Quantifying and estimating dependence via sensitivity of conditional distributions

Jonathan Ansari[1,a], Patrick B. Langthaler[1,2,b], Sebastian Fuchs[1,c], and Wolfgang Trutschnig[1,d]

[1] Department of Artificial Intelligence and Human Interfaces, University of Salzburg, Austria
[2] Department of Neurology, Christian Doppler Klinik, Paracelsus Medical University, Salzburg, Austria

[a] jonathan.ansari@plus.ac.at
[b] patrickbenjamin.langthaler@stud.plus.ac.at
[c] sebastian.fuchs@plus.ac.at
[d] wolfgang@trutschnig.net

Abstract

Recently established, directed dependence measures for pairs (X,Y) of random variables build upon the natural idea of comparing the conditional distributions of Y given $X=x$ with the unconditional distribution of Y. They assign pairs (X,Y) values in $[0,1]$, where the value is 0 if and only if X,Y are independent, and it is 1 exclusively for Y being a measurable function of X. Here we show that comparing randomly drawn conditional distributions with each other instead or, equivalently, analyzing how sensitive the conditional distribution of Y given $X=x$ is on x, opens the door to constructing novel families of dependence measures Λ_φ induced by general convex functions $\varphi:\mathbb{R}\to\mathbb{R}$, containing, e.g., Chatterjee's coefficient of correlation as special case. After establishing additional useful properties of Λ_φ we focus on continuous (X,Y), translate Λ_φ to the copula setting, consider the L^p-version and establish an estimator which is strongly consistent in full generality. A real data example and a simulation study illustrate the chosen approach and the performance of the estimator. Complementing the aforementioned results, we show how a slight modification of the sensitivity idea underlying Λ_φ can be used to define new measures of explainability generalizing the fraction of explained variance.

Keywords Chatterjee's correlation coefficient, conditional distributions, copula, dependence measure, explainability, Schur order, sensitivity

1 Introduction

In classical statistics, the comparison of two groups/populations is usually done by considering specific parameters like the mean or the variance of the two distributions $\mathbb{P}^{Y|X=x_1}$ and $\mathbb{P}^{Y|X=x_2}$. More precisely, one studies expressions of the form $\Delta(r(x_1),r(x_2))$ or $\Delta(v(x_1),v(x_2))$ for some functional/distance Δ, where $r(x)=\mathbb{E}[Y|X=x]$ denotes the regression function (conditional expectation) and $v(x)=\mathbb{V}(Y|X=x)$ the conditional variance of Y given $X=x$, respectively. Moving from the two-groups setting (modelled by a binary random variable X) to the general case of an arbitrary conditioning variable X, it seems natural to examine how much the conditional distribution of Y given X changes if X changes, i.e., how sensitive the conditional distribution $\mathbb{P}^{Y|X=x}$ of Y given $X=x$ is on x. Intuitively, knowing the extent of sensitivity should provide information on how much information on Y we gain by knowing X. Taking into account the distribution of the conditioning variable X, one natural approach for quantifying sensitivity is to consider the average/expected value of $\Delta(\mathbb{P}^{Y|X=x_1},\mathbb{P}^{Y|X=x_2})$ for a sample x_1,x_2 of X. Writing $\mathbb{P}^X\otimes\mathbb{P}^X$ for the product measure of $\mathbb{P}^X$ with itself, one could therefore consider (non-negative) functionals of the form

$$\Lambda_\Delta(Y|X):=\alpha_\Delta^{-1}\int_{\mathbb{R}^2}\Delta(\mathbb{P}^{Y|X=x_1},\mathbb{P}^{Y|X=x_2})\,\mathrm{d}(\mathbb{P}^X\otimes\mathbb{P}^X)(x_1,x_2), \tag{1}$$

for some normalizing constant $\alpha_\Delta\neq 0$. As a matter of fact, various well-known measures can be represented in the form (1), including the portion of explained variance or Chatterjee's coefficient of rank correlation,

see Section 2. Comparing conditional distributions is certainly not a new idea, see, e.g., [13, 14, 32] in the context of differential privacy. However, to the best of our knowledge, the (recently fast growing) literature on dependence measures (also known as measures of predictability, i.e., measures attaining values in [0,1] and being minimal/maximal exclusively for the case of independence/complete dependence) mainly focuses on comparing conditional and unconditional distributions, so on expressions of the form

$$\int_{\mathbb{R}} \Delta(\mathbb{P}^{Y|X=x}, \mathbb{P}^{Y})\, \mathrm{d}\mathbb{P}^{X}(x)\,, \tag{2}$$

where Δ refers either to a L^2-distance (see, e.g., [2, 5, 7, 15, 17, 35, 37, 11]), the L^1-distance (see, e.g., [20, 24, 38]), some optimal transport cost function (see, e.g., [31]), the maximum mean discrepancy (see, e.g., [22]), or the Wasserstein distance (see, e.g., [39]).

On the one hand, for Δ denoting the L^2-distance between univariate distribution functions, the quantities in (1) and (2) can be shown to coincide, see Example 1 in the next section. On the other hand, in general, these expressions may differ, so considering functionals of the form Λ_Δ leads to novel measures of association and, in particular, to new dependence measures, which are the main focus of this contribution. In a nutshell, the sensitivity idea provides an alternative perspective on known dependence measures and allows for the construction of broad new families of dependence measures.

The remainder of this paper is organized as follows: Section 2 first shows that Chatterjee's dependence measure can be represented in the form Λ_Δ and proposes a family of functionals $\Delta = \Delta_\varphi$ based on convex functions $\varphi : \mathbb{R} \to \mathbb{R}$. Before focusing on these measures for the rest of the paper, we show that using the same sensitivity idea but replacing the conditional distributions by the regression functions (i.e., by the conditional expectations) leads to a class of so-called explainability measures which can be seen as generalization of the well-known fraction of explained variance. Section 3 proves that the afore-mentioned functionals Λ_φ are in fact dependence measures, studies further (invariance and order) properties, then focuses on continuous (X, Y), and translates the dependence measures to the copula setting. Based on these results, in Section 4 a copula-based checkerboard estimator is proposed and shown to be strongly consistent in full generality, i.e., without any smoothness restrictions on the underlying bivariate copulas/distributions. A simulation study illustrating the performance of the estimator, complemented by a real data example in Section 5, rounds off the paper.

Various results presented in the sequel carry over to multivariate X without further ado. However, since there is no straightforward extension of the checkerboard estimator to the multivariate setting in what follows, we always consider real-valued general random variables X, Y defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We will refer to X as the exogenous and to Y as the endogenous variable, and assume that the latter is non-degenerate, i.e., $\mathbb{P}^Y$ is not a one-point distribution. Furthermore we will write $X' \overset{\mathrm{d}}{=} X$ if and only if X' and X have the same distribution, i.e., if $\mathbb{P}^{X'} = \mathbb{P}^X$ holds.

2 General setting and motivating examples

Let $\mathcal{M}$ denote the class of probability measures on $\mathbb{R}$. For studying the sensitivity of conditional distributions we consider functionals $\Delta : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ that may satisfy several desirable and natural properties such as symmetry, i.e., $\Delta(\mu, \nu) = \Delta(\nu, \mu)$ for all $\mu, \nu \in \mathcal{M}$, or positive definiteness, i.e., $\Delta(\mu, \nu) \geq 0$ for all $\mu, \nu \in \mathcal{M}$ with equality if and only if $\mu = \nu$.

Given a bivariate random vector (X, Y) on $(\Omega, \mathcal{A}, \mathbb{P})$, the mapping $\Lambda_\Delta(Y|X)$ is defined according to (1), whereby we implicitly assume that the mapping $(x_1, x_2) \mapsto \Delta\big(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}\big)$ is integrable. Moreover, for the rest of this paper, the normalizing constant α_Δ is defined by

$$\alpha_\Delta := \alpha_\Delta(\mathbb{P}^Y) = \sup\left\{ \int_{\mathbb{R}^2} \Delta(\mathbb{P}^{Y'|X'=x_1}, \mathbb{P}^{Y'|X'=x_2})\, \mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) : (X', Y') \in \mathcal{R}(Y) \right\}, \tag{3}$$

where $\mathcal{R}(Y)$ denotes the class of all random vectors $(X', Y')\colon \Omega \to \mathbb{R}^2$ with $Y' \overset{\mathrm{d}}{=} Y$. Simple expressions for α_Δ will be derived later on. Whenever the supremum is positive and finite, and whenever the integrals in (1) are non-negative, we obviously have that $\Lambda_\Delta \in [0,1]$.

In the sequel, we study properties of the functional Λ_Δ for several classes of Δ and characterize, in particular, the maximal elements determining the constant α_Δ . Notice that for Δ being positive definite, Λ_Δ is able to characterize independence since independence of X and Y means that the conditional distributions $(\mathbb{P}^{Y|X=x})_{x\in\mathbb{R}}$ do not depend on x, which for such Δ is equivalent to the condition $\Lambda_\Delta(Y|X) = 0$.

Remark 2.1. Considering the special case of Δ being a metric on $\mathcal{M}$, the functional $\Lambda_\Delta(Y|X)$ boils down to the average Δ-distance of two randomly (according to $\mathbb{P}^X$) drawn conditional distributions, $\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}$. For a discussion of the case where Δ corresponds to an L^p-metric we refer to Section 3.3.

The subsequent two examples illustrate the broadness of the Λ_Δ-approach according to (1) in the context of quantifying (directed) dependence as well as explainability in terms of the sensitivity of conditional distributions. They show that well-known statistical concepts can be expressed in terms of a functional Λ_Δ with adequately chosen Δ.

Example 1 (Chatterjee's coefficient of correlation). Consider Δ given by

$$\Delta\big(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}\big) := \int_{\mathbb{R}} \Big(\mathbb{P}(Y \geq y|X = x_1) - \mathbb{P}(Y \geq y|X = x_2)\Big)^2 \,\mathrm{d}\mathbb{P}^Y(y)\,.$$

Then Λ_Δ, defined by (1), coincides with Chatterjee's coefficient of correlation (see [7, 11]) since, using change of coordinates, we have

$$\begin{aligned}
\Lambda_\Delta(Y|X) &= \alpha_\Delta^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \Big(\mathbb{P}(Y \geq y|X = x_1) - \mathbb{P}(Y \geq y|X = x_2)\Big)^2 \,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \\
&= 2\,\alpha_\Delta^{-1} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{P}(Y \geq y|X = x)^2 \,\mathrm{d}\mathbb{P}^X(x) - \left(\int_{\mathbb{R}} \mathbb{P}(Y \geq y|X = x) \,\mathrm{d}\mathbb{P}^X(x) \right)^2 \right) \mathrm{d}\mathbb{P}^Y(y) \\
&= \frac{\int_{\mathbb{R}} \mathbb{V}(\mathbb{P}(Y \geq y|X))\,\mathrm{d}\mathbb{P}^Y(y)}{\int_{\mathbb{R}} \mathbb{V}(\mathbb{1}_{\{Y\geq y\}})\,\mathrm{d}\mathbb{P}^Y(y)}.
\end{aligned}$$

Thereby, according to the aforementioned references, again using a change of coordinates and Fubini's theorem, and letting Y_1, Y_2, Y be i.i.d., the normalizing constant α_Δ simplifies to (also see Theorem 3.2)

$$\begin{aligned}
\alpha_\Delta &= \left(2 \int_{\mathbb{R}} \mathbb{V}(\mathbb{1}_{\{Y\geq y\}})\,\mathrm{d}\mathbb{P}^Y(y) \right) \\
&= \left(\int_{\mathbb{R}} \int_{\Omega} \Big(\mathbb{1}_{[y,\infty)}(Y_1(\omega)) - \mathbb{P}(Y_1 \geq y)\Big)^2 + \Big(\mathbb{1}_{[y,\infty)}(Y_2(\omega)) - \mathbb{P}(Y_2 \geq y)\Big)^2 \,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\mathbb{P}^Y(y) \right) \\
&= \left(\int_{\mathbb{R}} \int_{\Omega} \Big(\mathbb{1}_{[y,\infty)}(Y_1(\omega)) - \mathbb{P}(Y_1 \geq y) - \mathbb{1}_{[y,\infty)}(Y_2(\omega)) + \mathbb{P}(Y_2 \geq y)\Big)^2 \,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\mathbb{P}^Y(y) \right) \qquad (4) \\
&= \left(\int_{\mathbb{R}^2} \int_{\mathbb{R}} \Big(\mathbb{1}_{[y,\infty)}(y_1) - \mathbb{1}_{[y,\infty)}(y_2)\Big)^2 \,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2) \right) \\
&= \int_{\mathbb{R}^2} \Delta\left(\delta_{y_1}, \delta_{y_2}\right) \,\mathrm{d}\mathbb{P}^Y \otimes \mathbb{P}^Y(y_1, y_2)\,,
\end{aligned}$$

where δ_x denotes the Dirac measure at $x \in \mathbb{R}$. Hence, by the properties of Chatterjee's coefficient of correlation, the above choice of Δ yields that the functional Λ_Δ is a measure of dependence, i.e., $\Lambda_\Delta(Y|X) \in [0,1]$, $\Lambda_\Delta(Y|X) = 0$ if and only if X and Y are independent, and $\Lambda_\Delta(Y|X) = 1$ exclusively if Y is completely dependent on X, i.e., there exists some measurable (not necessarily monotone) function $f\colon \mathbb{R} \to \mathbb{R}$ with $Y = f(X)$ almost surely.

Motivated by the above example, Section 3 introduces and investigates a class of functionals Λ_Δ, where Δ depends only on the difference of conditional distribution functions rescaled by some measurable function $\varphi\colon \mathbb{R} \to \mathbb{R}$:

$$\Delta_\varphi\big(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}\big) := \int_{\mathbb{R}} \varphi\left(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y)\right) \mathrm{d}\mathbb{P}^Y(y)\,. \tag{5}$$

We will show in Theorem 3.2 that for φ being convex and strictly convex at $0 = \varphi(0)$, the functional $\Lambda_{\Delta_\varphi} =: \Lambda_\varphi$ is a dependence measure (with normalizing constant α_{Δ_φ} according to (4)). Recall that a convex function $\varphi\colon \mathbb{R} \to \mathbb{R}$ is said to be strictly convex at 0 if $(\varphi(-\varepsilon) + \varphi(\varepsilon))/2 > \varphi(0)$ holds for every $\varepsilon > 0$.

Example 2 (Cramér-von-Mises indices; fraction of explained variance). Suppose that Y is square-integrable. Denote by $r(x) := \mathbb{E}(Y|X = x)$ the regression function of Y given $X = x$, and consider

$$\Delta\big(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}\big) \quad := \quad (\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2))^2 = (r(x_1) - r(x_2))^2\,.$$

Then the normalizing constant α_Δ is given by

$$\alpha_\Delta = \left(\int_{\mathbb{R}^2} (y_1 - y_2)^2 \,\mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2)\right) = 2\,\mathbb{V}(Y),$$

and the functional $\Lambda_\Delta(Y|X)$ in (1) coincides with the fraction of explained variance of Y given X, also known as the Sobol index or Cramér-von-Mises index (see [17]), i.e.,

$$\begin{aligned}
\Lambda_\Delta(Y|X) &= \alpha_\Delta^{-1} \int_{\mathbb{R}^2} (\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2))^2 \,\mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \\
&= 2\,\alpha_\Delta^{-1} \left(\int_{\mathbb{R}} r^2(x)\,\mathrm{d}\mathbb{P}^X(x) - \left(\int_{\mathbb{R}} r(x)\,\mathrm{d}\mathbb{P}^X(x)\right)^2\right) \\
&= 2\,\alpha_\Delta^{-1} \left(\mathbb{E}((r \circ X)^2) - (\mathbb{E}(r \circ X))^2\right) \\
&= 2\,\alpha_\Delta^{-1}\,\mathbb{V}(r \circ X) = \frac{\mathbb{V}(\mathbb{E}(Y|X))}{\mathbb{V}(Y)} = \frac{\int_{\mathbb{R}} (\mathbb{E}(Y|X = x) - \mathbb{E}(Y))^2 \,\mathrm{d}\mathbb{P}^X(x)}{\mathbb{V}(Y)}\,.
\end{aligned}$$

Recall that in this case $\Lambda_\Delta(Y|X) \in [0, 1]$ and we have $\Lambda_\Delta(Y|X) = 0$ if and only if $\mathbb{V}(r \circ X) = 0$ (which is the case if X and Y are independent, but not vice versa), and $\Lambda_\Delta(Y|X) = 1$ if and only if Y is completely dependent on X.

In the case of an affine regression function $r(x) = ax + b$, the quantity $\Lambda_\Delta(Y|X)$ then coincides with the squared Pearson correlation ρ of X and Y, i.e.,

$$\Lambda_\Delta(Y|X) = \frac{(\mathrm{Cov}(X, Y))^2}{\mathbb{V}(X)\mathbb{V}(Y)} = \rho^2(X, Y)\,.$$

As demonstrated in Example 2, $\Lambda_\Delta(Y|X)$ coincides with the L^2-Cramér-von-Mises index, i.e., it coincides with the expected squared distance between the conditional and the unconditional expectation. This motivates to investigate another subclass of functionals Λ_Δ , where Δ now depends only on the difference of conditional expectations, weighted again by some measurable function $\psi\colon \mathbb{R} \to \mathbb{R}$:

$$\Delta_\psi\big(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}\big) := \psi\left(\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2)\right). \tag{6}$$

Pursuing this idea leads to a novel class of Cramer-von-Mises indices solely based on the sensitivity of the conditional expectations. For some specific choices of ψ according to (6), the associated functional $\Lambda_{\Delta_\psi} =: \Lambda_\psi$ in (1) relates the variability of the conditional distribution to the variability of the unconditional distribution as in (2), see, e.g., [16] for several Gini-type measures of risk and variability.

For the special case where $\Delta = \Delta_\psi$ is of the form (6), the functional Λ_ψ is given by

$$\Lambda_\psi(Y|X) := \alpha_\psi^{-1} \int_{\mathbb{R}^2} \psi\left(\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2)\right) \, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \tag{7}$$

with normalizing constant

$$\alpha_\psi = \sup\left\{ \int_{\mathbb{R}^2} \psi\left(\mathbb{E}(Y'|X' = x_1) - \mathbb{E}(Y'|X' = x_2)\right) \, \mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) : (X', Y') \in \mathcal{R}(Y) \right\},$$

whenever the supremum is finite and positive.

The main focus of this article is to introduce new classes of dependence measures via sensitivity of conditional distributions (as studied in the next section). Underlining the broadness of the sensitivity idea, however, we first quickly formulate a general result on measures of explainability constructed via (7). To increase readability the proof of the resulting theorem is moved to the appendix. Notice, however, that (i) for the proof convexity of ψ and strict convexity in $0 = \psi(0)$ is key (also see Remark .1) and that (ii) in the subsequent theorem ψ is also allowed to attain negative values.

Theorem 2.2 (Measures of explainability)**.** *Suppose that $\psi(2Y)$ and $\psi(-2Y)$ are integrable and that $\psi\colon \mathbb{R} \to \mathbb{R}$ is convex and strictly convex at 0 with $\psi(0) = 0$. Then the constant α_ψ is given by*

$$\alpha_\psi = \int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right) \, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2)\,. \tag{8}$$

Furthermore, $\Lambda_\psi(Y|X)$, defined according to (7), is a measure of explainability, i.e., it fulfills the following three properties:

(i) $\Lambda_\psi(Y|X) \in [0, 1]$.

(ii) $\Lambda_\psi(Y|X) = 0$ *if and only if* $\mathbb{V}(\mathbb{E}(Y|X)) = 0$*, i.e.,* $\mathbb{E}(Y|X) = \mathbb{E}[Y]$ *almost surely.*

(iii) $\Lambda_\psi(Y|X) = 1$ *if and only if Y is completely dependent on X, i.e.,* $\mathbb{E}(Y|X) = Y$ *almost surely.*

3 A new class of dependence measures

For the rest of the paper we now focus on the functionals Λ_φ according to (5) and derive simple sufficient conditions for φ ensuring that $\Lambda_\Delta(Y|X)$ defined by (1) is a dependence measure. Following the recent literature, we refer to κ as a *(directed) dependence measure* (frequently also referred to as a measure of predictability) for bivariate random vectors (X, Y) if it satisfies the following axioms, see, e.g., [2, 6, 7, 10, 21, 36, 38] and the references therein:

(A1) $\kappa(Y|X) \in [0, 1]$.

(A2) $\kappa(Y|X) = 0$ if and only if Y and X are independent.

(A3) $\kappa(Y|X) = 1$ if and only if Y is completely dependent on X.

Due to the axioms (A2) and (A3), the values 0 and 1 of a measure of predictability κ have a clear interpretation. The meaning of κ attaining a specific value in the interval $(0, 1)$, however, is not that unequivocal—nevertheless, considering that the values in $[0, 1]$ are naturally ordered, they offer a clear comparative interpretation: the higher the value of κ the higher the extent of dependence of Y on X. This aspect motivates the study of dependence orderings $\prec$ that are compatible with κ in the following sense:

(P1) *Monotonicity*: If $(X, Y) \prec (X', Y')$, then $\kappa(Y|X) \leq \kappa(Y'|X')$.

It is further desirable—particularly in the context of estimation—to have an understanding of the measures' behavior in terms of convergence:

(P2) *Continuity*: If $(X_n, Y_n)_{n\in\mathbb{N}}$ converges (in some sense) to (X, Y), then $\lim_{n\to\infty} \kappa(Y_n|X_n) = \kappa(Y|X)$.

Moreover, the following invariance properties seem desirable and will be studied in the sequel:

(P3) *Invariance of* X: For bijective h, we have $\kappa(Y|X) = \kappa(Y|h(X))$.

(P4) *Version-Invariance*: If $(X, Y) \overset{\mathrm{d}}{=} (X', Y')$, then $\kappa(Y|X) = \kappa(Y'|X')$.

Property (P3) reflects the idea that the degree of predictability of Y through X only depends on the information represented by the σ-algebra generated by X. Note that property (P4) is also misleadingly referred to as law-invariance in the context of risk measures.

3.1 The class Λ_φ

For the special case where Δ is of the form (5), the functional Λ_φ is given by

$$\Lambda_\varphi(Y|X) := \alpha_\varphi^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y)) \, \mathrm{d}\mathbb{P}^Y(y) \, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \tag{9}$$

with normalizing constant α_φ, according to (3), defined by

$$\alpha_\varphi := \sup \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(F_{Y'|X'=x_1}(y) - F_{Y'|X'=x_2}(y)) \, \mathrm{d}\mathbb{P}^{Y'}(y) \, \mathrm{d}(\mathbb{P}^{Y'} \otimes \mathbb{P}^{Y'})(y_1, y_2) : (X', Y') \in \mathcal{R}(Y) \right\}, \tag{10}$$

whenever the supremum exists and is positive.

For deriving sufficient conditions on φ for Λ_φ in (9) to characterize independence and complete directed dependence, the following simple lemma is key.

Lemma 3.1. *Let U be a real-valued random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{B}$ and $\mathcal{C}$ be σ-algebras with $\mathcal{B} \subseteq \mathcal{C} \subseteq \mathcal{A}$. Then $\mathbb{E}[U|\mathcal{B}] \leq_{cx} \mathbb{E}[U|\mathcal{C}]$, where $\leq_{cx}$ denotes the convex order.*

Proof. Let φ be a convex function. Then, whenever the following expectations exist, one has

$$\mathbb{E}[\varphi(\mathbb{E}[U|\mathcal{B}])] = \mathbb{E}[\varphi(\mathbb{E}[\mathbb{E}[U|\mathcal{C}]|\mathcal{B}])] \leq \mathbb{E}[\mathbb{E}[\varphi(\mathbb{E}[U|\mathcal{C}])|\mathcal{B}]] = \mathbb{E}[\varphi(\mathbb{E}[U|\mathcal{C}])] \tag{11}$$

applying the tower property and Jensen's inequality for conditional expectation. □

Having this, we can now formulate the following main result of this section.

Theorem 3.2 (Measures of directed dependence)**.** *Assume that $\varphi : [-1, 1] \to \mathbb{R}$ is convex and strictly convex at 0 with $\varphi(0) = 0$ and that Y is non-degenerate. Then the normalizing constant α_φ in (10) is given by*

$$\alpha_\varphi = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}\right) \mathrm{d}\mathbb{P}^Y(y) \, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2) \tag{12}$$

and $\Lambda_\varphi(Y|X)$ defined according to (9) is a dependence measure, i.e., $\kappa = \Lambda_\varphi$ satisfies (A1)-(A3).

Proof. (A1): For every $y \in \mathbb{R}$ and every $c \in [0, 1]$, applying Lemma 3.1 directly yields

$$\int_\Omega \varphi\left(\mathbb{E}[\mathbb{1}_{\{Y\leq y\}}|X] - c\right) \mathrm{d}\mathbb{P} \leq \int_\Omega \varphi\left(\mathbb{E}[\mathbb{1}_{\{Y\leq y\}}|(X, Y)] - c\right) \mathrm{d}\mathbb{P} = \int_\Omega \varphi\left(\mathbb{1}_{\{Y\leq y\}} - c\right) \mathrm{d}\mathbb{P}.$$

Hence, using disintegration, Jensen's inequality, and Fubini's theorem, it follows that

$$
\begin{aligned}
0 = \varphi(0) &= \int_{\mathbb{R}} \mathbb{P}(Y \leq y) - \mathbb{P}(Y \leq y)\, \mathrm{d}\mathbb{P}^Y(y) \\
&= \varphi\left(\int_{\mathbb{R}} \int_{\mathbb{R}^2} (F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\, \mathrm{d}(P^X \otimes \mathbb{P}^X)(x_1, x_2)\, \mathrm{d}\mathbb{P}^Y(y) \right) \\
&\leq \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y) \right)\, \mathrm{d}\mathbb{P}^X(x_1)\, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^Y)(x_2, y) \\
&\leq \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - F_{Y|X=x_2}(y) \right)\, \mathrm{d}\mathbb{P}^Y(y_1)\, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^Y)(x_2, y) \\
&= \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - F_{Y|X=x_2}(y) \right)\, \mathrm{d}\mathbb{P}^X(x_2)\, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y) \\
&\leq \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}} \right)\, \mathrm{d}\mathbb{P}^Y(y_2)\, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y) \\
&= \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}} \right)\, \mathrm{d}\mathbb{P}^Y(y)\, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2) \\
&= \sup \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(F_{Y'|X'=x_1}(y) - F_{Y'|X'=x_2}(y))\, \mathrm{d}\mathbb{P}^{Y'}(y)\, \mathrm{d}(\mathbb{P}^{Y'} \otimes \mathbb{P}^{Y'})(y_1, y_2) : (X', Y') \in \mathcal{R}(Y) \right\} \\
&= \alpha_\varphi \,,
\end{aligned}
$$

where the supremum is attained whenever $Y = f(X)$ almost surely for some measurable f. This implies (12) and yields $\Lambda_\varphi(Y|X) \in [0, 1]$.

(A2): X and Y are independent, if and only if $F_{Y|X=x}(\cdot) = F_Y(\cdot)$ for $\mathbb{P}^X$-almost all x. The latter, however, is equivalent to the inner integral in (9) being zero $\mathbb{P}^X \otimes \mathbb{P}^X$-almost everywhere since φ is strictly convex at 0 and $\varphi(0) = 0$ holds.

(A3): Suppose that $Y = f(X)$ almost surely for some measurable $f : \mathbb{R} \to \mathbb{R}$. Then we obviously have $F_{Y|X=x}(y) = \mathbb{1}_{\{f(x) \leq y\}}$, so applying Fubini's theorem and change of coordinates yields

$$
\begin{aligned}
\alpha_\varphi\, \Lambda_\varphi(Y|X) &= \int_{\mathbb{R}} \int_{\mathbb{R}^2} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2)\, \mathrm{d}\mathbb{P}^Y(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^2} \varphi(\mathbb{1}_{\{f(x_1) \leq y\}} - \mathbb{1}_{\{f(x_2) \leq y\}})\, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2)\, \mathrm{d}\mathbb{P}^Y(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^2} \varphi(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}})\, \mathrm{d}(\mathbb{P}^{f \circ X} \otimes \mathbb{P}^{f \circ X})(y_1, y_2)\, \mathrm{d}\mathbb{P}^Y(y) \\
&= \alpha_\varphi .
\end{aligned} \tag{13}
$$

Since Y is non-degenerate, property (A2) and (12) imply that α_φ is positive and finite, so $\Lambda_\varphi(Y|X) = 1$ holds.

It remains to prove the reverse implication, i.e., that $\Lambda_\varphi(Y|X) = 1$ implies $Y = f(X)$ almost surely for some f, which can be done as follows: First, we show that the condition $\Lambda_\varphi(Y|X) = 1$ implies the existence of some $\mathbb{P}^Y$-null set N such that for fixed $y \in N^c$ there exists a $\mathbb{P}^X$-null set N_y with

$$
F_{Y|X=x}(y) \in \{0, 1\} \quad \text{for all } x \in N_y^c \,. \tag{14}
$$

We then prove that there exists some $\mathbb{P}^X$-null set N_* fulfilling

$$
F_{Y|X=x}(y) \in \{0, 1\} \quad \text{for all } y \in \mathbb{R} \text{ and for all } x \in N_*^c \,. \tag{15}
$$

Having this, proceeding like in the proofs of [5, Theorem 9.2. (iii)] or [38, Lemma 12] yields complete dependence of Y on X, which completes the proof. In order to show (14), fix $y \in \mathbb{R}$ and consider $a : \mathbb{R} \to \mathbb{R}$, defined by $a(x) = \mathbb{P}(Y \leq y \mid X = x)$. We first establish an upper bound for the mapping $(x_1, x_2) \mapsto \varphi\big(a(x_1) - a(x_2)\big)$ and procced as follows: Setting

$$\begin{aligned} c_1 &:= a(x_1) - \min\{a(x_1), a(x_2)\} \in [0,1] \\ c_{-1} &:= a(x_2) - \min\{a(x_1), a(x_2)\} \in [0,1] \\ c_0 &:= 1 - a(x_1) - a(x_2) + 2\min\{a(x_1), a(x_2)\} \in [0,1]\,, \end{aligned}$$

we obviously have $c_1 + c_{-1} + c_0 = 1$, so using convexity yields

$$\begin{aligned} &\varphi(a(x_1) - a(x_2)) \\ &= \varphi\big(1 \cdot (a(x_1) - \min\{a(x_1), a(x_2)\}) - 1 \cdot (a(x_2) - \min\{a(x_1), a(x_2)\}) + 0 \cdot c_0\big) \\ &\leq \varphi(1)(a(x_1) - \min\{a(x_1), a(x_2)\}) + \varphi(-1)(a(x_2) - \min\{a(x_1), a(x_2)\}) + \underbrace{\varphi(0)}_{=0} \cdot c_0 \\ &= \varphi(1)\, a(x_1) + \varphi(-1)\, a(x_2) - \underbrace{(\varphi(1) + \varphi(-1))}_{>0} \underbrace{\min\{a(x_1), a(x_2)\}}_{\geq a(x_1)a(x_2)} \\ &\leq \varphi(1)\, a(x_1) + \varphi(-1)\, a(x_2) - (\varphi(1) + \varphi(-1))\, a(x_1)a(x_2) \\ &= \varphi(1)\, a(x_1)(1 - a(x_2)) + \varphi(-1)\, a(x_2)(1 - a(x_1)) \\ &= \int_{\mathbb{R}^2} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}\right) \mathrm{d}(\mathbb{P}^{Y|X=x_1} \otimes \mathbb{P}^{Y|X=x_2})(y_1, y_2). \end{aligned} \tag{16}$$

Integrating both sides over $\mathbb{R}^2$ with respect to $\mathbb{P}^X \otimes \mathbb{P}^X$ and using disintegration, it follows that

$$\begin{aligned} &\int_{\mathbb{R}^2} \varphi(a(x_1) - a(x_2))\, \mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1, x_2) \\ &\leq \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}\right) \mathrm{d}(\mathbb{P}^{Y|X=x_1} \otimes \mathbb{P}^{Y|X=x_2})(y_1, y_2)\, \mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1, x_2) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi\left(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}\right) \mathrm{d}\mathbb{P}^{Y|X=x_1}(y_1)\, \mathrm{d}\mathbb{P}^X(x_1)\, \mathrm{d}\mathbb{P}^{Y|X=x_2}(y_2)\, \mathrm{d}\mathbb{P}^X(x_2) \\ &= \int_{\mathbb{R}^2} \varphi(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}})\, \mathrm{d}\mathbb{P}^Y \otimes \mathbb{P}^Y(y_1, y_2). \end{aligned} \tag{17}$$

Consequently, if equality holds in (17), then for $\mathbb{P}^X \otimes \mathbb{P}^X$-almost all $(x_1, x_2) \in \mathbb{R}^2$ inequality (16) becomes an equality. In particular, there exists some Borel set $G \subseteq \mathbb{R}^2$ with $\mathbb{P}^X \otimes \mathbb{P}^X(G) = 1$ such that

$$\min\{a(x_1), a(x_2)\} = a(x_1)a(x_2)$$

holds for all $(x_1, x_2) \in G$. Considering that the latter is equivalent to the fact that $a(x_1) \in \{0,1\}$ or $a(x_2) \in \{0,1\}$ for all $(x_1, x_2) \in G$, we conclude that $a(x) \in \{0,1\}$ for $\mathbb{P}^X$-almost every $x \in \mathbb{R}$. Summing up, for every fixed $y \in \mathbb{R}$, we have shown that equality in (17) implies that $a(x) = \mathbb{P}(Y \leq y \mid X = x) = F_{Y|X=x}(y) \in \{0,1\}$ for $\mathbb{P}^X$-almost every $x \in \mathbb{R}$.

Now, to prove (14), suppose that $\Lambda_\varphi(Y|X) = 1$, i.e., that we have

$$\begin{aligned} &\int_{\mathbb{R}} \int_{\mathbb{R}^2} \varphi(a(x_1) - a(x_2))\, \mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1, x_2)\, \mathrm{d}\mathbb{P}^Y(y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^2} \varphi(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}\, \mathrm{d}\mathbb{P}^Y \otimes \mathbb{P}^Y(y_1, y_2)\, \mathrm{d}\mathbb{P}^Y(y)\,. \end{aligned} \tag{18}$$

Then, due to (17), there exists some Borel set $E \subseteq \mathbb{R}$ with $\mathbb{P}^Y(E) = 1$ such that for every $y \in E$ we have equality in (17). Due to the previous step, for $N := E^c$ (14) holds true for all $y \in E$, i.e., for every $y \in E$ there exists a $\mathbb{P}^X$-null set N_y with $F_{Y|X=x}(y) \in \{0,1\}$ for all $x \in N_y^c$.

In order to prove (15) first note that by disintegration we also have $\mathbb{P}^{Y|X=x}(E) = 1$ for $\mathbb{P}^X$-almost every $x \in \mathbb{R}$. Hence we may proceed as follows: Define the set $N \in \mathcal{B}(\mathbb{R}^2)$ by

$$N := \{(x,y) \in \mathbb{R}^2 : F_{Y|X=x}(y) \in (0,1)\}$$

and suppose that $\mathbb{P}^{(X,Y)}(N) > 0$. Setting $N_x := \{y \in \mathbb{R} : (x,y) \in N\}$ disintegration yields

$$0 < \mathbb{P}^{(X,Y)}(N) = \int_{\mathbb{R}} \mathbb{P}^{Y|X=x}(N_x)\, \mathrm{d}\mathbb{P}^X(x),$$

so the set $O \subseteq \mathbb{R}$, given by $O = \{x \in \mathbb{R} : \mathbb{P}^{Y|X=x}(N_x) > 0 \text{ and } \mathbb{P}^{Y|X=x}(E) = 1\}$ fulfills $\mathbb{P}^X(O) > 0$. For every $x \in O$ we therefore have $N_x \cap E \neq \emptyset$, which is absurd. Consequently, $\mathbb{P}^{(X,Y)}(N) = 0$, whence $\mathbb{P}^{(X,Y)}(N^c) = 1$ holds. Disintegrating once more yields a $\mathbb{P}^X$-null set N_* such that for all $x \in N_*^c$ the conditional distribution function $y \mapsto F_{Y|X=x}(y)$ only attains values in $\{0,1\}$, which proves (15). □

Remark 3.3. (a) If $\varphi(x) = x^2$, then Λ_φ in (9) reduces to Chatterjee's coefficient of correlation, see Example 1. However, in general, the comparison of conditional distributions via Λ_φ does not boil down to comparing conditional and unconditional distributions, see Example 4. In other words, Theorem 3.2 provides broad new classes of dependence measures.

(b) In Theorem 3.2 the assumption of strict convexity of φ in 0 is, on the one hand, necessary for the characterization of independence. On the other hand, it is also used in the proof that $\Lambda_\varphi(Y|X) = 1$ implies complete dependence. Note that Theorem 3.2 only requires that φ is strictly convex at 0 but not necessarily strictly convex on the full interval $[-1,1]$. Hence, each of the functions $x \mapsto \max\{x,0\}$, $x \mapsto \exp(cx) - 1$ for $c \neq 0$, and $x \mapsto |x|^p$, $p \geq 1$, satisfy the assumptions of Theorem 3.2. Further, as it is the case for ψ in Theorem 2.2, also φ may attain negative values. If Y has no point masses (or, equivalently, if Y has continuous distribution function) then a straightforward calculation shows $\alpha_\varphi = (\varphi(-1) + \varphi(1))/6$. Hence, for $\varphi(x) = |x|^p$, $p \geq 1$, the constant α_φ is $1/3$ for every choice of p.

(c) The second part of the proof of Theorem 3.2 can also be established along the lines of the proof of the second part of Theorem 2.2(iii). However, we opted for a more constructive proof providing additional insight on properties of Λ_φ. Instead of the denseness argument here we use the fact that for independent X_1 and X_2 with $X_1 \overset{\mathrm{d}}{=} X_2 \overset{\mathrm{d}}{=} X$, the random variable $F_{Y|X=X_1}(y) - F_{Y|X=X_2}(y)$ obtained from (13) is bounded on $[-1,1]$ for every $y \in \mathbb{R}$ and the support of the extremal elements with respect to convex order is $\mathbb{P}^Y$-almost surely the set $\{-1,0,1\}$.

3.2 Additional properties of Λ_φ

Convexity of the function φ yields the so-called *data processing inequality*, which states that applying a transformation to the explanatory variable X cannot increase the extent of dependence of Y on X (see [8] for a detailed discussion). Note that strict convexity at 0 is not required for the following results.

Proposition 3.4 (Data processing inequality)**.** *Assume that φ is convex on $[-1,1]$. Then Λ_φ, defined according to eq.* (9)*, fulfills the data processing inequality, i.e.,*

$$\Lambda_\varphi(Y|Z) \leq \Lambda_\varphi(Y|X)$$

holds for all X, Y, Z such that Y and Z are conditionally independent given X. In particular,

$$\Lambda_\varphi(Y|h(X)) \leq \Lambda_\varphi(Y|X)$$

for all measurable functions $h : \mathbb{R} \to \mathbb{R}$.

Proof. Assume that Y and Z are conditionally independent given X. Proceeding as in the proof of the second assertion of Lemma 3.5 in [20], and using disintegration, it follows that for every fixed $y \in \mathbb{R}$, we have

$$F_{Y|Z=z}(y) = \int_{\mathbb{R}} F_{Y|Z=z,X=x}(y)\,\mathrm{d}\mathbb{P}^{X|Z=z}(x) = \int_{\mathbb{R}} F_{Y|X=x}(y)\,\mathrm{d}\mathbb{P}^{X|Z=z}(x)$$

for $\mathbb{P}^Z$-almost all $z \in \mathbb{R}$. Hence, using convexity of φ, Jensen's inequality, and disintegration once more yields

$$\begin{aligned}
&\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y|Z=z_1}(y) - F_{Y|Z=z_2}(y))\,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^Z\otimes\mathbb{P}^Z)(z_1,z_2)\\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi\left(\int_{\mathbb{R}} F_{Y|X=x_1}(y)\,\mathrm{d}\mathbb{P}^{X|Z=z_1}(x_1) - \int_{\mathbb{R}} F_{Y|X=x_2}(y)\,\mathrm{d}\mathbb{P}^{X|Z=z_2}(x_2)\right)\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^Z\otimes\mathbb{P}^Z)(z_1,z_2)\\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi\left(\int_{\mathbb{R}^2} F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y)\,\mathrm{d}(\mathbb{P}^{X|Z=z_1}\otimes\mathbb{P}^{X|Z=z_2})(x_1,x_2)\right)\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^Z\otimes\mathbb{P}^Z)(z_1,z_2)\\
&\le \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^{X|Z=z_1}\otimes\mathbb{P}^{X|Z=z_2})(x_1,x_2)\,\mathrm{d}(\mathbb{P}^Z\otimes\mathbb{P}^Z)(z_1,z_2)\\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}\mathbb{P}^{(X,Z)}(x_1,z_1)\,\mathrm{d}\mathbb{P}^{(X,Z)}(x_2,z_2)\\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^X\otimes\mathbb{P}^X)(x_1,x_2)\,,
\end{aligned}$$

which completes the proof. □

As a consequence of the data processing inequality, Λ_φ satisfies the property of *self-equitability* introduced in [26] (see also [12]), which roughly states that for (X,Y) in a regression setting $Y = f(X)+\varepsilon$, the function Λ_φ should only depend on the distribution of the noise ε and not on the specific form of the function f, see [36].

Corollary 3.5 (Self-equitability)**.** *Assume that φ is convex on $[-1,1]$. Then Λ_φ defined according to* (9) *is* self-equitable*, i.e.,*

$$\Lambda_\varphi(Y|h(X)) = \Lambda_\varphi(Y|X)$$

for all (X,Y) and all measurable functions h such that Y and X are conditionally independent given $h(X)$.

The following result provides additional useful invariance properties for Λ_φ. Note that for these properties to hold we do not even need φ to be convex.

Proposition 3.6 (Invariances)**.** Λ_φ *satisfies the invariance properties (P3) and (P4), i.e.,*

(i) $\Lambda_\varphi(Y|X) = \Lambda_\varphi(Y|h(X))$ *for all bijective functions* $h:\mathbb{R}\to\mathbb{R}$*,*

(ii) $\Lambda_\varphi(Y|X) = \Lambda_\varphi(Y'|X')$ *whenever* $(X,Y) \overset{\mathrm{d}}{=} (X',Y')$*.*

Additionally,

(iii) $\Lambda_\varphi(Y|X) = \Lambda_\varphi(g(Y)|X)$ *for all strictly increasing functions* $g:\mathbb{R}\to\mathbb{R}$*.*

Proof. (i): Let (X',Y') be an independent copy of (X,Y). Since h is bijective, the σ-algebras generated by X and $h(X)$ coincide. Using change of coordinates it therefore follows that

$$\begin{aligned}
\Lambda_\varphi(Y|X) &= \int_{\mathbb{R}}\int_{\Omega} \varphi\left(\mathbb{E}[\mathbb{1}_{\{Y\le y\}}|X](\omega) - \mathbb{E}[\mathbb{1}_{\{Y'\le y\}}|X'](\omega)\right)\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\mathbb{P}^Y(y)\\
&= \int_{\mathbb{R}}\int_{\Omega} \varphi\left(\mathbb{E}[\mathbb{1}_{\{Y\le y\}}|h(X)](\omega) - \mathbb{E}[\mathbb{1}_{\{Y'\le y\}}|h(X')](\omega)\right)\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\mathbb{P}^Y(y) = \Lambda_\varphi(Y|h(X))\,.
\end{aligned}$$

Statement (ii) is trivial. (iii): Since there exists some measurable function $k\colon \mathbb{R}^2 \to [0,1]$ such that $F_{Y|X=x}(y) = k(x, F_Y(y))$ for $\mathbb{P}^X$-almost all $x \in \mathbb{R}$ and for all $y \in \mathbb{R}$, see [3, Theorem 2.2], using change of coordinates and that g is strictly increasing, it follows that (λ denoting the Lebesgue measure, F^- the quasi-inverse of a distribution function F)

$$\begin{aligned}\Lambda_\varphi(Y|X) &= \int_{\mathbb{R}^2}\int_{(0,1)} \varphi\left(k\left(x_1, F_Y \circ F_Y^-(u)\right) - k(x_2, F_Y \circ F_Y^-(u))\right) \mathrm{d}\lambda(u)\,\mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1,x_2)\\ &= \int_{\mathbb{R}^2}\int_{(0,1)} \varphi\left(k\left(x_1, F_{g(Y)} \circ F_{g(Y)}^-(u)\right) - k\left(x_2, F_{g(Y)} \circ F_{g(Y)}^-(u)\right)\right) \mathrm{d}\lambda(u)\,\mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1,x_2)\\ &= \Lambda_\varphi(g(Y)|X)).\end{aligned}$$

□

Considering that $\Lambda(Y|X)$ according to Theorem 3.6(i) is invariant with respect to bijective transformations of X, in order to establish some general ordering results for Λ_φ, we need to work with rearrangement-invariant orders. To this end, for integrable functions $f, g\colon (0,1) \to \mathbb{R}$ with $\int_0^1 f(t)\,\mathrm{d}\lambda(t) = \int_0^1 g(t)\,\mathrm{d}\lambda(t)$ consider the Schur order $\prec_S$, defined by

$$f \prec_S g \quad :\Longleftrightarrow \quad \int_0^x f^*(t)\,\mathrm{d}t \le \int_0^x g^*(t)\,\mathrm{d}t \quad \text{for all } x \in (0,1)\,, \tag{19}$$

where h^* denotes the decreasing rearrangement of a function $h\colon (0,1) \to \mathbb{R}$, i.e., the λ-almost everywhere unique decreasing function h^* such that $\lambda(\{h^* \le t\}) = \lambda(\{h \le t\})$ for all $t \in \mathbb{R}$. Denote by $q_Z(u) := \inf\{z \in \mathbb{R} :|\, F_Z(z) \ge u\} = F_Z^-(u)$ the quasi-inverse of the distribution function F_Z of Z. Then, for Y and Y' having continuous distribution functions, the Schur order $(Y|X) \le_S (Y'|X')$ for conditional distributions is defined by

$$\mathbb{E}\left[\mathbb{1}_{\{Y \le q_Y(v)\}} \mid X = q_X(\cdot)\right] \le_S \mathbb{E}\left[\mathbb{1}_{\{Y' \le q_{Y'}(v)\}} \mid X' = q_{X'}(\cdot)\right] \quad \text{for all } v \in (0,1)\,, \tag{20}$$

see [2] as well as [1, 3, 34, 36]. Hence, the Schur order for conditional distributions compares the variability of conditional distribution functions with respect to the conditioning variable.

For bivariate random vectors (U,V) and (U',V'), denote by $(U,V) \le_{lo} (U',V')$ the lower orthant order, i.e., the pointwise comparison of the associated distribution functions. Further, the bivariate random vector (U,V) is said to be stochastically increasing (SI) if the conditional distribution $V|U=u$ is stochastically increasing in u, i.e., $\mathbb{E}[f(V)|U=u]$ is increasing in u for all increasing functions f such that the expectations exist. If (U',V') is SI and if $U \overset{\mathrm{d}}{=} U'$ and $V \overset{\mathrm{d}}{=} V'$, then $(V|U) \le_S (V'|U')$ implies $(U,V) \le_{lo} (U',V')$. If additionally (U,V) is SI, then $(V|U) \le_S (V'|U')$ and $(U,V) \le_{lo} (U',V')$ are equivalent, see [3, Lemma 3.16]. The following result provides simple conditions for monotonicity of Λ_φ.

Proposition 3.7 (Monotonicity)**.** *Suppose that $(Y|X) \le_S (Y'|X')$ and that Y, Y' have continuous distribution functions. Then $\Lambda_\varphi(Y|X) \le \Lambda_\varphi(Y'|X')$ for all convex functions φ such that the integrals exist.*

Proof. By the Hardy-Littlewood-Polya theorem (see, e.g., [33, Theorem 3.21] we have that $(Y|X) \le_S (Y'|X')$ is equivalent to

$$\int_0^1 \phi(F_{Y|X=q_X(u)}(q_Y(v)))\,\mathrm{d}\lambda(u) \le \int_0^1 \phi(F_{Y'|X'=q_{X'}(u)}(q_{Y'}(v)))\,\mathrm{d}\lambda(u)$$

for all v and for all convex ϕ such that the integrals exist. Since the functions $z \mapsto \varphi(z-c)$ and $z \mapsto \varphi(c-z)$ are convex whenever φ is convex, using Fubini's theorem and change of coordinates repeatedly, it follows

that

$$\begin{aligned}
&\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y))\,\mathrm{d}\mathbb{P}^Y(y)\,\mathrm{d}(\mathbb{P}^X\otimes\mathbb{P}^X)(x_1,x_2)\\
&= \int_{(0,1)}\int_{(0,1)}\int_{(0,1)} \varphi\left(F_{Y|X=q_X(u_1)}(q_Y(v)) - F_{Y|X=q_X(u_2)}(q_Y(v))\right)\,\mathrm{d}\lambda(u_1)\,\mathrm{d}\lambda(v)\,\mathrm{d}\lambda(u_2)\\
&\le \int_{(0,1)}\int_{(0,1)}\int_{(0,1)} \varphi\left(F_{Y'|X'=q_{X'}(u_1)}(q_{Y'}(v)) - F_{Y'|X'=q_{X'}(u_2)}(q_{Y'}(v))\right)\,\mathrm{d}\lambda(v)\,\mathrm{d}\lambda(u_1)\,\mathrm{d}\lambda(u_2)\\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi(F_{Y'|X'=x_1}(y) - F_{Y'|X'=x_2}(y))\,\mathrm{d}\mathbb{P}^{Y'}(y)\,\mathrm{d}(\mathbb{P}^{X'}\otimes\mathbb{P}^{X'})(x_1,x_2)\,,
\end{aligned}$$

which implies $\Lambda_\varphi(Y|X) \le \Lambda_\varphi(Y'|X')$. □

Remark 3.8. (a) Minimal and maximal elements with respect to the Schur order for conditional distributions characterize independence and complete directed dependence, respectively, see [2, Theorem 3.5]. Hence, as a consequence of Theorem 3.7, if Y is completely dependent on X and if φ is convex, then $\alpha_\varphi \in (0,\infty)$ and $\Lambda_\varphi(Y|X) = 1$. If additionally φ is strictly convex at 0 and if $\varphi(0) = 0$, then Theorem 3.2 provides the reverse implication, i.e., that $\Lambda_\varphi(Y|X) = 1$ implies that $Y = f(X)$ almost surely.

(b) Various well-known (sub-)families of bivariate distributions are stochastically increasing and increasing with respect to the lower orthant order (implying that the associated conditional distributions are increasing with respect to the Schur order for conditional distributions). Examples include the bivariate normal distribution with respect to the non-negative correlation parameter, several Archimedean copula families such as the Clayton, Gumbel, or Frank family, or various extreme value copula families, see [4].

3.3 Continuous setting and L^p-version

From now on we will assume that (X,Y) has continuous bivariate distribution function since continuity is necessary for having uniqueness in Sklar's theorem (see [30]) and both, for deriving continuity results of the functional Λ_φ and for the estimation procedure studied in the next section. Furthermore, in the sequel (following [24]) we will write conditional distributions as Markov kernels, i.e., $\mathbb{P}^{Y|X=x}(\cdot) = K(x,\cdot)$, let C denote the (unique) copula underlying (X,Y) and $K_C(\cdot,\cdot)$ the Markov kernel of C. Notice that, according to [28, Lemma 1] (also see [3, Theorem 2.2]), starting with a Markov kernel $K_C(\cdot,\cdot)$ and setting

$$K(x,(-\infty,y]) := K_C(F_X(x),[0,F_Y(y)])$$

yields a (version of the) Markov kernel of (X,Y). For continuous marginal distribution functions, the functional Λ_φ only depends on the underlying copula C, since for every continuous $\varphi\colon [-1,1]\to\mathbb{R}$ using a change of coordinates, we have

$$\begin{aligned}
\Lambda_\varphi(Y|X) &= \alpha_\varphi^{-1}\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi\big(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y)\big)\,\mathrm{d}\mathbb{P}^X\otimes\mathbb{P}^X(x_1,x_2)\,\mathrm{d}\mathbb{P}^Y(y)\\
&= \alpha_\varphi^{-1}\int_{\mathbb{R}^2}\int_{\mathbb{R}} \varphi\big(K_C(F_X(x_1),[0,F_Y(y)]) - K_C(F_X(x_2),[0,F_Y(x)])\big)\,\mathrm{d}\mathbb{P}^X\otimes\mathbb{P}^X(x_1,x_2)\mathrm{d}\mathbb{P}^Y(y)\\
&= \alpha_\varphi^{-1}\int_0^1\int_0^1\int_0^1 \varphi\big(K_C(u_1,[0,v]) - K_C(u_2,[0,v]\big)\,\mathrm{d}\lambda(v)\,\mathrm{d}\lambda(u_1)\,\mathrm{d}\lambda(u_2). \qquad (21)
\end{aligned}$$

Following [25], we say that a sequence $(C_n)_{n\in\mathbb{N}}$ of bivariate copulas *converges weakly conditionally* to a bivariate copula C and write $C_n \xrightarrow{wcc} C$, if there exists some Borel set $\Gamma\subseteq[0,1]$ with $\lambda(\Gamma)=1$ such that the sequence $(K_{C_n}(u,\cdot))_{n\in\mathbb{N}}$ of probability measures converges weakly to the probability measure $K_C(u,\cdot)$ for every $u\in\Gamma$. The following result provides sufficient conditions for continuity of the functional Λ_φ. The proof is straightforward using eq. (21) and dominated convergence.

Proposition 3.9 (Continuity)**.** *Let $(X,Y),(X_1,Y_1),(X_2,Y_2),\dots$ be bivariate random vectors with continuous distribution functions and corresponding copulas $C, C_1, C_2, \dots$, and let $\varphi\colon [-1,1] \to \mathbb{R}$ be continuous. Then $C_n \xrightarrow{wcc} C$ implies convergence of $(\Lambda_\varphi(Y_n|X_n))_{n\in\mathbb{N}}$ to $\Lambda_\varphi(Y|X)$.*

Example 3. A particular family of dependence measures fulfilling all of the aforementioned properties is obtained when considering $\varphi(x) = |x|^p$ for $p \geq 1$, in which caseΛ_φ corresponds to the L^p-distance of conditional distribution functions. In fact, using eq. (21) we have

$$\Lambda_p(Y|X) := 3 \int_{[0,1]^3} |K_C(u_1,[0,z]) - K_C(u_2,[0,z])|^p \, \mathrm{d}\lambda^3(u_1,u_2,z) = \Lambda_p(C)\,, \tag{22}$$

where the normalizing constant $\alpha_\varphi = 1/3$ in (12) does not depend on p. It follows that the mapping $p \mapsto \Lambda_p(Y|X)$ is decreasing in $p \in [1,\infty)$.

For $p = 2$ the functional Λ_p compares two conditional distributions as in (1) and coincides with Chatterjee's rank correlation (see Example 1) which compares conditional and unconditional distributions as in (2). For $p = 1$, however, the following examples illustrates that the comparison of two conditional distributions as in (1) only coincides for some specific families with the L^1-analogue of Chatterjee's coefficient of correlation.

Example 4 (FGM and Fréchet copulas). For a random vector (X,Y) with continuous distribution function and underlying copula C, the functional $\zeta_1(Y|X)$ defined by

$$\zeta_1(Y|X) := 3 \int_{[0,1]^2} |K_C(u_1,[0,z]) - z| \, \mathrm{d}\lambda^2(u_1,z) \tag{23}$$

is a dependence measure introduced and studied in [38].

(a) For $\alpha \in [0,1]$ the copula $C_\alpha(u,v) := uv + \alpha\, u(1-u)v(1-v)$ is a member of the Farlie-Gumbel-Morgenstern (FGM) family. Let (X,Y) be a random vector with continuous distribution function and underlying copula C_α. Then straightforward calculations yield

$$\Lambda_1(Y|X) = \frac{\alpha}{3} \qquad \text{and} \qquad \zeta_1(Y|X) = \frac{\alpha}{4},$$

so $\Lambda_1(Y|X) = \frac{4}{3}\,\zeta_1(Y|X) \neq \zeta_1(Y|X)$.

(b) For $\alpha \in [0,1]$, the copula $C_\alpha(u,v) := \alpha\, M(u,v) + (1-\alpha)\,\Pi(u,v)$ is a member of the Fréchet copula family. In this case, for continuous (X,Y) with underlying copula C_α, we get

$$\Lambda_1(Y|X) = \alpha = \zeta_1(Y|X)\,.$$

4 Estimating the functionals

In this section we propose a copula-based so-called *checkerboard estimator* as plug-in estimator for the functional Λ_φ in (9) and start with recalling its construction and some relevant related facts.

For every bivariate copula A and $N \in \mathbb{N}$, the N-checkerboard approximation of A (see [27, 24]) will be denoted by $\mathfrak{Cb}_N(A)$ and N will be referred to as resolution. Loosely speaking, $\mathfrak{Cb}_N(A)$ is a uniform aggregation of A on N^2 squares of equal size. As a consequence of [27, Corollary 3.2], for every copula A the sequence $(\mathfrak{Cb}_N(A))_{N\in\mathbb{N}}$ converges weakly conditionally to A. The afore-mentioned checkerboard estimator is the checkerboard approximation of the empirical copula E_n (i.e., of the bilinear interpolation of the subcopula resulting from Sklar's theorem applied to the empirical distribution functions, see, e.g., [18, 20]). More precisely, suppose that $(X_1, Y_1), (X_2, Y_2), \ldots$ is a random sample from (X, Y) with copula A and let E_n denote the the empirical copula of the first n observations. Then the checkerboard estimator (i.e., the checkerboard approximation of the empirical copula E_n, again see [20]) fulfills

$$\mathfrak{Cb}_{N(n)}(E_n)(u,v) = \int_0^u \int_0^v N(n)^2 \sum_{i,j=1}^{N(n)} \frac{\beta_{ij}}{n} \mathbb{1}_{(\frac{i-1}{N(n)}, \frac{i}{N(n)}]\times(\frac{j-1}{N(n)}, \frac{j}{N(n)}]}(x,y)\, \mathrm{d}\lambda^2(x,y) \tag{24}$$

with $\beta_{ij} := \#\{k \in \{1,\ldots,n\} \mid (R_k/n, S_k/n) \in (\frac{i-1}{N(n)}, \frac{i}{N(n)}] \times (\frac{j-1}{N(n)}, \frac{j}{N(n)}]\}$, where R_k and S_k denote the ranks of X_k and Y_k, respectively, i.e., the number of ℓ such that $X_\ell \le X_k$ resp. $Y_\ell \le Y_k$.

For an illustration of the checkerboard estimator we refer to Figure 5 which depicts the density of $\mathfrak{Cb}_{N(n)}(E_n)$ (left panel) as well as the corresponding conditional distribution functions of the stripes (right panel) for the case of a real data example from medicine analyzed in Section 5.2. The following lemma shows strong consistency of the checkerboard estimator.

Lemma 4.1. *Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a random sample from (X, Y) and assume that (X, Y) has a continuous distribution function, underlying copula A, and that E_n denotes the empirical copula (of the first n observations). Then setting $N(n) := \lfloor n^s \rfloor$ for some $s \in (0, \frac{1}{2})$ the sequence $(\mathfrak{Cb}_{N(n)}(E_n))_{n\in\mathbb{N}}$ converges weakly conditionally to A with probability* 1.

Proof. Choose $\tilde{\Gamma} \in \mathcal{B}([0,1])$ with $\lambda(\tilde{\Gamma}) = 1$ in such a way that for every $x \in \tilde{\Gamma}$ we have that $(K_{\mathfrak{Cb}_N(A)}(x,\cdot))_{N\in\mathbb{N}}$ converges weakly to $K_A(x,\cdot)$. Setting $\Gamma := \tilde{\Gamma} \setminus \bigcup_{N=2}^{\infty}\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$ obviously yields $\lambda(\Gamma) = 1$. Fix $x \in \Gamma$ and consider some $y \in (0,1)$ with

$$\lim_{N\to\infty} K_{\mathfrak{Cb}_N(A)}(x,[0,y]) = K_A(x,[0,y]). \tag{25}$$

Plugging in the empirical copula, setting $I_n := |K_{\mathfrak{Cb}_{N(n)}(E_n)}(x,[0,y]) - K_A(x,[0,y])|$, and using the triangle inequality yields

$$I_n \;\le\; |K_{\mathfrak{Cb}_{N(n)}(E_n)}(x,[0,y]) - K_{\mathfrak{Cb}_{N(n)}(A)}(x,[0,y])| + |K_{\mathfrak{Cb}_{N(n)}(A)}(x,[0,y]) - K_A(x,[0,y])|.$$

The second summand converges to 0 for $n \to \infty$ due to (25). In order to show that the first summand converges to 0 almost surely, denote by $i_N(x)$, $N \ge 2$, the unique integer in $\{1,\ldots,N\}$ fulfilling $x \in (\frac{i_N(x)-1}{N}, \frac{i_N(x)}{N})$.

Since $\mathfrak{Cb}_{N(n)}(E_n)$ and $\mathfrak{Cb}_{N(n)}(A)$ are checkerboard copulas, we have

$$\begin{aligned} K_{\mathfrak{Cb}_{N(n)}(E_n)}(x,[0,y]) &= N(n)\left(E_n\left(\frac{i_{N(n)}(x)}{N(n)}, y\right) - E_n\left(\frac{i_{N(n)}(x)-1}{N(n)}, y\right)\right) \\ K_{\mathfrak{Cb}_{N(n)}(A)}(x,[0,y]) &= N(n)\left(A\left(\frac{i_{N(n)}(x)}{N(n)}, y\right) - A\left(\frac{i_{N(n)}(x)-1}{N(n)}, y\right)\right). \end{aligned}$$

This, however, implies

$$\limsup_{n\to\infty} |K_{\mathfrak{Cb}_{N(n)}(E_n)}(x,[0,y]) - K_{\mathfrak{Cb}_{N(n)}(A)}(x,[0,y])| \leq 2 \limsup_{n\to\infty} N(n)\, d_\infty(E_n, A) = 0\,,$$

$\mathbb{P}$-almost surely, whereby the last identity is a consequence of the definition of $N(n)$ and the fact that, according to [23, Lemma 1], the empirical copula E_n satisfies

$$d_\infty(E_n, A) = O\left(\sqrt{\frac{\log(\log(n))}{n}}\right) \tag{26}$$

with probability 1. Altogether it follows that $\lim_{n\to\infty} I_n = 0$ almost surely. Since all but at most countably many $y \in (0,1)$ fulfill (25) and weak convergence is equivalent to convergence on a dense set, we have shown weak convergence of $(K_{\mathfrak{Cb}_{N(n)}(E_n)}(x,\cdot))_{n\in\mathbb{N}}$ to $(K_A(x,\cdot))_{n\in\mathbb{N}}$ for every $x \in \Gamma$, which completes the proof. □

Having this, the following main result of this section is straightforward to prove.

Theorem 4.2 (Strongly consistent estimator for $\Lambda_\varphi(A)$)**.** *Let $(X_1,Y_1),(X_2,Y_2),\ldots$ be a random sample from (X,Y), assume that (X,Y) has a continuous distribution function, and let A denote the copula of (X,Y). Furthermore choose s and the resolution $N(n)$ as in Lemma 4.1. If φ is continuous, then $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ is a strongly consistent estimator for $\Lambda_\varphi(A)$, i.e., with probability 1, we have*

$$\lim_{n\to\infty} \Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n)) = \Lambda_\varphi(A).$$

Proof. According to (the proof of) Lemma 4.1 we have that $(K_{CB_{N(n)}}(u,\cdot))_{n\in\mathbb{N}}$ converges weakly to $K_A(u,\cdot)$ for every $u \in \Gamma$. As a direct consequence, for $u_1,u_2 \in \Gamma$ it follows that

$$\lim_{n\to\infty} |K_{CB_{N(n)(E_n)}}(u_1,[0,y])) - K_{CB_{N(n)(E_n)}}(u_2,[0,y]))| \quad = \quad |K_A(u_1,[0,y])) - K_A(u_2,[0,y]))|,$$

hence, considering $\lambda_2(\Gamma\times\Gamma) = 1$, using eq. (22), and applying dominated convergence we have

$$\begin{aligned}
&\lim_{n\to\infty} \int_{[0,1]^2}\int_{[0,1]} |K_{\mathfrak{Cb}_{N(n)(E_n)}}(u_1,[0,y])) - K_{\mathfrak{Cb}_{N(n)(E_n)}}(u_2,[0,y]))|^p d\lambda(y)\, d\lambda_2(u_1,u_2)\\
&= \int_{[0,1]^2}\int_{[0,1]} |K_A(u_1,[0,y])) - K_A(u_2,[0,y]))|^p d\lambda(y)\, d\lambda_2(u_1,u_2) = \Lambda_p(A).
\end{aligned}$$

□

Remark 4.3. Notice that, according to Theorem 4.2, strong consistency holds in full generality, i.e., without any smoothness restrictions for the copula A. Motivated by numerous simulations we conjecture that the estimator $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ is asymptotically normal; a proof, however, seems still out of reach, see, e.g., [36].

Remark 4.4. The checkerboard estimator can also be used to estimate $\Lambda_\Delta(Y|X)$ in the more general case where the distance function Δ is at least component-wise continuous with respect to weak convergence and invariant with respect to strictly increasing transformations of Y. Moreover, the checkerboard estimator may also be employed for estimating measures of explainability (see Theorem 2.2) in the case of knowing the distribution of Y.

5 Simulation study and real data example

In this section we first study the performance of the checkerboard estimator $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ from Theorem 4.2 and then briefly discuss a real data example. An R implementation allowing to reproduce all the results gathered in this section is available at https://github.com/plplus/depsenscond.

5.1 Simulation study

We consider the continuous setting as studied in the previous section, focusing on the following choices for the convex function φ – notice that φ is convex and strictly convex at $0 = \varphi(0)$, so Λ_φ is a dependence measure:

(i) $\varphi(x) = |x|^p$ for $p \in \{1, 2, 3\}$,

(ii) $\varphi(x) = e^{cx} - 1$ for $c \in \{1/5, 1, 5\}$,

(iii) $\varphi(x) = e^{|cx|} - 1$ for $c \in \{1/5, 1, 5\}$.

As dependence structures we choose copulas A from the Marshall-Olkin (MO) family, defined by

$$M_{\alpha,\beta}(u, v) := \begin{cases} u^{1-\alpha}v, & \text{if} \quad u^\alpha \geq v^\beta, \\ uv^{1-\beta}, & \text{if} \quad u^\alpha < v^\beta, \end{cases}$$

with $\alpha, \beta \in [0, 1]$, and consider $(\alpha, \beta) \in \{(1, 0), (1, 1), (0.2, 0.7), (0.3, 1)\}$. Figures 1 to 4 illustrate the behavior of the estimator $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ from Theorem 4.2 for each of the chosen copulas A and with sample sizes given by $n \in \{10; 50; 100; 500; 1\,000; 5\,000; 10\,000\}$, where each scenario is simulated 1 000 times. Estimation of the empirical checkerboard copula is performed using the command *ECBC* from the R package *qad* with the default of $s = 1/2$ (see [24, Remark 3.14] for a detailed discussion of this choice). Figures 1 to 4 can be interpreted as follows:

- The measures with convex function φ of type (ii) perform best in all cases in terms of speed of convergence. One possible explanation of this observation is asymmetry of $\varphi(x) = e^{cx} - 1$: The number of distinct values over which the integral runs is potentially twice as large and might therefore speed up convergence. As a drawback, however, for these functions, $\Lambda_\varphi(Y|X)$ often attains small values, see Figures 3 and 4, which, however should not be confused with independence as considered in Figure 1.
- $\Lambda_\varphi(Y|X)$ is decreasing in the parameters p and c (compare with Example 3). Consequently, in the case of complete dependence, the measures with parameter $p = 1$ and $c = 1/5$ perform best because deviations are less penalized by the shape of the function φ than by the other cases. In contrast to that, if the parameters p and c are large, then the measure Λ_φ is more sensitive for detecting independence.
- The L^2-version $\varphi(x) = x^2$ as well as the asymmetric version $\varphi(x) = e^x - 1$ perform best over all considered cases.

5.2 Real data example

As a real data example, we consider a dataset gathered by [9] and available online at PhysioNet.org, see [19]. The dataset contains, among other variables, 101 complete observations of pregnant women's age, BMI, blood pressure, blood glucose level and thickness of visceral adipose tissue, as well as the birth weight of the child. We choose the birth weight of the child as the endogenous variable Y and each of the other

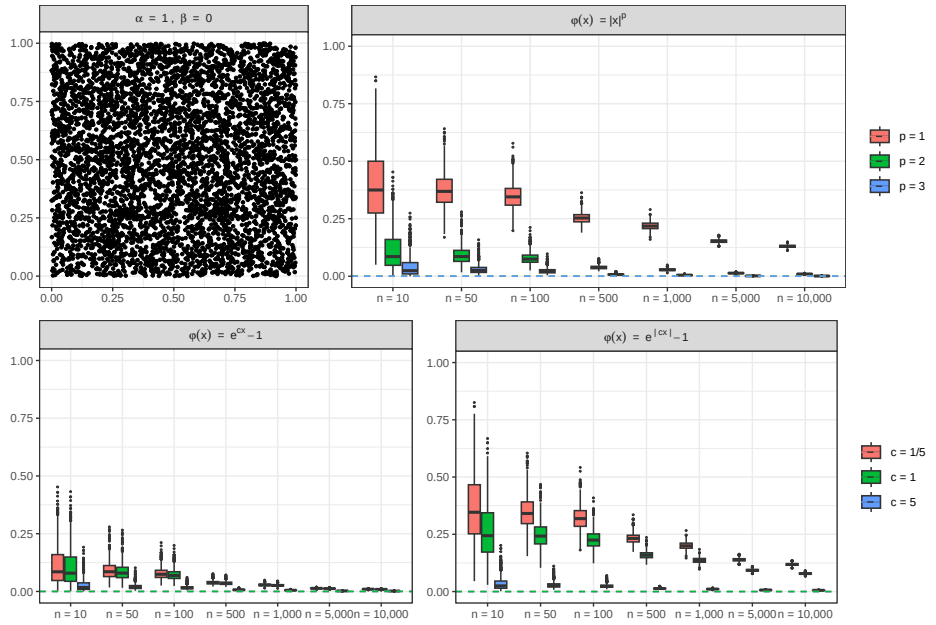

Figure 1 Simulation results for the MO copula with parameters $\alpha = 1$ and $\beta = 0$, i.e., for the independence copula. The top left panel depicts a random sample of size $n = 5.000$ from this distribution. The other panels contain boxplots of the estimations $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ according to Theorem 4.2 based on $1,000$ runs for each sample size n and for each convex function φ. The dashed lines represent the true values of $\Lambda_\varphi(A)$.

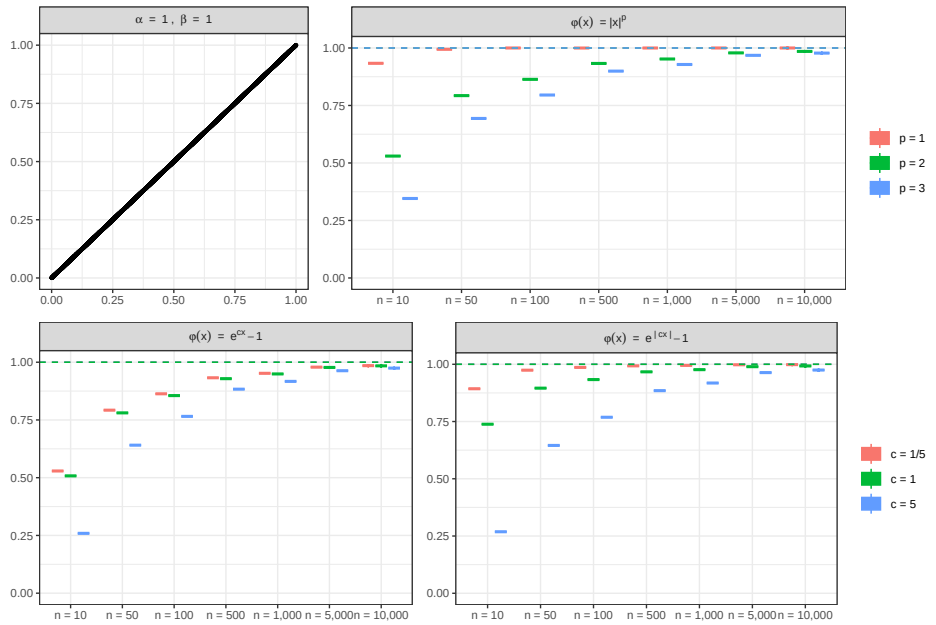

Figure 2 Simulation results for the MO Copula with parameters $\alpha = 1$ and $\beta = 1$, i.e., for the case of complete positive dependence.The top left panel depicts a random sample of size $n = 5.000$ from this distribution. The other panels contain boxplots of the estimations $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ according to Theorem 4.2 based on $1,000$ runs for each sample size n and for each convex function φ. The dashed lines represent the true values of $\Lambda_\varphi(A)$.

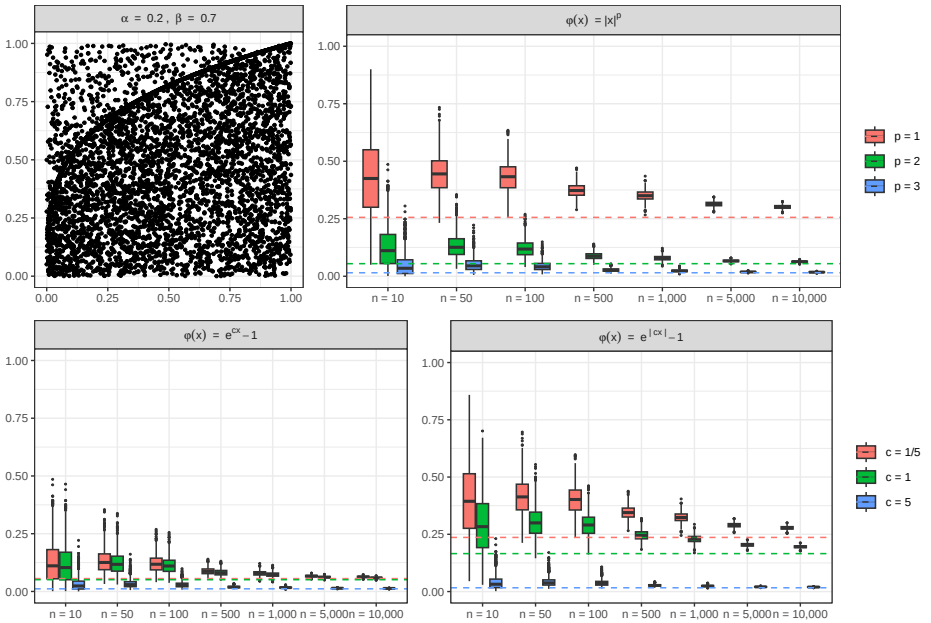

Figure 3 Simulation results for the MO Copula with parameters $\alpha = 0.2$ and $\beta = 0.7$. The top left panel depicts a random sample of size $n = 5.000$ from this distribution. The other panels contain boxplots of the estimations $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ according to Theorem 4.2 based on $1,000$ runs for each sample size n and for each convex function φ. The dashed lines represent the true values of $\Lambda_\varphi(A)$.

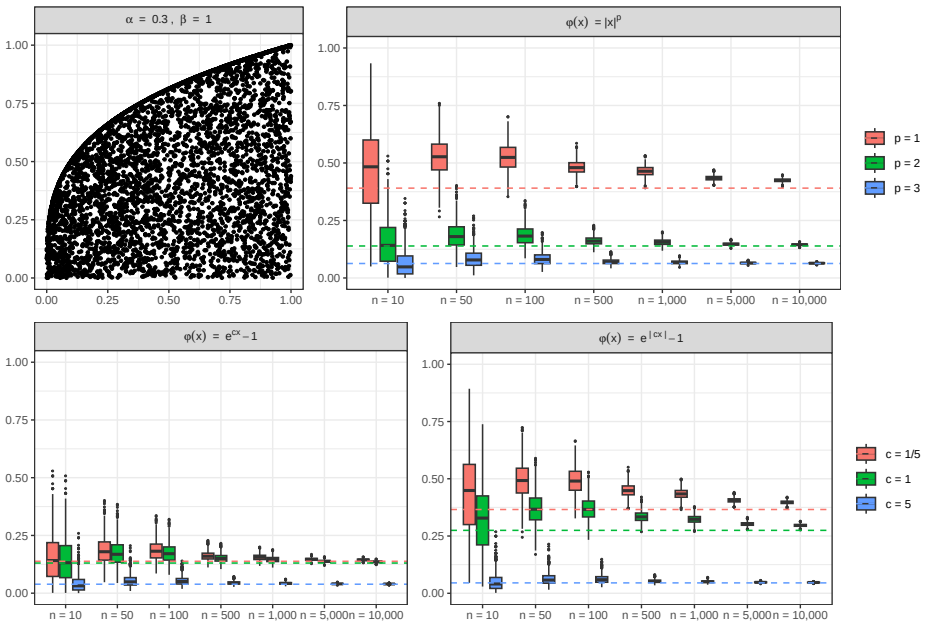

Figure 4 Simulation results for the MO Copula with $\alpha = 0.3$ and $\beta = 1$. The top left panel depicts a random sample of size $n = 5.000$ from this distribution. The other panels contain boxplots of the estimations $\Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n))$ according to Theorem 4.2 based on $1,000$ runs for each sample size n and for each convex function φ. The dashed lines represent the true values of $\Lambda_\varphi(A)$.

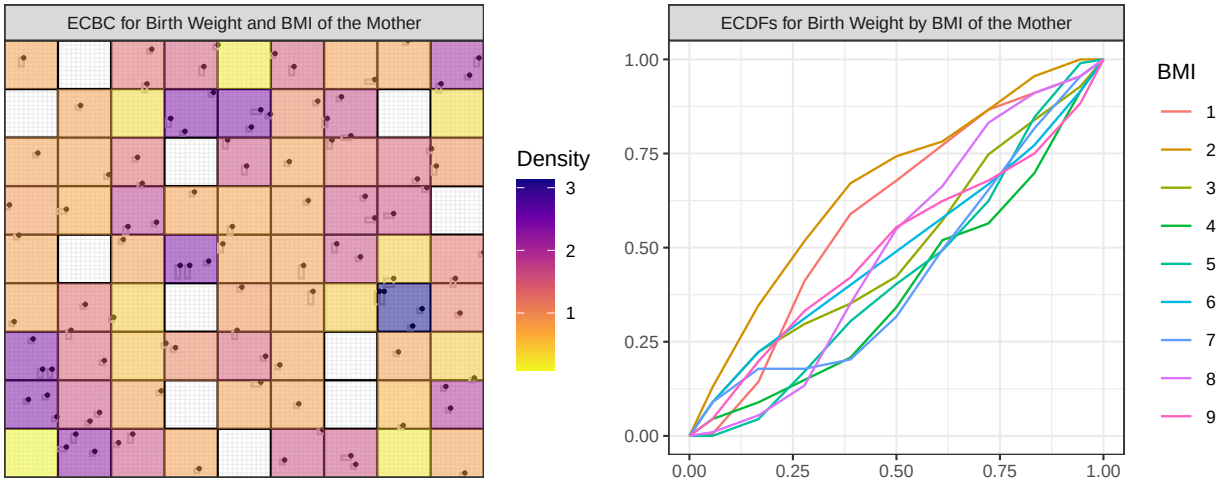

Figure 5 The left panel depicts the density of the empirical checkerboard copula $\mathfrak{Cb}_{N(n)}(E_n)$) with resolution $N = 9$, together with the normalized ranks (a.k.a. psuedo-observations) for Birth Weight (y-axis) and BMI of the mother (x-axis) as points. The small gray squares/rectangles to the bottom-left of each pseudo-observation are the regions where the empirical copula (bilinear interpolation of the subcopula resulting from Sklar's theorem) has positive density. The reason for not only having squares are the ties in the dataset; for details see the documentation of the *ECBC* function in the *qad* package). The right panel contains the resulting conditional distribution functions for the 9 different BMI regions corresponding to the 9 vertical stripes on the left plot. $\Lambda_\Delta(Y|X)$ is computed by taking the average pairwise distances $\Delta(F_i, F_j)$, $i, j \in \{1, \ldots, 9\}$ over these 9 ECDFs and subsequently dividing by α_Δ.

variables as the exogenous variable. Figure 5 illustrates the empirical checkerboard copula $\mathfrak{Cb}_{N(n)}(E_n)$ for BMI as exogenous variable. Table 1 summarizes the obtained values of $\Lambda_\varphi(Y|X)$ for all exogenous variables. Interpretation of the different values is not straightforward since they depend on the shape of φ (unless they are exactly 0 or 1). Due to the small sample size, the calculated estimates might still differ from the underlying true value. Remarkably, however, when sorting the variables by their estimated $\Lambda_\varphi(Y|X)$ value, the order is the same for all convex functions φ; that is, for all choices of φ the variable importance (w.r.t. Y) is the same.

Table 1 Values of $\Lambda_\varphi(Y|X)$ for various choices of the exogenous variable X and for different convex functions φ. In each of the cases the endogenous variable Y is the birth weight.

	$\varphi(x)$		
Exogenous Variable	$\lvert x\rvert$	$e^{\lvert x\rvert} - 1$	$e^x - 1$
Diastolic BP	0.354	0.231	0.072
BMI	0.326	0.210	0.060
Visceral Adipose Tissue	0.304	0.195	0.052
Systolic BP	0.291	0.186	0.049
Blood Glucose	0.214	0.134	0.027
Age	0.137	0.083	0.012

Appendix

Remark .1. Our proof of assertion (iii) in Theorem 2.2 requires the assumption of strict convexity of ψ in 0 and uses the fact that Jensen's inequality, applied to the difference of two non-degenerate i.i.d. random variables, is strict if φ is strictly convex at 0. Hence, equality can only hold if the underlying random variables are degenerate. Due to the integral with respect to the product measure $\mathbb{P}^X \otimes \mathbb{P}^X$ in (7), it has to be shown that Jensen's inequality holds for the integrand $\mathbb{P}^X \otimes \mathbb{P}^X$-almost surely.

Proof of Theorem 2.2. We first prove property (i) and eq. (8) and proceed as follows: Let Y_1 and Y_2 be independent copies of Y. Then, using convexity of ψ and the integrability assumption, the right-hand side of eq. (8) is finite since

$$\begin{aligned}\mathbb{E}\left|\psi(Y_1 - Y_2)\right| &= \mathbb{E}\left|\psi\left(\frac{2Y_1 + (-2Y_2)}{2}\right)\right| \\ &\leq \mathbb{E}\left|\psi(2Y_1) + \psi(-2Y_2)\right| \leq \mathbb{E}|\psi(2Y)| + \mathbb{E}|\psi(-2Y)| < \infty.\end{aligned}$$

If $Y = f(X)$ almost surely, then using change of coordinates, it yields

$$\begin{aligned}&\int_{\mathbb{R}^2} \psi\left(\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2)\right) \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \\ &= \int_{\mathbb{R}^2} \psi\left(f(x_1) - f(x_2)\right) \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \\ &= \int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right) \mathrm{d}(\mathbb{P}^{f(X)} \otimes \mathbb{P}^{f(X)})(y_1, y_2) \\ &= \int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right) \mathrm{d}(\mathbb{P}^{Y} \otimes \mathbb{P}^{Y})(y_1, y_2).\end{aligned} \tag{27}$$

Hence, using disintegration and Jensen's inequality, it follows for all $(X', Y') \in \mathcal{R}(Y)$ that

$$\begin{aligned}0 &= \psi(0) = \psi\left(\mathbb{E}(Y') - \mathbb{E}(Y')\right) \\ &= \psi\left(\int_{\mathbb{R}^2} \mathbb{E}[Y'|X' = x_1] - \mathbb{E}[Y'|X' = x_2] \,\mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2)\right) \\ &\leq \int_{\mathbb{R}^2} \psi\left(\mathbb{E}[Y'|X' = x_1] - \mathbb{E}[Y'|X' = x_2]\right) \mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) \qquad (28) \\ &= \int_{\mathbb{R}^2} \psi\left(\int_{\mathbb{R}^2} y_1 - y_2 \,\mathrm{d}(\mathbb{P}^{Y'|X'=x_1} \otimes \mathbb{P}^{Y'|X'=x_2})(y_1, y_2)\right) \mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) \\ &\leq \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right) \mathrm{d}(\mathbb{P}^{Y'|X'=x_1} \otimes \mathbb{P}^{Y'|X'=x_2})(y_1, y_2) \,\mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) \\ &= \int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right) \mathrm{d}(\mathbb{P}^{Y'} \otimes \mathbb{P}^{Y'})(y_1, y_2) \qquad (29) \\ &\leq \sup\left\{\int_{\mathbb{R}^2} \psi\left(\mathbb{E}(Y'|X' = x_1) - \mathbb{E}(Y'|X' = x_2)\right) \mathrm{d}(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) : (X', Y') \in \mathcal{R}(Y)\right\} \\ &= \alpha_\psi.\end{aligned}$$

Using eq. (27) and applying eq. (28) to eq. (29) on the elements over which the supremum is taken shows that the third inequality becomes an equality. This proves eq. (8) and implies $\Lambda_\psi(Y|X) \in [0, 1]$.

(ii): Obviously $\mathbb{V}(\mathbb{E}(Y|X)) = 0$ if and only if $\mathbb{E}(Y|X = x_1) = \mathbb{E}(Y) = \mathbb{E}(Y|X = x_2)$ for $\mathbb{P}^X \otimes \mathbb{P}^X$-almost all $x_1, x_2 \in \mathbb{R}$. The latter, however, is equivalent to the integrand in (7) being zero almost everywhere, since

ψ is strictly convex at 0 and $\psi(0) = 0$. (iii): If Y is completely dependent on X, then (27) and (8) imply

$$\begin{aligned}\alpha_\psi\, \Lambda_\psi(Y|X) &= \int_{\mathbb{R}^2} \psi\left(\mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_2)\right)\, \mathrm{d}(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2) \\ &= \int_{\mathbb{R}^2} \psi\left(y_1 - y_2\right)\, \mathrm{d}(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2) = \alpha_\psi.\end{aligned}$$

Since $\mathbb{P}^Y$ is non-degenerate, complete dependence of Y on X yields $\mathbb{E}(Y|X) = Y$ and we do not have $Y = \mathbb{E}(Y)$ almost surely. Hence, using (i) and (ii) the constant α_ψ is positive and $\Lambda_\psi(Y|X) = 1$ follows.

To show the reverse implication, first recall that by Jensen's inequality and eq. (8), we have

$$\begin{aligned}\alpha_\psi \Lambda_\psi(Y|X) &= \int_{\mathbb{R}^2} \psi\Big(\int_{\mathbb{R}^2} (y_1 - y_2)\, \mathrm{d}\mathbb{P}^{Y|X=x_1} \otimes \mathbb{P}^{Y|X=x_2}(y_1, y_2)\Big)\, \mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1, x_2) \\ &\leq \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \psi(y_1 - y_2)\, \mathrm{d}\mathbb{P}^{Y|X=x_1} \otimes \mathbb{P}^{Y|X=x_2}\, \mathrm{d}\mathbb{P}^X \otimes \mathbb{P}^X(x_1, x_2) = \alpha_\psi.\end{aligned} \tag{30}$$

For $\Lambda_\psi(Y|X) = 1$, the above inequality becomes an equality if there exists some Borel set $G \subseteq \mathbb{R}^2$ with $\mathbb{P}^X \otimes \mathbb{P}^X(G) = 1$ such that for $(x_1, x_2) \in G$ and $Y_i \sim \mathbb{P}^{Y|X=x_i}$, $i \in \{1, 2\}$, Y_1 and Y_2 independent,

$$\text{either} \quad Y_1 \leq Y_2 \ \text{ almost surely} \quad \text{or} \quad Y_1 \geq Y_2 \ \text{ almost surely.} \tag{31}$$

In fact, otherwise Jensen's inequality would be strict due to strict convexity of ψ at $0 = \psi(0)$, which contradicts $\Lambda_\psi(Y|X) = 1$. Letting $G_x := \{y \in \mathbb{R} \colon (x, y) \in G\}$ denote the x-cut of G, defining $W := \{x \in \mathbb{R} \colon \mathbb{P}^X(G_x) = 1\}$ and using disintegration, obviously yields $\mathbb{P}^X(W) = 1$. To simplify notation, let $l(x)$ and $u(x)$ denote the infimum and supremum of the support of $\mathbb{P}^{Y|X=x}$, i.e.,

$$l(x) := \inf(\mathrm{supp}(\mathbb{P}^{Y|X=x})) \leq \sup(\mathrm{supp}(\mathbb{P}^{Y|X=x})) =: u(x). \tag{32}$$

Then, according to (31), for each $(x_1, x_2) \in G$ the open intervals fulfil $(l(x_1), u(x_1)) \cap (l(x_2), u(x_2)) = \emptyset$. Finally, define

$$V := \{x \in \mathbb{R} \colon \mathbb{P}^{Y|X=x} \text{ is non-degenerate}\} = \{x \in \mathbb{R} \colon \mathbb{V}(Y|X = x) > 0\}. \tag{33}$$

For completing the proof, it suffices to show

$$\mathbb{P}^X(V) = 0 \tag{34}$$

since it then follows that $\mathbb{P}^{Y|X=x}$ is degenerate for $\mathbb{P}^X$-almost all $x \in \mathbb{R}$, implying that Y is completely dependent on X.

Assume, on the contrary, that $\mathbb{P}^X(V) > 0$ holds. Then considering $\mathbb{P}^X(W) = 1$, it follows that

$$\mathbb{P}^X(V \cap W) = \mathbb{P}^X(V) > 0\,. \tag{35}$$

Let $x \in V \cap W$ be arbitrary but fixed. Then, by construction, we have

$$(l(x), u(x)) \cap (l(z), u(z)) = \emptyset \tag{36}$$

for all $z \in \mathbb{R}$ with $(x, z) \in G$, and, in particular, for all $z \in V \cap W \cap G_x$. Considering $\mathbb{P}^X(G_x) = 1 = \mathbb{P}^X(W)$, it follows that $0 < \mathbb{P}^X(V) = \mathbb{P}^X(V \cap W \cap G_x)$. Hence, (36) holds, in particular, for $\mathbb{P}^X$-almost every $z \in V \cap W$.

Finally, let $q_1, q_2, q_3, \ldots$ be an enumeration of the rational numbers in $\mathbb{R}$. Then, defining M_i by $M_i := \{x \in V \cap W \colon q_i \in (l(x), u(x))\}$ yields $\bigcup_{i=1}^\infty M_i = V \cap W$, since the rationals are dense in $\mathbb{R}$. Using (35) as well as sub-σ-additivity, yields

$$0 < \mathbb{P}^X(V \cap W) = \mathbb{P}^X\Big(\bigcup_{i=1}^\infty M_i\Big) \leq \sum_{i=1}^\infty \mathbb{P}^X(M_i), \tag{37}$$

so there exists some $i_0 \in \mathbb{N}$ with $\mathbb{P}^X(M_{i_0}) > 0$. The latter, however, contradicts (36), so $\mathbb{P}^X(V) > 0$ cannot hold, and the proof is complete. □

Funding
The authors gratefully acknowledge the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP). The first and the third author further acknowledge the support of the Austrian Science Fund (FWF) project P 36155-N *ReDim: Quantifying Dependence via Dimension Reduction.*

References

[1] J. Ansari. *Ordering risk bounds in partially specified factor models.* Freiburg im Breisgau: Univ. Freiburg, Fakultät für Mathematik und Physik (Diss.), 2019.

[2] J. Ansari, S. Fuchs. A simple extension of azadkia & chatterjee's rank correlation to a vector of endogenous variables. arXiv:2212.01621, 2023.

[3] J. Ansari, L. Rüschendorf. Sklar's theorem, copula products, and ordering results in factor models. *Depend. Model.*, 9:267–306, 2021.

[4] J. Ansari, M. Rockel. Dependence properties of bivariate copula families. *Depend. Model.* 12:1-36, 2024.

[5] M. Azadkia and S. Chatterjee. A simple measure of conditional dependence. *Ann. Stat.*, 49(6):3070–3102, 2021.

[6] P.J. Bickel. Measures of independence and functional dependence. arxiv:2206.13663v1, 2022.

[7] S. Chatterjee. A new coefficient of correlation. *J. Amer. Statist. Ass.*, 116(536):2009–2022, 2020.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Hoboken, 2006.

[9] A. d. S Rocha, L. von Diemen, D. Kretzer, S. Matos, J. A. Rombaldi Bernardi, J. & Magalhaes. Visceral adipose tissue measurements during pregnancy (version 1.0.0), 2020. https://doi.org/10.13026/p729-7p53.

[10] N. Deb, P. Ghosal, and B. Sen. Measuring association on topological spaces using kernels and geometric graphs. arXiv:2010.01768, 2020.

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

[12] A.A. Ding, J.G. Dy, Y Li, and Y. Chang. A robust-equitable measure for feature ranking and selection. *J. Mach. Learn. Res.*, 18:1–46, 2017.

[13] C. Dwork, F. McSherry, K. Nissim, A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography. Third theory of cryptography conference, TCC 2006, New York, NY, USA, March 4–7, 2006. Proceedings.*, pages 265–284. Berlin: Springer, 2006.

[14] A. Evfimievski, J. Gehrke, R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 211–222, 2003.

[15] S. Fuchs. Quantifying directed dependence via dimension reduction. *J. Multivariate Anal.,*, 201:105266, 2024.

[16] E. Furman, R. Wang, R. Zitikis. Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, 83:70–84, 2017.

[17] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[18] C. Genest, J.G. Neslehová, B. Rémillard. Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivariate Anal.*, 159:82–110, 2017.

[19] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, B. Moody, G., and K. & Stanley H. E. Peng, C. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, 101(23):e215–e220, 2000. https://doi.org/10.13026/p729-7p53.

[20] F. Griessenberger, R.R. Junker, and W. Trutschnig. On a multivariate copula-based dependence measure and its estimation. *Electron. J. Statist.*, 16:2206–2251, 2022.

[21] F. Han and Z. Huang. Azadkia-Chatterjee's correlation coefficient adapts to manifold data. *Ann. Appl. Probab.*, 34:5172–5210, 2022.

[22] Z. Huang, N. Deb, B. Sen. Kernel partial correlation coefficient — a measure of conditional dependence. *J. Mach. Learn. Res.*, 23(216):1–58, 2022.

[23] P. Janssen, J. Swanepoel, N. Veraverbeke. Large sample behavior of the Bernstein copula estimator. *J. Stat. Plann. Inference*, 142(5):1189–1197, 2012.

[24] R.R. Junker, F. Griessenberger, W. Trutschnig. Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.*, 153:107058, 22 pages, 2020.

[25] T. Kasper, S. Fuchs, and W. Trutschnig. On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation. *Bernoulli*, 27:2217–2240, 2021.

[26] J.B. Kinney and G.S Atwal. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA*, 111:3354–3359, 2014.

[27] X. Li, P. Mikusiński, and M. D. Taylor. Strong approximation of copulas. *J. Math. Anal. Appl.*, 225(2):608–623, 1998.

[28] T. Mroz, J. Fernández-Sánchez, S. Fuchs, W. Trutschnig. On distributions with fixed marginals maximizing the joint or the prior default probability, estimation, and related results. *J. Stat. Plann. Inference*, 223:33–52, 2023.

[29] A. Müller, M. Scarsini. Archimedean copulae and positive dependence. *J. Multivariate Anal.*, 93(2):434–445, 2005.

[30] R. B. Nelsen. *An Introduction to Copulas.* Springer, New York, 2nd edition, 2006.

[31] T.G. Nies, T. Staudt, and A. Munk. Transport dependency: Optimal transport based dependency measures. arxiv:2105.02073, 2023.

[32] A. Rohde, L. Steinberger. Geometrizing rates of convergence under local differential privacy constraints. *Ann. Stat.*, 48(5):2646–2670, 2020.

[33] L. Rüschendorf. *Mathematical risk analysis. Dependence, risk bounds, optimal allocations and portfolios.* Springer Ser. Oper. Res. Financ. Eng. Berlin: Springer, 2013.

[34] M. Shaked, M. A. Sordo, A. Suárez-Llorens. A global dependence stochastic order based on the presence of noise. In *Stochastic Orders in Reliability and Risk. In Honor of Professor Moshe Shaked.*, pages 3–39. New York, NY: Springer, 2013.

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

[36] C. Strothmann, H. Dette, and K.F. Siburg. Rearranged dependence measures. *Bernoulli*, 30:1055-1078, 2024.

[37] E. A. Sungur. A note on directional dependence in regression setting. *Comm. Statist. Theory Methods*, 34:1957–1965, 2005.

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

[39] J.C.W. Wiesel. Measuring association with Wasserstein distances. *Bernoulli*, 28:2816–2832, 2022.