \(X\). The copula \(\psi(A)\) itself contains the information

(i) whether \(Y\) is completely dependent on the \(d\)-dimensional random vector \(X\), and

(ii) whether \(Y\) and the \(d\)-dimensional random vector \(X\) are independent.

and hence provides information about the degree of predictability or explainability of \(Y\) given \(X\). From this perspective both initial \((d + 1)\)-dimensional problems - (1) quantifying the scale-invariant extent of dependence of \(Y\) on \(X\), and (2) evaluating the proportion of the variance of \(Y\) that is explained by the regression model - reduce to a 2-dimensional one.

For the bivariate copula \(\psi(A)\) we propose an estimator whose form is reminiscent of the empirical copula, but which is actually based on the graph-based so-called ‘simple measure of conditional dependence’ introduced in Azadkia and Chatterjee [1]. By applying the tools given in [1], we show that the copula estimator is strongly consistent from which strong consistency of the plug-in estimates of Spearman’s footrule (that is \(T\)) and Spearman’s rho (that is the distribution-free \(R^2\)) can be derived.

The rest of this contribution is organized as follows: Section 2 gathers preliminaries and notations that will be used throughout the paper. In Section 3 we formally define the transformation \(\psi\) mapping every \((d + 1)\)-dimensional copula \(A\) to an exchangeable bivariate one, and show that this bivariate copula allows to detect independence and complete dependence, is invariant with respect to a variety of transformations of the original dependence structure \(A\) and satisfies the so-called information gain inequality along the diagonal. In Section 4 we then provide an estimator for \(\psi(A)\), prove its strong consistency, and illustrate its small/moderate sample performance. The connection between \(T\) and Spearman’s footrule of \(\psi(A)\) and between the distribution-free \(R^2\) and Spearman’s rho of \(\psi(A)\) is presented in Section 5. This way of looking at \(T\) and \(R^2\) allows the introduced copula estimate to be used as plug-in estimator for \(T\) (via Spearman’s footrule) and as plug-in estimator for \(R^2\) (via Spearman’s rho). Finally, we present a real data example to illustrate our methodology (Section 6).

## 2 Preliminaries

Throughout this paper we will write \(I := [0,1]\) and let \(d \geq 1\) be an integer which will be kept fixed. Bold symbols will be used to denote vectors, e.g., \(x = (x_1, \ldots, x_d) \in \mathbb{R}^d\). The \(d\)-dimensional Lebesgue measure will be denoted by \(\lambda^d\), in case of \(d = 1\) we simply write \(\lambda\).
We will let $C^{d+1}$ denote the family of all $(d+1)$-dimensional copulas, $M$ will denote the comonotonicity copula, $\Pi$ the independence copula and, for $d=1$, $W$ will denote the countermonotonicity copula (we omit the index indicating the dimension since no confusion will arise). For every $A \in C^{d+1}$ the corresponding probability measure will be denoted by $\mu_A$, i.e. $\mu_A([0,u] \times [0,v]) = A(u,v)$ for all $(u,v) \in I^d \times I$; for more background on copulas and copula measures we refer to [6, 24]. For every metric space $(\Omega, \delta)$ the Borel $\sigma$-field on $\Omega$ will be denoted by $B(\Omega)$.

For a copula $A \in C^{d+1}$ we denote by $A^K$ the $|K|$-dimensional marginal of $A$ with respect to the coordinates in $K \subseteq \{1, \ldots, d+1\}$, and for the $k$-dimensional marginal of $A$ ($k \geq 2$) with respect to the first $k \in \{1, \ldots, d\}$ coordinates we simply write $A^{1:k} := A^{(1,\ldots,k)}$.

In what follows Markov kernels will play a decisive role: A Markov kernel from $\mathbb{R}^d$ to $B(\mathbb{R})$ is a mapping $K : \mathbb{R}^d \times B(\mathbb{R}) \to I$ such that for every fixed $F \in B(\mathbb{R})$ the mapping $x \mapsto K(x,F)$ is measurable and for every fixed $x \in \mathbb{R}^d$ the mapping $F \mapsto K(x,F)$ is a probability measure. Given a real-valued random variable $Y$ and a real-valued $d$-dimensional random vector $X$ on a probability space $(\Omega, A, P)$ we say that a Markov kernel $K$ is a regular conditional distribution of $Y$ given $X$ if

$$K(X(\omega), F) = E(1_F \circ Y \mid X)(\omega)$$

holds $P$-almost surely for every $F \in B(\mathbb{R})$. It is well-known that for each random vector $(X,Y)$ a regular conditional distribution $K(\ldots)$ of $Y$ given $X$ always exists and is unique for $P^X$-a.e. $x \in \mathbb{R}^d$, where $P^X$ denotes the push-forward of $P$ under $X$. If $(X,Y)$ has distribution function $H$ (in which case we will also write $(X,Y) \sim H$) and let $\mu_H$ denote the corresponding probability measure on $B(\mathbb{R}^{d+1}))$ we will let $K_H$ (a version of) the regular conditional distribution of $Y$ given $X$ and simply refer to it as Markov kernel of $H$.

If $A \in C^{d+1}$ is a copula then we will consider the Markov kernel of $A$ (with respect to the first $d$ coordinates) automatically as mapping $K_A : I^d \times B(I) \to I$. Defining the $u$-section of a set $G \in B(I^{d+1})$ as $G_u := \{v \in I : (u,v) \in G\}$ the so-called disintegration theorem yields

$$\mu_A(G) = \int_{I^d} K_A(u, G_u) \, d\mu_{A^{1:d}}(u)$$

(1)

so, in particular, we have

$$\mu_A(I^d \times F) = \int_{I^d} K_A(u, F) \, d\mu_{A^{1:d}}(u) = \lambda(F)$$

For more background on conditional expectation and general disintegration we refer to [15, 19]; for more information on Markov kernels in the context of copulas we refer to [6, 18, 23].

We say that $Y$ is completely dependent on $X$ if there exists a measurable function $f$ such that $P(Y = f(X)) = 1$. If the marginal distribution functions $F_i$ of $X_i$, $i \in \{1,\ldots,d\}$, and $G$ of $Y$ are continuous with connecting copula $A$, then (denoting $U_i := F_i(X_i)$, $i \in \{1,\ldots,d\}$ and $V := G(Y)$) the following statements are equivalent (see [12]):

(a) $Y$ is completely dependent on $X$.

(b) There exists a $\mu_{A^{1:d}} - \lambda$ preserving transformation $h : I^d \to I$ (i.e. $(\mu_{A^{1:d}})^h = \lambda$) such that $V = h(U)$ a.s.

(c) There exists a $\mu_{A^{1:d}} - \lambda$ preserving transformation $h : I^d \to I$ such that $K(u,F) := 1_F(h(u))$ is a regular conditional distribution of $A$.

For more properties of complete dependence we refer to [20] as well as to [7] and the references therein.
3 A bivariate copula capturing independence and complete dependence between a random variable and a random vector

Consider a \((d+1)\)-dimensional random vector \((X, Y)\) with continuous marginal distribution functions \(F_i\) of \(X_i\), \(i \in \{1, \ldots, d\}\), and \(G\) of \(Y\) and connecting copula \(A\). Then (denoting \(U_i := F_i(X_i), i \in \{1, \ldots, d\}\), and \(V := G(Y)\)) we have \((U, V) \sim A\).

In what follows we construct an exchangeable bivariate copula that allows to detect

(i) whether \(Y\) and \(X\) (or, equivalently, \(V\) and \(U\)) are independent.

(ii) whether \(Y\) is completely dependent on \(X\) (or, equivalently, \(V\) is completely dependent on \(U\)).

To this end, we extend the random vector \((U, V)\) and consider the \((d+2)\)-dimensional random vector \((U, V, V')\) with uniform univariate marginals such that \((U, V) \sim A\) and \((U, V') \sim A\) and assume that \(V\) and \(V'\) are conditionally independent given \(U = u\). Then, using disintegration, the distribution function of \((V, V')\) is a copula and equals

\[
P(V \leq s, V' \leq t) = \frac{E(P(V \leq s, V' \leq t \mid U))}{E(P(V \leq s \mid U) P(V' \leq t \mid U))}
= \int_{\mathbb{R}^d} P(V \leq s \mid U = u) P(V' \leq t \mid U = u) \, d\mu_{A^{1,d}}(u)
= \int_{\mathbb{R}^d} K_A(u, [0, s]) K_A(u, [0, t]) \, d\mu_{A^{1,d}}(u)
= \Pi(s, t) + \text{cov}(K_A(U, [0, s]), K_A(U, [0, t]))
\]

for all \((s, t) \in \mathbb{R}^2\). The map \(\psi : C^{d+1} \to C^2\) given by

\[
\psi(A)(s, t) := \int_{\mathbb{R}^d} K_A(u, [0, s]) K_A(u, [0, t]) \, d\mu_{A^{1,d}}(u)
\]

for all \((s, t) \in \mathbb{R}^2\), then transforms every \((d+1)\)-dimensional copula to an exchangeable bivariate copula; see Figures 1, 2 and 3 for an illustration.

Figure 1: Sample of size \(n = 10,000\) drawn from a Clayton copula \(A\) (see [6, Example 6.5.17]) with parameter \(\alpha = -0.75\) (left panel), together with a sample of equal size from \(\psi(A)\) (right panel).
Figure 2: Sample of size $n = 10,000$ drawn from a Marshall-Olkin copula $A$ (see [6, Example 6.6.8]) with parameter $\alpha = 0.9$ and $\beta = 0.3$ (left panel), together with a sample of equal size from $\psi(A)$ (right panel).

Figure 3: Sample of size $n = 10,000$ drawn from a copula $A$ simulated from a standard normal random variable $X$ and random variable $Y = 2X^2 + 2X + \varepsilon$ with standard normal noise $\varepsilon$ (left panel), together with a sample of equal size from $\psi(A)$ (right panel).

As mentioned above, the copula $\psi(A)$ can be interpreted as the distribution function of two random variables that are conditionally independent and share the same conditional distribution. The copula $\psi(A)$ may also be interpreted as a generalization of the well-known bivariate Markov product of copulas: For two copulas $B, C \in \mathcal{C}^2$ the Markov product $B \ast C : \mathbb{I}^2 \to \mathbb{I}$ given by

$$(B \ast C)(s, t) := \int_{\mathbb{I}} K_B(u, [0, s]) K_C(u, [0, t]) \, d\lambda(u)$$

is a copula (see [2, 28] and [6, Chapter 5]); here, $B^t$ denotes the transpose of $B$. In the bivariate case, i.e. for $d = 1$, we therefore have $\psi(A) = A^t \ast A$. Notice that, in this case, the fix points of $\psi$ are exactly those copulas that are idempotent with respect to the Markov product (i.e. those copulas $A$ satisfying $A \ast A = A$).

The following theorem shows that the copula $\psi(A)$ characterises complete dependence of $Y$ on $X$; notice that Theorem 3.1 generalizes Theorem 11.1 in Darsow et al. [4].
3.1 Theorem. Consider a \((d+1)\)-dimensional random vector \((X, Y)\) with continuous marginals and connecting copula \(A\). Then the following statements are equivalent:

(a) \(Y\) is completely dependent on \(X\).

(b) \(\psi(A) = M\).

Proof. Assume that \(Y\) is completely dependent on \(X\). Then there exists some \(\mu_{A_{1:d}}\lambda\)-preserving map \(h\) such that \(K_A(u, [0, v]) = \mathbb{I}_{[0,v]}(h(u))\) for \(\mu_{A_{1:d}}\)-almost all \(u \in \mathbb{I}^d\) and every \(v \in \mathbb{I}\) (see [12]). In this case \(K_A(u, [0, s]) K_A(u, [0, t]) = \mathbb{I}_{[0,\lambda]}(h(u)) \mathbb{I}_{[0,\lambda]}(h(u)) = \mathbb{I}_{[0,\lambda]}(h(u)) = K_A(u, [0, M(s, t)])\) and hence

\[
\psi(A)(s, t) = \int_{\mathbb{I}^d} K_A(u, [0, M(s, t)]) d\mu_{A_{1:d}}(u) = M(s, t)
\]

for all \((s, t) \in \mathbb{I}^2\).

Now, assume that \(\psi(A) = M\). Then

\[
v = M(v, v) = \psi(A)(v, v) = \int_{\mathbb{I}^d} K_A(u, [0, v])^2 d\mu_{A_{1:d}}(u) \leq \int_{\mathbb{I}^d} K_A(u, [0, v]) d\mu_{A_{1:d}}(u) = v
\]

for all \(v \in \mathbb{I}\) which implies

\[
\int_{\mathbb{I}^d} K_A(u, [0, v]) - K_A(u, [0, v])^2 d\mu_{A_{1:d}}(u) = 0
\]

Since the integrand is positive we obtain \(K_A(u, [0, v]) - K_A(u, [0, v])^2 = 0\) or, equivalently, either \(K_A(u, [0, v]) = 0\) or \(K_A(u, [0, v]) = 1\) for \(\mu_{A_{1:d}}\)-almost every \(u \in \mathbb{I}^d\). Since \(v \mapsto K_A(u, [0, v])\) is increasing, it can be shown in exactly the same manner as in [27] Lemma 12 that there exists some \(\mu_{A_{1:d}} - \lambda\) preserving transformation \(h : \mathbb{I}^d \to \mathbb{I}\) such that \(K_A(u, F) = \mathbb{I}_F(h(u))\) is a regular conditional distribution of \(A\). This proves the result. \(\square\)

For a copula \(C \in C^2\), we denote by \(\delta_C\) its diagonal, i.e. \(\delta_C(t) := C(t, t)\) for all \(t \in \mathbb{I}^1\). The copula \(\psi(A)\) also characterises independence of \(Y\) and \(X\):

3.2 Theorem. Consider a \((d+1)\)-dimensional random vector \((X, Y)\) with continuous marginals and connecting copula \(A\). Then the following statements are equivalent:

(a) \(Y\) and \(X\) are independent.

(b) \(\psi(A) = \Pi\).

(c) \(\delta_{\psi(A)} = \delta_{\Pi}\).

Proof. Assume that \(Y\) and \(X\) are independent. Then

\[
\psi(A)(s, t) = \int_{\mathbb{I}^d} K_A(u, [0, s]) K_A(u, [0, t]) d\mu_{A_{1:d}}(u) = \int_{\mathbb{I}^d} s t d\mu_{A_{1:d}}(u) = \Pi(s, t)
\]

for all \((s, t) \in \mathbb{I}^2\) and hence (b), which then implies (c).

Now, assume that \(\delta_{\psi(A)} = \delta_{\Pi}\). Applying Hölder’s inequality yields

\[
v^2 = \delta_{\Pi}(v) = \delta_{\psi(A)}(v) = \int_{\mathbb{I}^d} K_A(u, [0, v])^2 d\mu_{A_{1:d}}(u) \geq \left( \int_{\mathbb{I}^d} K_A(u, [0, v]) d\mu_{A_{1:d}}(u) \right)^2 = v^2
\]
hence we have
\[
0 = \int_{\mathbb{I}^d} K_A(u, [0, v])^2 - v^2 \, d\mu_{A^{1:d}}(u) = \int_{\mathbb{I}^d} (K_A(u, [0, v]) - v)^2 \, d\mu_{A^{1:d}}(u)
\]
for all \( v \in \mathbb{I} \). Since the integrand is positive we obtain \( K_A(u, [0, v]) = v \) for \( \mu_{A^{1:d}} \)-almost every \( u \in \mathbb{I}^d \) and all \( v \in \mathbb{I} \), hence \( A(u, v) = A^{1:d}(u) \, v \) for all \( (u, v) \in \mathbb{I}^d \times \mathbb{I} \). This proves the result.

Since \( \psi(A) \) is a bivariate copula we have \( \psi(A) \leq M \). Applying Hölder’s inequality directly shows that the diagonal of \( \psi(A) \) exceeds the diagonal of \( \Pi \), i.e. \( \delta_{\psi(A)} \geq \delta_{\Pi} \) (Lemma 3.3). Example 3.4 however, illustrates that the inequality \( \psi(A)(s, t) \geq \Pi(s, t) \) fails to hold for every \( (s, t) \in \mathbb{I}^2 \).

3.3 Lemma. The diagonal \( \delta_{\psi(A)} \) of copula \( \psi(A) \) fulfills \( \delta_M \geq \delta_{\psi(A)} \geq \delta_{\Pi} \).

3.4 Example. Consider a \((1 + 1)\)-dimensional random vector \((U, V)\) with uniform univariate marginals such that \((U, V) \sim A = (M + W)/2\). Then
\[
\psi(A) = \frac{M + W}{2} = A
\]
and there exists some \((s, t) = (1/4, 3/4) \in \mathbb{I}^2 \) such that \( \psi(A)(s, t) = 2/16 < 3/16 = \Pi(s, t) \).

Due to Example 3.4, the copula \( \psi(A) \) fails to be stochastically increasing, left tail decreasing (LTD) and positive quadrant dependent (PQD), in general; see [24] for more information on these dependence properties.

The values of \( \psi(A) \) outside the diagonal are bounded from above by the values along the diagonal:

3.5 Lemma. For every \( A \in C^{d+1} \), the inequality
\[
\psi(A)(s, t) \, \psi(A)(t, s) = \psi(A)(s, t)^2 \leq \psi(A)(s, s) \, \psi(A)(t, t)
\]
holds for all \((s, t) \in \mathbb{I}^2 \).

Proof. The result is immediate from Hölder’s inequality.

The next result shows that the copula \( \psi(A) \) is invariant under a variety of measurable and bijective transformations of the first \( d \) coordinates of \( A \); this includes permutations and reflections of copulas (see [3] for more background on permutations and reflections of copulas):

3.6 Lemma. For \( A \in C^{d+1} \), consider the identity map \( \text{id} : \mathbb{I} \rightarrow \mathbb{I} \) and some measurable bijective transformation \( \zeta : \mathbb{I}^d \rightarrow \mathbb{I}^d \) such that the push-forward \( (\mu_{A^{1:d}})^\zeta \) is again a copula measure. Denoting \( A_{(\zeta, \text{id})} \) the copula of \( (\mu_{A})^{(\zeta, \text{id})} \), we have \( \psi(A_{(\zeta, \text{id})}) = \psi(A) \).

Proof. Since \( \zeta \) is bijective, the Markov kernel of \( A_{(\zeta, \text{id})} \) satisfies \( K_{A_{(\zeta, \text{id})}}(u, [0, v]) = K_{A}(\zeta^{-1}(u), [0, v]) \) for \( \mu_{A^{1:d}} \)-almost all \( u \in \mathbb{I}^d \) and all \( v \in \mathbb{I} \) and hence
\[
\psi(A_{(\zeta, \text{id})})(s, t) = \int_{\mathbb{I}^d} K_{A_{(\zeta, \text{id})}}(u, [0, s]) \, K_{A_{(\zeta, \text{id})}}(u, [0, t]) \, d\mu_{(A_{(\zeta, \text{id})})^{1:d}}(u)
\]
Then the inequality holds for all 
which proves the result.

3.7 Remark. In terms of a random vector \((X, Y)\) with continuous marginals and connecting copula \(A\) Lemma [3.6] implies that the transformation \(\psi\) is invariant

1. with respect to permutations of \(X\), and

2. with respect to coordinatewise continuous and strictly increasing (or decreasing) transformations of \(X\).

The copula \(\psi(A)\) satisfies the so-called information gain inequality (see, e.g., [12]) along the diagonal, i.e. the more conditioning variables are involved, the larger the value of \(\psi(A)\) along the diagonal.

3.8 Theorem. Consider some \(K \subseteq \{1, \ldots, d\}\) with \(1 \leq |K| \leq d - 1\) and some \(k \in \{1, \ldots, d\}\setminus K\). Then the inequality

\[
\psi(A^{K \cup \{d+1\}})(t, t) \leq \psi(A^{K \cup \{k\} \cup \{d+1\}})(t, t)
\]

holds for all \(t \in \mathbb{I}\). In particular, the inequality

\[
\psi(A^{1,d+1})(t, t) \leq \psi(A^{1:2,d+1})(t, t) \leq \cdots \leq \psi(A^{1:k,d+1})(t, t) \leq \cdots \leq \psi(A)(t, t) \tag{2}
\]

holds for all \(k \in \{1, \ldots, d\}\) and all \(t \in \mathbb{I}\).

Proof. Due to Lemma [3.6] w.l.o.g. we may assume that \(K = \{1, \ldots, k\}\) for some \(k \in \{1, \ldots, d-1\}\). Then, with the notation used at the beginning of this section

\[
\psi(A^{1:k,d+1})(t, t) = E(E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k))^2)
\]

\[
\psi(A^{1:(k+1),d+1})(t, t) = E(E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k, U_{k+1}))^2)
\]

for all \(t \in \mathbb{I}\). Considering the right hand side as conditional expectation

\[
\psi(A^{1:k,d+1})(t, t) = E((\mathbf{1}_{[0,t]} \circ V)^2) - E\left(E\left((\mathbf{1}_{[0,t]} \circ V - E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k)))^2\right)\right)
\]

\[
\psi(A^{1:(k+1),d+1})(t, t) = E((\mathbf{1}_{[0,t]} \circ V)^2) - E\left(E\left((\mathbf{1}_{[0,t]} \circ V - E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k, U_{k+1})))^2\right)\right)
\]

and applying Hilbert’s projection theorem yields

\[
E\left((\mathbf{1}_{[0,t]} \circ V - E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k, U_{k+1})))^2\right) \leq E\left((\mathbf{1}_{[0,t]} \circ V - E(\mathbf{1}_{[0,t]} \circ V \mid (U_1, \ldots, U_k)))^2\right)
\]

and hence

\[
\psi(A^{1:k,d+1})(t, t) \leq \psi(A^{1:(k+1),d+1})(t, t)
\]

for all \(t \in \mathbb{I}\). This proves the assertion.
In general, the information gain inequality does not hold outside the diagonal.

3.9 Example. Consider the copula \( A : \mathbb{I}^3 \to \mathbb{I} \) given by
\[
A(u_1, u_2, v) := u_1 \frac{M+W}{2}(u_2, v)
\]
Then \( A^{13} = \Pi \), hence \( \psi(A^{13}) = \Pi \), and
\[
\psi(A)(s, t) = \int_{\mathbb{I}^2} K_A(u, [0, s]) K_A(u, [0, t]) \, d\mu_{A^{1:2}}(u)
\]
for all \((s, t) \in \mathbb{I}^2\). In accordance with Theorem 3.8 and Lemma 3.3 we obtain
\[
\psi(A^{13})(t, t) \leq \psi(A)(t, t)
\]
for all \( t \in \mathbb{I} \), however, by Example 3.4 there exists some \((s, t) = (1/4, 3/4) \in \mathbb{I}^2\) such that
\[
\psi(A)(s, t) = 2/16 < 3/16 = \psi(A^{13})(s, t).
\]

For copulas satisfying the so-called conditional independence assumption (w.r.t. the \( k \)-th coordinate, see [2] [12])
\[
K_A\left(\bigotimes_{i \in \{1, \ldots, d+1\} \setminus \{k\}} E_i \bigotimes_{i \in \{1, \ldots, d+1\} \setminus \{k\}} K_{A^k}(u_k, E_i) \right) = \prod_{i \in \{1, \ldots, d+1\} \setminus \{k\}} K_{A^k}(u_k, E_i) \tag{3}
\]
for \( \lambda \)-a.e. \( x_k \in \mathbb{I} \) and \( E_1, \ldots, E_{d+1} \in \mathcal{B}(\mathbb{I}) \), the information gain inequality [2] becomes an equality:

3.10 Theorem. Consider a \((d + 1)\)-dimensional random vector \((X, Y)\) with continuous marginals and connecting copula \( A \), and assume that \( A \) satisfies the conditional independence assumption. Then \( \psi(A) = \psi(A^{k,d+1}) \).

Proof. W.l.o.g. consider \( k = 1 \). Then \( A \) satisfies the conditional independence assumption [3] and [12] Lemma 3.5 yields
\[
\psi(A)(s, t) = \int_{\mathbb{I}^d} K_A(u, [0, s]) K_A(u, [0, t]) \, d\mu_{A^{1:2}}(u)
\]
for all \((s, t) \in \mathbb{I}^2\). This proves the assertion.

It has been recognized in [12] that the family of copulas satisfying equation (3) is dense in \((\mathcal{C}^{d+1}, d_\infty)\) where \( d_\infty \) denotes the uniform metric on \( \mathcal{C}^{d+1} \).
Notice that, in terms of a random vector \((X, Y)\) with continuous marginals and connecting copula \(A\), the conditional independence assumption holds whenever there exists some \(k \in \{1, \ldots, d\}\) such that \(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_d, Y\) are conditionally independent given \(X_k\).

**3.11 Remark.** If \(A\) is absolutely continuous with Lebesgue density \(a\), then \(\psi(A)\) is absolutely continuous as well and satisfies
\[
\psi(A)(s, t) = \int_{I^d} K_A(u, [0, s]) K_A(u, [0, t]) \, d\mu_{A^1:d}(u)
\]
\[
= \int_{\{u \in I^{d:1:d}(u) > 0\}} K_A(u, [0, s]) K_A(u, [0, t]) \, a^{1:d}(u) \, d\lambda^d(u)
\]
\[
= \int_{[0,s] \times [0,t]} \int_{\{u \in I^{d:1:d}(u) > 0\}} \frac{a(u, p) a(u, q)}{a^{1:d}(u)} \, d\lambda^d(u) \, d\lambda^2(p, q)
\]
for all \((s, t) \in I^2\).

For an illustration, consider the EFGM copula (see [6, Section 6.3]) \(A \in C^{d+1}\) (see Figure 4) with parameter \(\alpha \in [-1, 1]\) given by
\[
A(u, v) := \Pi(u, v) + \alpha v(1 - v) \prod_{i=1}^d u_i(1 - u_i)
\]
\[
= \int_{[0,u] \times [0,v]} 1 + \alpha(1 - 2z) \prod_{i=1}^d (1 - 2w_i) \, d\lambda^{d+1}(w, z)
\]

Then \(a^{1:d}(u) = 1\) for all \(u \in I^d\) and hence
\[
\psi(A)(s, t) = \int_{[0,s] \times [0,t]} \int_{I^d} \left( 1 + \alpha(1 - 2p) \prod_{i=1}^d (1 - 2u_i) \right) \left( 1 + \alpha(1 - 2q) \prod_{i=1}^d (1 - 2u_i) \right) \, d\lambda^d(u) \, d\lambda^2(p, q)
\]
\[
= \Pi(s, t) + \int_{[0,s] \times [0,t]} \alpha^2(1 - 2p)(1 - 2q) \, d\lambda^2(p, q) \int_{I^d} \prod_{i=1}^d (1 - 2u_i)^2 \, d\lambda^d(u)
\]
\[
= \Pi(s, t) + \frac{\alpha^2}{3^d} s(1 - s) t(1 - t)
\]
for all \((s, t) \in I^2\). Therefore, \(\psi(A)\) where \(A\) is EFGM with parameter \(\alpha\) is again of EFGM type but with parameter \(\alpha^2/3^d\).

In the context of Markov products, it has been recognized in [6, Theorem 5.2.10] that, for \(d = 1\), the map \(A \mapsto A^t \ast A\) fails to be jointly continuous with respect to the topology of uniform convergence, and hence uniform convergence of a sequence of copulas \((A_n)_{n \in \mathbb{N}}\) to \(A\) does not imply uniform convergence of the sequence \((\psi(A_n))_{n \in \mathbb{N}}\) to \(\psi(A)\) (see also [26, Theorem 6]).

## 4 Estimation

In this section, we consider a \((d + 1)\)-dimensional random vector \((X, Y)\) with continuous univariate marginal distribution functions \(F_i\) of \(X_i, i \in \{1, \ldots, d\}\), and \(G\) of \(Y\) and connecting copula \(A\).

Further, let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. copies of \((X, Y)\). Since the univariate marginals are continuous ties only occur with probability 0. For each \(i\), we denote by \(N(i)\) the index \(j\) such that
Figure 4: Sample of size \( n = 10,000 \) drawn from an EFGM copula \( A \) (see [6, Section 6.3]) with parameter \( \alpha = 1 \) (left panel), together with a sample of equal size from \( \psi(A) \) (right panel).

\( X_j \) is the nearest neighbour of \( X_i \) with respect to the Euclidean metric on \( \mathbb{R}^d \). Since there may exist several nearest neighbours of \( X_i \) ties are broken at random.

For \( (s,t) \in \mathbb{R}^2 \), we define

\[
D_n(s,t) := \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{[0,s]}(G_n^*(Y_k)) \mathbf{1}_{[0,t]}(G_n^*(Y_{N(k)})
\]

where \( G_n^* \) denotes a renormalised version of the empirical distribution function of \( Y_1, \ldots, Y_n \), i.e.

\[
G_n^*(y) = \frac{1}{n+1} \sum_{k=1}^{n} \mathbf{1}_{(-\infty,y]}(Y_k)
\]

and note that

\[
\psi(A)(s,t) = \int_{\mathbb{R}^d} K_A(u,[0,s])K_A(u,[0,t]) \, d\mu_A(u) = E\left(P\left(G(Y) \leq s \mid X\right) \, P\left(G(Y) \leq t \mid X\right)\right)
\]

By adapting the ideas used in Azadkia and Chatterjee [1], in what follows we show that, for \( (s,t) \in \mathbb{R}^2 \),

\[
\lim_{n \to \infty} D_n(s,t) = \psi(A)(s,t)
\]

almost surely. However, for proving (5) we use a modification of \( D_n \) for which we choose the usual normalisation of the ranks. For \( (s,t) \in \mathbb{R}^2 \), define

\[
C_n(s,t) := \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{[0,s]}(G_n(Y_k)) \mathbf{1}_{[0,t]}(G_n(Y_{N(k)})
\]

where \( G_n \) denotes the empirical distribution function of \( Y_1, \ldots, Y_n \), i.e.

\[
G_n(y) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{(-\infty,y]}(Y_k)
\]

Due to Proposition 4.1 below, the estimates \( C_n \) and \( D_n \) are asymptotically equivalent (see also (7) below).
For a realization $X_1, \ldots, X_n$ and each $i \in \{1, \ldots, n\}$, let $K_{n,i}$ be the number of $j$ such that $X_i$ is the nearest neighbour of $X_j$. The following result is given in [1]:

4.1 Proposition. [1] Lemma 11.4
There exists a constant $c(d)$ such that $K_{n,1} \leq c(d)$. Notice that the upper bound $c(d)$ used in Proposition 4.1 only depends on dimension $d$.

The following two results are key for proving consistency of (6):

4.2 Lemma. For every $(s, t) \in \mathbb{R}^2$ we have $\lim_{n \to \infty} E(C_n(s, t)) = \psi(A)(s, t)$.

Proof. For $(s, t) \in \mathbb{R}^2$, we define

$$C_n^*(s, t) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{[0, s]}(G(Y_k)) \mathbb{I}_{[0, t]}(G(Y_{N(k)}))$$

and show that

1. $\lim_{n \to \infty} E(|C_n(s, t) - C_n^*(s, t)|) = 0$ and
2. $\lim_{n \to \infty} E(C_n^*(s, t)) = \psi(A)(s, t)$.

We first prove (1). For $(s, t) \in \mathbb{R}^2$, we have

$$|C_n(s, t) - C_n^*(s, t)| \leq \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{[0, s]}(G_n(Y_k)) \mathbb{I}_{[0, t]}(G_n(Y_{N(k)})) - \mathbb{I}_{[0, s]}(G(Y_k)) \mathbb{I}_{[0, t]}(G(Y_{N(k)})) \right|$$

$$\leq \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{[0, s]}(G_n(Y_k)) - \mathbb{I}_{[0, s]}(G(Y_k)) \right| + \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{[0, t]}(G_n(Y_{N(k)})) - \mathbb{I}_{[0, t]}(G(Y_{N(k)})) \right|$$

$$=: I_{1,s} + I_{2,t}$$

Since $G$ is continuous, we first obtain

$$\frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{(-\infty, G_n^+(s))}(Y_k) - \mathbb{I}_{(-\infty, G^+(s))}(Y_k) \right|$$

$$= \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{(-\infty, G_n^+(s))}(Y_k) + \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{(-\infty, G^+(s))}(Y_k) - 2 \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{(-\infty, G_n^+(s))}(Y_k) \mathbb{I}_{(-\infty, G^+(s))}(Y_k)$$

$$= |G_n \circ G_n^+(s) - G_n \circ G^+(s)|$$

$$\leq |G_n \circ G_n^+(s) - s| + |G \circ G^+(s) - G_n \circ G^+(s)|$$

By Glivenko-Cantelli, the second term converges to 0 almost surely and, for $(l - 1)/n < s \leq l/n$, the first term reduces to $|G_n \circ G_n^+(s) - s| = (l/n - s) \leq 1/n$. Therefore,

$$\lim_{n \to \infty} E(I_{1,s}) = \lim_{n \to \infty} E \left( \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{[0, s]}(G_n(Y_k)) - \mathbb{I}_{[0, s]}(G(Y_k)) \right| \right)$$

$$\leq \lim_{n \to \infty} \left( \frac{1}{n} + E \left( \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{(-\infty, G_n^+(s))}(Y_k) - \mathbb{I}_{(-\infty, G^+(s))}(Y_k) \right| \right) \right)$$

$$= 0$$
that for any

This proves (1).

Further, let

In order to prove key result (2) we adapt the ideas used in the proof of [1, Lemma 11.8]. Let \( F \) be the \( \sigma \)-field generated by \( X_1, \ldots, X_n \) and the random variables used for breaking nearest neighbours. Then, for \( (s, t) \in \mathbb{I}^2 \), we have

Further, let \( f_t(X) := P(G(Y) \leq t|X) \). Then \( f_t \) is measurable and, by [1] Lemma 11.7, \( f_t(X_{N(1)}) - f_t(X_1) \) tends to 0 in probability, and since \( |f_s(X)| = P(G(Y) \leq s|X_1) \leq 1 \) it follows that (see, e.g., [9])

and hence \( \lim_{n \to \infty} E(C_n^*(s, t)) = \psi(A)(s, t) \). This proves (2) from which the assertion follows. \( \square \)

The proof of the following lemma is similar to that of [1] Lemma 11.9:

**4.3 Lemma.** There are positive constants \( M_1 \) and \( M_2 \) depending only on dimension \( d \) such that for any \( n \in \mathbb{N} \) and any \( \eta \in (0, \infty) \)

for every \( (s, t) \in \mathbb{I}^2 \).
Proof. We prove the result using the bounded difference concentration inequality (see, e.g., [21]). For that purpose, we consider a realization \((X_1, Y_1), \ldots, (X_n, Y_n)\) and i.i.d. uniformly U[0,1]-distributed random variables \(U_1, \ldots, U_n\) where \(U_i\) is used for breaking ties when selecting \(X_i\)’s nearest neighbour. We replace \((X_i, Y_i, U_i)\) by some alternative value \((X_i^0, Y_i^0, U_i^0)\) and find a bound on the maximum possible change in \(C_n\). After such a replacement, each \(G_n(Y_j), j \neq i\), can change by at most 1/n. Now, for \((s, t) \in \mathbb{R}^2\), define

\[
a_k := \mathbb{I}_{[0,\infty]}(G_n(Y_k)) \quad b_k := \mathbb{I}_{[0,\infty]}(G_n(Y_k))
\]

Then \(C_n(s, t) = \frac{1}{n} \sum_{k=1}^{n} a_k b_{N(k)}\). When replacing \(Y_i\) by \(Y_i^0\) there exist at most two indices \(l\) such that \(a_l\) changes by 1 and, by Proposition 4.1, there are at most \(2 \cdot c(d)\) indices \(l\) such that \(b_{N(l)}\) changes by 1. When replacing \(X_i\) by \(X_i^0\), by Proposition 4.1 there are at most \(2 \cdot c(d) + 1\) indices \(l\) such that \(b_{N(l)}\) changes by 1. And when replacing \(U_i\) by \(U_i^0\) there exists at most one index \(l\) such that \(b_{N(l)}\) changes by 1. In sum, there are at most \(4 \cdot c(d) + 4\) indices \(l\) such that \(a_l\) or \(b_{N(l)}\) changes by 1, and hence \(C_n(s, t)\) changes by at most \((4 \cdot c(d) + 4)/n\). Applying the bounded difference concentration inequality and replacing \((4 \cdot c(d) + 4)/n\) by \(c(d)\), for every \(\eta \in (0, \infty)\), we obtain

\[
P(\{|C_n(s, t) - E(C_n(s, t))| \geq \eta\}) \leq 2 \exp(-c(d)\eta^2/n)
\]

As the above inequality holds for all \((s, t) \in \mathbb{R}^2\), this completes the proof. \(\Box\)

Combining Lemma 4.2 and Lemma 4.3 proves consistency of \(C_n\):

4.4 Lemma. Let \((X_1, Y_1), (X_2, Y_2), \ldots\) be a random sample from \((X, Y)\) with continuous marginals and connecting copula \(A \in \mathcal{C}^{d+1}\). Then, for all \((s, t) \in \mathbb{R}^2\), we have

\[
\lim_{n \to \infty} C_n(s, t) = \psi(A)(s, t)
\]

almost surely.

Due to Proposition 4.1, we have

\[
|C_n(s, t) - D_n(s, t)| \leq \frac{k(d) + 1}{n}
\]

implying that \(C_n\) and \(D_n\) are asymptotically equivalent. We are now in the position to prove Equation 5:

4.5 Theorem. Let \((X_1, Y_1), (X_2, Y_2), \ldots\) be a random sample from \((X, Y)\) with continuous marginals and connecting copula \(A \in \mathcal{C}^{d+1}\). Then, for all \((s, t) \in \mathbb{R}^2\), we have

\[
\lim_{n \to \infty} D_n(s, t) = \psi(A)(s, t)
\]

almost surely. In particular, in the bivariate case, i.e. when \(d = 1\), the estimates \(C_n\) and \(D_n\) are consistent estimators for the Markov product \(A^1 * A\).
4.1 Simulation study

We now illustrate the small and moderate sample performance of our estimator $D_n$ for the two dependence structures discussed in Remark 3.11 and Example 5.8: (1) the EFGM copula, and (2) the so-called tent copula; see also Figures 4 and 7. We restrict ourselves to the case $d = 1$ in order to be able to interpret the results obtained also as estimates of the Markov product $A^t \ast A$.

If $X$ and $Y$ have EFGM copula $A$ with parameter $\alpha = 1$ as connecting copula, by Remark 3.11, the copula $\psi(A)$ is EFGM as well with parameter $1/3$. To test the performance of our estimator $D_n$ in this setting, we generated samples of size $n \in \{20, 50, 100, 200, 500, 1,000, 5,000, 10,000\}$ and calculated $D_n$. These steps were repeated $R = 1,000$ times. Figure 5 then depicts the $d_\infty$-distance between our estimate and the true copula evaluated on a grid of size 50.

If $X$ and $Y$ have the so-called tent copula $A$ as connecting copula (see Example 5.8 and Figure 7), the copula $\psi(A)$ equals $(M + W)/2$. To test the performance of our estimator $D_n$ in this setting, we generated samples of size $n \in \{20, 50, 100, 200, 500, 1,000, 5,000, 10,000\}$ and calculated $D_n$. These steps were repeated $R = 1,000$ times. Figure 6 then depicts the $d_\infty$-distance between our estimate and the true copula evaluated on a grid of size 50.

As can be seen from Figures 5 and 6, the copula estimate converges rather fast to the true copula.

Figure 5: Boxplots summarizing the 1,000 obtained $d_\infty$-distances between our estimate and the true copula. Samples of size $n$ are drawn from an EFGM copula $A$ with parameter $\alpha = 1$. 
5 Applications

5.1 A simple measure of conditional dependence

Quite recently, Azadkia and Chatterjee [11] introduced their so-called ‘simple measure of conditional dependence’ which is based on [3, 5] and, for a \((d + 1)\)-dimensional random vector \((X, Y)\), given by

\[
T(Y \mid X) := \frac{\int_{\mathbb{R}} \text{var}(P(Y \geq y \mid X)) \, dP_Y(y)}{\int_{\mathbb{R}} \text{var}(1_{\{Y \geq y\}}) \, dP_Y(y)}
\]

The map \(T\) quantifies the scale-invariant extent of dependence of the random variable \(Y\) on the random vector \(X\) and thus belongs to a class of indices capable of detecting complete dependence and independence which also include [7, 12, 14, 27].

In our setting, i.e. considering random vectors \((X, Y)\) with continuous marginals, \(T\) possesses an alternative representation in terms of a well-known dependence measure:

**5.1 Theorem.** Consider a \((d + 1)\)-dimensional random vector \((X, Y)\) with continuous marginals and connecting copula \(A\). Then \(T\) satisfies

\[
T(Y \mid X) = \phi(\psi(A))
\]

where \(\phi : C^2 \to \mathbb{R}\) denotes Spearman’s footrule given by

\[
\phi(C) = 6 \int_\frac{1}{2} \int C(t, t) \, d\lambda(t) - 2
\]

**Proof.** Denote by \(H\) the distribution function of \((X, Y)\). Since \((X, Y)\) has continuous marginals, the random vector \((F(X), G(Y))\) has distribution function \(A\). By change of measure we first obtain

\[
\int_{\mathbb{R}} \text{var}(1_{\{Y \geq y\}}) \, dP_Y(y) = \int_\frac{1}{2} P(Y \geq G^+(v))[1 - P(Y \geq G^+(v))] \, d\lambda(v) = \int_1 (1 - v) \, v \, d\lambda(v) = \frac{1}{6}
\]
Finally, $T(Y|X) = 6 \int_1 \psi(A)(t,t) \, d\lambda(t) - 2 = \phi(\psi(A))$ which proves the assertion. \hfill \Box

The following properties of $T$ can be derived from Theorem 5.1 and the results in Section 5.

**5.2 Theorem.** Consider a $(d+1)$-dimensional random vector $(X,Y)$ with continuous marginals and connecting copula $A$. Then the following properties hold:

1. $T(Y|X) = 1$ if and only if $Y$ is completely dependent on $X$.
2. $T(Y|X) = 0$ if and only if $Y$ and $X$ are independent.
3. $T(Y|X)$ is invariant with respect to permutations of $X_1, \ldots, X_d$.
4. $T(Y|X)$ is invariant with respect to continuous and strictly monotone transformations of $X_1, \ldots, X_d$.
5. $T$ fulfills the information gain inequality, i.e.

$$T(Y|X_1) \leq T(Y|(X_1, X_2)) \leq \cdots \leq T(Y|(X_1, \ldots, X_k)) \leq \cdots \leq T(Y|(X_1, \ldots, X_d))$$

for all $k \in \{1, \ldots, d\}$. Additionally, if $X_2, \ldots, X_d, Y$ are conditionally independent given $X_1$, then $T(Y|X) = T(Y|X_1)$.

Properties (1) and (2) in Theorem 5.2 also hold for non-continuous random vectors $(X,Y)$; see [1, Theorem 2.1].

Finally, consider an i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X,Y)$. For each $i$, let further $N(i)$ be the index $j$ such that $X_j$ is the nearest neighbor of $X_i$ with respect to the Euclidean metric on $\mathbb{R}^d$ (ties broken uniformly at random). As estimate of $T$ Akzadia & Chatterjee [1] propose to use the statistic

$$T_n(Y|X) = \frac{\sum_{i=1}^n (n \min\{R_i, R_{N(i)}\} - L_i^2)}{\sum_{i=1}^n L_i(n-L_i)}$$
where \( R_i \) denotes the rank of \( Y_i \) among \( Y_1, \ldots, Y_n \), i.e. the number of \( j \) such that \( Y_j \leq Y_i \), and \( L_i \) denotes the number of \( j \) such that \( Y_j \geq Y_i \) (see also [13]). Straightforward calculation yields

\[
T_n(Y|X) = 1 - \frac{3}{n^2 - 1} \sum_{i=1}^{n} |R_i - R_{N(i)}| + \frac{3}{n^2 - 1} \left( \sum_{i=1}^{n} R_{N(i)} + \sum_{i=1}^{n} R_i - n(n+1) \right)
\]

Notice that there may exist more than one index \( i \) such that \( X_j \) is a nearest neighbour of \( X_i \) implying that \( \sum_{i=1}^{n} R_{N(i)} \) may fail to equal \( \frac{n(n+1)}{2} \).

It turns out that \( T_n(Y|X) \) is a strongly consistent estimator for \( T(Y|X) \):

5.3 Proposition. [1, Theorem 2.2] Suppose that \( Y \) is not almost surely a constant. Then \( \lim_{n \to \infty} T_n(Y|X) = T(Y|X) \) almost surely. 

In [25] the authors showed asymptotic normality of \( \sqrt{n}T_n \) under independence and for some regularity conditions; for a comprehensive summary of properties for \( T_n \) we refer to Han [13] and the references therein.

Motivated by Theorem 5.1, for \((d+1)\)-dimensional random vectors \((X, Y)\) with continuous marginals and connecting copula \( A \), we propose to use the estimate

\[
S_n(Y|X) := 1 - \frac{3}{n(n+1)} \sum_{i=1}^{n} |R_i - R_{N(i)}| + \frac{3}{n(n+1)} \left( n(n+1) - \sum_{i=1}^{n} R_{N(i)} - \sum_{i=1}^{n} R_i \right)
\]

which equals the plug-in estimate \( \phi(D_n) \) of Spearman’s footrule (see [9, 11, 24]) with \( D_n \) being the consistent estimator for \( \psi(A) \) defined in Section 4. By Theorem 4.5, the estimate \( S_n(Y|X) \) is another consistent estimator for \( T(Y|X) \):

5.4 Corollary. Let \((X_1, Y_1), (X_2, Y_2), \ldots\) be a random sample from \((X, Y)\) with continuous marginals and connecting copula \( A \in \mathcal{C}^{d+1} \). Then

\[
\lim_{n \to \infty} S_n(Y|X) = T(Y|X)
\]

almost surely.

5.2 Nonparametric coefficient of determination

Let \((X, Y)\) be a \((d+1)\)-dimensional random vector such that \( Y \in L^2 \). It is well known that the variance \( \text{var}(Y) \) of \( Y \) can be decomposed via

\[
\text{var}(Y) = \text{var}(E(Y|X)) + \text{var}(Y - E(Y|X))
\]

\[
= \text{var}(E(Y|X)) + E(\text{var}(Y|X))
\]

where the variance \( \text{var}(E(Y|X)) \) equals the part of the variance of \( Y \) explained by the regression function \( r(x) := E(Y|X = x) \), and

\[
R^2 := R^2(Y|X) := \frac{\text{var}(E(Y|X))}{\text{var}(Y)} = 1 - \frac{\text{var}(Y - E(Y|X))}{\text{var}(Y)} = 1 - \frac{E((Y - E(Y|X))^2)}{\text{var}(Y)}
\]

(also known as Sobol index; see [10]) denotes the proportion of the variance that is explained by the regression function \( r \).
5.5 Theorem. Consider a \((d+2)\)-dimensional random vector \((\mathbf{X}, Y, Y')\) such that \((\mathbf{X}, Y')\) and \((\mathbf{X}, Y')\) are identically distributed, and assume that \(Y\) and \(Y'\) are conditionally independent given \(\mathbf{X}\). Then \(R^2\) of \(Y\) given \(\mathbf{X}\) fulfills
\[
R^2(Y|\mathbf{X}) = \rho_p(Y, Y')
\]
where \(\rho_p\) denotes Pearson correlation coefficient.

Proof. Since \(E(E(Y|\mathbf{X})^2) = E(E(Y|\mathbf{X})E(Y'|\mathbf{X})) = E(E(YY'|\mathbf{X})) = E(YY')\) we obtain
\[
R^2(Y|\mathbf{X}) = \frac{\text{var}(E(Y|\mathbf{X}))}{\text{var}(Y)} = \frac{E(E(Y|\mathbf{X})^2) - E(Y)^2}{\text{var}(Y)} = \frac{E(YY') - E(Y)E(Y')}{{\text{var}(Y)}^{\frac{1}{2}}{\text{var}(Y')}^{\frac{1}{2}}} = \rho_p(Y, Y')
\]
This proves the result. \(\square\)

Again, assume that the random vector \((\mathbf{X}, Y)\) has continuous marginals \(F_i\) of \(X_i\), \(i \in \{1, \ldots, d\}\), and \(G\) of \(Y\), distribution function \(H\) and connecting copula \(A\). Then (denoting \(U_i := F_i(X_i)\), \(i \in \{1, \ldots, d\}\), and \(V := G(Y)\)) the distribution-free \(R^2(V|U)\) possesses a representation in terms of another well-known dependence measure:

5.6 Theorem. Consider a \((d+1)\)-dimensional random vector \((\mathbf{X}, Y)\) with continuous marginals and connecting copula \(A\). Then \(R^2\) of \(V\) given \(U\) fulfills
\[
R^2(V|U) = \varrho_S(\psi(A))
\]
where \(\varrho_S : C^2 \to \mathbb{R}\) denotes Spearman’s rank correlation coefficient (a.k.a. Spearman’s rho) given by
\[
\varrho_S(C) = 12 \int_{I^2} C(s,t) \, d\lambda^2(s,t) - 3 = 12 \int_{I^2} st \, d\mu_C(s,t) - 3
\]
For the bivariate case, i.e. \(d = 1\), it has been recognized in Shih and Emura [25, Theorem 1] that \(R^2(V|U)\) can be represented as Spearman’s rho of the Markov product \(A^t \ast A\), which in this case coincides with \(\psi(A)\).

The following properties of \(R^2(V|U)\) can be derived from Theorem 5.6 and the results in Section 3. The information gain inequality (i.e. reducing the number of conditioning variables reduces \(R^2\)) does not follow from Theorem 3.8 but from Hilbert’s projection theorem.

5.7 Theorem. Consider a \((d+1)\)-dimensional random vector \((\mathbf{X}, Y)\) with continuous marginals and connecting copula \(A\). Then the following properties hold:

1. \(R^2(V|U) = 1\) if and only if \(Y\) is completely dependent on \(\mathbf{X}\).
2. \(R^2(V|U) = 0\) if and only if \(E(V|U = \mathbf{u}) = \frac{1}{2}\) for \(\mu_{A^{d\times d}}\)-almost all \(\mathbf{u} \in \mathbb{R}^d\). In particular, \(R^2(V|U) = 0\) whenever \(Y\) and \(\mathbf{X}\) are independent.
3. \(R^2(V|U)\) is invariant with respect to permutations of \(X_1, \ldots, X_d\).
4. \(R^2(V|U)\) is invariant with respect to continuous and strictly monotone transformations of \(X_1, \ldots, X_d\).
5. \( R^2(V|U) \) fulfills the information gain inequality, i.e.
\[
R^2(V|U_1) \leq R^2(V|(U_1, U_2)) \leq \cdots \leq R^2(V|(U_1, \ldots, U_k)) \leq \cdots \leq R^2(V|(U_1, \ldots, U_d))
\]
for all \( k \in \{1, \ldots, d\} \). Additionally, if \( X_2, \ldots, X_d, Y \) are conditionally independent given \( X_1 \), then \( R^2(V|U) = R^2(V|U_1) \).

Proof. It remains to prove property (2) which is immediate from the identity
\[
R^2(V|U) = 12 \int_{I^d} \left( E(V|U = u) - \frac{1}{2} \right)^2 d\mu_{A_1 \cdots A_d}(u)
\]
This proves the assertion. \( \square \)

For \( d = 1 \), properties (1) and (2) in Theorem [5.7] are given in [26, Proposition 2].

Notice that \( R^2(Y|X) \) and its distribution-free version \( R^2(V|U) \) may differ:

5.8 Example. Consider the case \( d = 1 \) and the uniformly distributed random variable \( V \sim U[0, 1] \) and define
\[
U := 2V 1_{[0,0.5]}(V) + (2 - 2V) 1_{(0.5,1]}(V) \sim U[0, 1]
\]
Then \( E(V|U = u) = \frac{1}{2} \) for \( \lambda \)-almost all \( u \in I \) and by Theorem [5.7] we obtain \( R^2(V|U) = 0 \).
Now, consider the random vector \((X, Y) := (U, V^2)\). Since the map \( v \mapsto v^2 \) is continuous and strictly increasing on \( I \), the random vectors \((X, Y)\) and \((U, V)\) share the same copula and hence \((F(X), G(Y)) \sim (U, V)\). In addition,
\[
R^2(Y|X) = \frac{\text{var}(E(Y|X))}{\text{var}(Y)} = \frac{\text{var}(E(V^2|U))}{\text{var}(V^2)} = \frac{45}{4} \int_{I^3} (E(V^2|U = u) - E(V^2))^2 d\lambda(u) = \frac{1}{16}
\]
and hence \( R^2(V|U) = 0 < \frac{1}{16} = R^2(Y|X) \).
\( \square \)

Figure 7: Sample of size \( n = 10,000 \) drawn from the random vector \((U, V) \sim A\) discussed in Example [5.8] (left panel), together with a sample of equal size from \( \psi(A) \) (right panel).
Finally, consider an i.i.d. sample \((X_1, Y_1), \ldots, (X_n, Y_n)\) of \((X, Y)\) with continuous marginals. For each \(i\), let further \(N(i)\) be the index \(j\) such that \(X_j\) is the nearest neighbor of \(X_i\) with respect to the Euclidean metric on \(\mathbb{R}^d\) (ties broken uniformly at random). Motivated by Theorem 5.6, as estimate for \(R^2(V|U)\) we propose to use the statistic

\[
R_n^2(V|U) := \frac{12}{n(n+1)^2} \sum_{i=1}^{n} R_i R_{N(i)} - 3 + \frac{12}{n(n+1)} \left( n(n+1) - \sum_{i=1}^{n} R_{N(i)} - \sum_{i=1}^{n} R_i \right)
\]

which equals the plug-in estimate \(\rho_S(D_n)\) of Spearman’s rho (see [24]) with \(D_n\) being the consistent estimator for \(\psi(A)\) defined in Section 4. By Theorem 4.5, the estimate \(R_n^2(V|U)\) is a strongly consistent estimator for \(R^2(V|U)\):

**5.9 Theorem.** Let \((X_1, Y_1), (X_2, Y_2), \ldots\) be a random sample from \((X, Y)\) with continuous marginals and connecting copula \(A \in C^{d+1}\). Then

\[
\lim_{n \to \infty} R_n^2(V|U) = R^2(V|U)
\]

almost surely. For the bivariate case, i.e. \(d = 1\), Gamboa et al. [10] introduced an estimate for \(R^2(Y|X)\) based on Pearson’s correlation coefficient using the technique introduced by Chatterjee [3].

6 Real data example

Finally, we present a real data example and use the introduced methodology to visualize and quantify the degree of predictability and explainability of a single random variable using the information contained in a set of potential explanatory random variables.

6.1 Analysis of global climate data

We consider a data set of bioclimatic variables for \(n = 1862\) locations homogeneously distributed over the global landmass from CHELSEA ([16, 17]) and want to analyse the influence of a set of thermal variables on annual precipitation.

We first focus on the case \(d = 1\) and use the variable Annual Mean Temperature (AMT) to predict/explain the Annual Precipitation (AP). Figure 8 depicts the original dependence structure of AMT and AP (left panel) and the estimate \(D_n\) of \(\psi(A)\) (right panel). The estimate \(S_n\) for \(T\), i.e. for the scale-invariant extent of dependence of Annual Precipitation on Annual Mean Temperature, equals 0.5603 and the estimate \(R_n^2\) for \(R^2(V|U)\), i.e. for the distribution-free coefficient of determination in this model, equals 0.6393.

In a second step, we extend the number of potential explanatory thermal variables to \(d = 11\) (including Annual Mean Temperature (AMT), Mean Diurnal Range (MDR), Isothermality (IT), Temperature Seasonality (TS), Max Temperature of Warmest Month (MTWM), Min Temperature of Coldest Month (MTCM), Temperature Annual Range (TAR), Mean Temperature of Wettest Quarter (MTWeQ), Mean Temperature of Driest Quarter (MTDQ), Mean Temperature of Coldest Quarter (MTCQ), Mean Temperature of Warmest Quarter (MTWaQ)) to predict/explain the Annual Precipitation (AP). Figure 9 depicts the estimate \(D_n\) of \(\psi(A)\) where \(A\) describes the dependence structure of variables AMT, MDR, IT, TS, MTWM, MTCM, TAR, MTWeQ, MTDQ, MTCQ, MTWaQ and AP. The estimate \(S_n\) for \(T\), i.e. for the scale-invariant extent of dependence of AP
Figure 8: Original dependence structure of AMT and AP (left panel), together with the estimate $D_n$ of $\psi(A)$ (right panel).

on AMT, MDR, IT, TS, MTWM, MTCM, TAR, MTWeQ, MTDQ, MTCQ and MTWaQ, equals 0.6616 and the estimate $R^2_n$ for $R^2(V|U)$, i.e. for the distribution-free coefficient of determination in this model, equals 0.8105.

Figure 9: together with the estimate $D_n$ of $\psi(A)$ (right panel).

The increase in the estimated values for $T$ and $R^2$ with increasing number of explanatory variables is in accordance with Theorem 5.2 and Theorem 5.7.

Acknowledgement

The author gratefully acknowledges the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).
References

[1] Azadkia, M. and S. Chatterjee (2021). A simple measure of conditional dependence. *Ann. Stat.*, to appear.

[2] Bayramoglu, I. (2010). On conditionally independent random variables, copula and order statistics. *Comm. Statist. Theory Meth.* 43(10-12), 2105–2117.

[3] Chatterjee, S. (2020). A new coefficient of correlation. *J. Amer. Statist. Ass.*, to appear.

[4] Darsow, W., B. Nguyen, and E. Olsen (1992). Copulas and Markov processes. *Illinois Journal of Mathematics* 36(4), 600–642.

[5] Dette, H., K. F. Siburg, and P. A. Stoimenov (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Statist.* 40(1), 21–41.

[6] Durante, F. and C. Sempi (2016). *Principles of Copula Theory*. CRC Press, Boca Raton FL.

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

[8] Fuchs, S. (2014). Multivariate copulas: Transformations, symmetry, order and measures of concordance. *Kybernetika* 50(5), 725–743.

[9] Fuchs, S. and Y. McCord (2019). On the lower bound of Spearman’s footrule. *Depend. Model.* 7, 121–129.

[10] Gamboa, F., P. Gremaud, T. Klein, and A. Lagnoux (2020). Global sensitivity analysis: a new generation of mighty estimators based on rank statistics. *Available at https://arxiv.org/abs/2003.01772*.

[11] Genest, C., J. Nešlehová, and N. Ben Ghorbal (2010). Spearman’s footrule and Gini’s gamma: a review with complements. *J. Nonparametr. Stat.* 22(8), 937–954.

[12] Griessenberger, F., R. Junker, and W. Trutschnig (2021). On a multivariate copula-based dependence measure and its estimation. *Available at https://arXiv:2109.12883v1*.

[13] Han, F. (2021). On extensions of rank correlation coefficients to multivariate spaces. *Bernoulli* 28, 7–11.

[14] Junker, R., F. Griessenberger, and W. Trutschnig (2020). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.* 153, Article ID 107058, 22 pages.

[15] Kallenberg, O. (2002). *Foundations of Modern Probability*. New York.

[16] Karger, D., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. Soria-Auza, N. Zimmermann, H. Linder, and M. Kessler (2017). Climatologies at high resolution for the earth’s land surface areas. *Sci. Data* 4, Article ID 170122.

[17] Karger, D., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. Soria-Auza, N. Zimmermann, H. Linder, and M. Kessler (2018). Data from: Climatologies at high resolution for the earth’s land surface areas. [dataset].
[18] Kasper, T., S. Fuchs, and W. Trutschnig (2021). On weak conditional convergence of bivariate Archimedean and Extreme Value copulas, and consequences to nonparametric estimation. Bernoulli 27, 2217–2240.

[19] Klenke, A. (2008). Wahrscheinlichkeitstheorie. Heidelberg.

[20] Lancaster, H. O. (1963). Correlation and complete dependence of random variables. Ann. Math. Statist. 34(4), 1315–1321.

[21] McDiarmid, C. (1989). On the method of bounded differences. In J. Siemons (Ed.), Surveys in combinatorics, pp. 144–188. Cambridge University Press.

[22] Mroz, T., J. Fernández-Sánchez, S. Fuchs, and W. Trutschnig (2021). Distributions with fixed marginals maximizing the mass of the endograph of a function. Available at https://arxiv.org/abs/1602.05807.

[23] Mroz, T., S. Fuchs, and W. Trutschnig (2021). How simplifying and flexible is the simplifying assumption in pair-copula constructions “analytic answers in dimension three and a glimpse beyond. Electron. J. Statist. 15(1), 1951–1992.

[24] Nelsen, R. B. (2006). An Introduction to Copulas. (Second ed.). New York: Springer.

[25] Shi, H., M. Drton, and F. Han (2021). On Azadkia-Chatterjee’s conditional dependence coefficient. Available at https://arxiv.org/abs/2108.06827v1.

[26] Shih, J.-H. and T. Emura (2021). On the copula correlation ratio and its generalization. J. Multivariate Anal. 182, Article ID 104708.

[27] Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. 384(2), 690–705.

[28] Trutschnig, W. and J. Fernández-Sánchez (2012). Idempotent and multivariate copulas with fractal support. J. Statist. Plann. Inference 142, 3086–3096.