

Mean and quantile regression in the copula setting: properties, sharp bounds and a note on estimation

Henrik Kaiser^a, Wolfgang Trutschnig^{a,*}

^a*Department for Artificial Intelligence and Human Interfaces, University Salzburg,
Hellbrunnerstrasse 34, A-5020 Salzburg, Austria
Tel: +43 662 8044 5326 Fax: +43 662 8044 137*

Abstract

Driven by the interest on how uniformity of marginal distributions propagates to properties of regression functions, in this contribution we tackle the following questions: Given a $(d-1)$ -dimensional random vector $\mathbf{X}$ and a random variable Y such that all univariate marginals of $(\mathbf{X}, Y)$ are uniformly distributed on $[0, 1]$, how large can the average absolute deviation of the mean and the quantile regression function of Y given $\mathbf{X}$ from the value $\frac{1}{2}$ be, and how much mass may sets with large deviation have? We answer these questions by deriving sharp inequalities, both in the mean as well as in the quantile setting, and sketch some cautionary consequences to nowadays quite popular pair copula constructions involving the so-called simplifying assumption. Rounding off our results, working with the so-called empirical checkerboard estimator in the bivariate setting, we show strong consistency for both regression types and illustrate the speed of convergence in terms of a simulation study.

Keywords: copulas, regression, quantiles, estimation

2020 MSC: 60E05, 62G08, 62H20, 60E15

1. Introduction

Regression methods, particularly mean and quantile regression, play a fundamental role throughout all quantitative fields. Traditionally, the focus

*Corresponding author

Email addresses: henrik.kaiser@plus.ac.at (Henrik Kaiser),
wolfgang@trutschnig.net (Wolfgang Trutschnig)

lies on estimating the conditional mean of a response variable Y , given the values of an ensemble of explanatory variables $\mathbf{X}$, with the aim to summarize the relationship between covariates and outcome or to predict the value of Y , based on new observations of $\mathbf{X}$. Trying to capture more information on the distribution of the outcome Y given an observation of $\mathbf{X}$, quantile regression (see, e.g., [12, 13]) provides a more comprehensive understanding of the response distribution across its entire range.

According to Sklar's famous theorem (see, [4, 20]), copulas constitute the link between (continuous) multivariate distributions functions and their univariate marginals - as such, they capture all scale-invariant dependence between random variables. Motivated by this fact, we here study mean and quantile regression in the context of d -dimensional copulas, interpreting the first $d - 1$ coordinates as covariates and the d -th coordinate as outcome. Our focus is not on separate estimation of the marginals and the underlying copulas as done in [3] - inspired by curiosity about how uniformity of the marginals translates/propagates to characteristics of regression functions, the goal of this contribution is twofold: firstly, to derive best-possible upper bounds for the maximal L^p -deviation of the regression function from the value $\frac{1}{2}$ corresponding to the regression function describing independence. And, secondly, to determine best-possible bounds for the mass of sets with large deviation. In other words, assuming that $\mathbf{X}$ is $(d - 1)$ -dimensional random vector and Y is a random variable on a joint probability space $(\Omega, \mathcal{A}, \mathbb{P})$, such that $(\mathbf{X}, Y)$ has copula C as distribution function (restricted to $[0, 1]^d$), we want to know, how large

$$\int_{\mathbb{I}^{d-1}} |\mathbb{E}(Y|\mathbf{X} = \mathbf{x}) - \tfrac{1}{2}|^p d\mathbb{P}^{\mathbf{X}}(\mathbf{x}), \quad p \in [1, \infty),$$

as well as

$$\mathbb{P}^{\mathbf{X}} \left(\left\{ \mathbf{x} \in \mathbb{I}^{d-1} : |\mathbb{E}(Y|\mathbf{X} = \mathbf{x}) - \tfrac{1}{2}| \geq a \right\} \right)$$

can possibly be. After providing sharp inequalities for mean regression we ask (and answer) the analogous questions for quantile regression. To the best of our knowledge, these natural properties have not been investigated in the literature yet. While our original intention was to provide answers to the afore-mentioned question primarily for the family of so-called linkages, modeling the situation in which the covariates are independent (which is particularly useful for constructing multivariate dependence measures (see [6]), we decided to state and prove our results for the general setting, since very little additional technical effort is required.

Throughout the past decade pair copula constructions in combination with the so-called simplifying assumption (praised for their flexibility also in

the high-dimensional setting, see [1, 8, 19]) have become more and more popular. As shown in [17] (also see the very recent survey [18]), these constructions may suffer from very poor approximation quality and should therefore be handled with care. Our results on regression underline the fact that the afore-mentioned warning extends to the regression setting.

The remainder of this paper is organized as follows: After gathering notation and preliminaries in Section 2, we study the absolute deviation of the mean regression function from its mean $\frac{1}{2}$ and answer the afore-mentioned two questions by providing best-possible bounds. Section 4 focuses on quantile regression and again establishes sharp bounds. Finally, in Section 5, we study the bivariate setting and prove strong consistency of the empirical checkerboard estimator for the mean and the quantile regression function, without any regularity conditions for the copula C . A small simulation study, illustrating the obtained convergence results, rounds off the paper.

2. Notation and Preliminaries

For every metric space (S, d) we will let $\mathcal{B}(S)$ denote the Borel σ -field on S . Throughout this article, $\mathbb{I} := [0, 1]$ denotes the closed unit interval, the dimension is denoted by $d \in \mathbb{N} \setminus \{1\}$, and bold symbols refer to vectors, e.g., $\mathbf{x} := (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$. For $\mathbf{x} := (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ and $l \in \{1, \dots, d\}$, we will let $\mathbf{x}_{1:l}$ denote the vector $(x_1, \dots, x_l) \in \mathbb{R}^l$. Since the main focus of this contribution is regression, we will in particular write $(\mathbf{x}, y) = (x_1, \dots, x_{d-1}, y) \in \mathbb{R}^d$ for all $\mathbf{x} \in \mathbb{R}^{d-1}$ and $y \in \mathbb{R}$. Furthermore, λ_d stands for the d -dimensional Lebesgue-measure on $\mathcal{B}(\mathbb{R}^d)$ or $\mathcal{B}(\mathbb{I}^d)$, with $\lambda := \lambda_1$, for brevity. The family of d -dimensional copulas is indicated by $\mathcal{C}^d$, the uniform metric on $\mathcal{C}^d$ is defined by

$$d_\infty(A, B) := \max_{(\mathbf{x}, y) \in \mathbb{I}^d} |A(\mathbf{x}, y) - B(\mathbf{x}, y)| \quad (A, B \in \mathcal{C}^d).$$

It is well known that $(\mathcal{C}^d, d_\infty)$ is a compact metric space (see [4, 20]). Specifically Π_d and M_d denote the d -dimensional product (or independence) and minimum copula, respectively, i.e., $\Pi(\mathbf{x}, y) = y \prod_{j=1}^{d-1} x_j$ and $M_d(\mathbf{x}, y) = \min\{x_1, \dots, x_{d-1}, y\}$. In the bivariate case, we simply write $\Pi := \Pi_2$ and $M := M_2$.

Given $C \in \mathcal{C}^d$, the corresponding d -stochastic measure is denoted by μ_C , i.e., $\mu_C([0, \mathbf{x}] \times [0, y]) := C(\mathbf{x}, y)$ for all $(\mathbf{x}, y) \in \mathbb{I}^d$, where $[0, \mathbf{x}] := \times_{i=1}^{d-1} [0, x_i]$. For every vector $\mathbf{j} = (j_1, \dots, j_l) \in \{1, \dots, d\}^l$, with $j_1 < j_2 < \dots < j_l$, we will let $C_{\mathbf{j}}$ denote the marginal copula of C with respect to the coordinates

$\mathbf{j}$. In other words, defining the projection $\pi_{\mathbf{j}} : \mathbb{I}^d \rightarrow \mathbb{I}^l$ by

$$\pi_{\mathbf{j}}(x_1, \dots, x_d) = (x_{j_1}, x_{j_2}, \dots, x_{j_l}),$$

we have $C_{\mathbf{j}}$ corresponds to the push-forward $\mu_C^{\pi_{\mathbf{j}}}$ of μ_C via $\pi_{\mathbf{j}}$. To keep notation simple (and in accordance with $x_{1:l}$ from above), in the case of $\mathbf{j} = (1, 2, \dots, d-1)$, we will simply write $C_{\mathbf{j}} = C_{1:(d-1)} \in \mathcal{C}^{d-1}$ and refer to $C_{1:(d-1)}$ as marginal copula of C with respect to the first $d-1$ coordinates; analogously, in the case of $\mathbf{j} = (1, 2, \dots, d-2, d)$, we will write $C_{\mathbf{j}} = C_{1:(d-2),d} \in \mathcal{C}^{d-1}$. Finally, the d -flipped (flipped with respect to the last coordinate) of $C \in \mathcal{C}^d$ is defined by $\bar{C}(\mathbf{x}, y) := C_{1:(d-1)}(\mathbf{x}) - C(\mathbf{x}, 1-y)$, obviously we have $\bar{C} \in \mathcal{C}^d$.

Besides d -stochastic measures, in what follows (regular) conditional distributions of a copula will be of special importance. Suppose that $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces. Then, a map $K : \Omega_1 \times \mathcal{A}_2 \rightarrow \mathbb{I}$ is called Markov kernel (a.k.a. transition probability) from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$, if the map $\omega_1 \mapsto K(\omega_1, A_2)$ is $\mathcal{A}_1$ - $\mathcal{B}(\mathbb{R})$ -measurable, for every fixed $A_2 \in \mathcal{A}_2$, and the map $A_2 \mapsto K(\omega_1, A_2)$ is a probability measure on $\mathcal{A}_2$, for every fixed $\omega_1 \in \Omega_1$. Considering a random variable Y and a $(d-1)$ -dimensional random vector $\mathbf{X}$ on a joint probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a Markov kernel $K : \mathbb{R}^{d-1} \times \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{I}$ is said to be a regular conditional distribution of Y given $\mathbf{X}$, if for every $F \in \mathcal{B}(\mathbb{R})$ the following assertion holds: for $\mathbb{P}$ -almost every $\omega \in \Omega$ we have

$$K(\mathbf{X}(\omega), F) = \mathbb{E}(\mathbf{1}_F \circ Y | \mathbf{X})(\omega).$$

For each random vector $(\mathbf{X}, Y)$, it is well-known that a regular conditional distribution K of Y given $\mathbf{X}$ exists and is unique for $\mathbb{P}^{\mathbf{X}}$ -a.e. $\mathbf{x} \in \mathbb{R}^{d-1}$, with $\mathbb{P}^{\mathbf{X}}$ denoting the push-forward of $\mathbb{P}$ under $\mathbf{X}$. It is also well known that K only depends on $\mathbb{P}^{(\mathbf{X}, Y)}$. We will write $(\mathbf{X}, Y) \sim C$ if the copula C is the distribution function of $(\mathbf{X}, Y)$ restricted to $\mathbb{I}^d$. Finally, without loss of generality, we interpret the Markov kernel K_C of $C \in \mathcal{C}^d$ (with respect to the first $(d-1)$ -coordinates) as a mapping $K_C : \mathbb{I}^{d-1} \times \mathcal{B}(\mathbb{I}) \rightarrow \mathbb{I}$. For further background on copulas, d -stochastic measures, conditional expectation and Markov kernels we refer to [4, 11, 20, 10].

Generally speaking, disintegration theorems refer to integral representations of multivariate measures in terms of marginals and conditional distributions. In the case of $d \geq 3$ many such representations exist (see [10]). In its simplest form, in the copula setting we have

$$\mu_C(\mathbf{B} \times F) = \int_{\mathbf{B}} K_C(\mathbf{x}, F) \, d\mu_{C_{1:(d-1)}}(\mathbf{x}), \quad (1)$$

for every $\mathbf{B} \in \mathcal{B}(\mathbb{I}^{d-1})$ and every $F \in \mathcal{B}(\mathbb{I})$, implying in particular that the d -th univariate marginal is uniform on $\mathbb{I}$. In the sequel, we will especially use the following property on the relationship between projections and Markov kernels, the proof of which is provided in Appendix A.

Lemma 1. *Suppose that $d \geq 3$, that $C \in \mathcal{C}^d$ and let $F \in \mathcal{B}(\mathbb{I})$ be arbitrary but fixed. Then, for λ_{d-2} -a.e. $\mathbf{x}_{1:d-2} \in \mathbb{I}^{d-2}$, we have*

$$\int_{\mathbb{I}} K_C(\mathbf{x}_{1:d-1}, F) K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}) = K_{C_{1:(d-2),d}}(\mathbf{x}_{1:d-2}, F). \quad (2)$$

The (mean) regression function r_C of a copula $C \in \mathcal{C}^d$ (with respect to the first $d-1$ coordinates), i.e., the function $\mathbf{x} \mapsto \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ for $(\mathbf{X}, Y) \sim C$, can be expressed in terms of the Markov kernel K_C as

$$r_C(\mathbf{x}) := \int_{\mathbb{I}} y K_C(\mathbf{x}, dy) \quad (\mathbf{x} \in \mathbb{I}^{d-1}), \quad (3)$$

or, equivalently as

$$r_C(\mathbf{x}) = \int_{\mathbb{I}} K_C(\mathbf{x}, (y, 1]) dy \quad (\mathbf{x} \in \mathbb{I}^{d-1}). \quad (4)$$

Obviously, for the flipped copula $\overline{C}$, we have $r_{\overline{C}} = 1 - r_C$.

For fixed $C \in \mathcal{C}^d$, $\mathbf{x} \in \mathbb{I}^{d-1}$ and $\tau \in [0, 1]$, the τ -quantile function Q_C^τ is defined by

$$Q_C^\tau(\mathbf{x}) := \inf\{y \in \mathbb{I} : K_C(\mathbf{x}, [0, y]) \geq \tau\} = (F_C^\mathbf{x})^-(\tau), \quad (5)$$

where $(F_C^\mathbf{x})^-$ denotes the quasi-inverse of the conditional distribution function $y \mapsto F_C^\mathbf{x}(y) = K_C(\mathbf{x}, [0, y])$. In the sequel, we will only consider $\tau \in (0, 1]$ since, by definition, for $\tau = 0$ we have $Q_C^\tau(\mathbf{x}) = 0$ for every $C \in \mathcal{C}^d$ and every $\mathbf{x} \in \mathbb{I}^{d-1}$. It is well known (and straightforward to verify) that $y_0 < Q_C^\tau(\mathbf{x})$ if, and only if $K_C(\mathbf{x}, [0, y_0]) < \tau$. Moreover, $Q_C^\tau(\mathbf{x}) < y_0$ implies $K_C(\mathbf{x}, [0, y_0]) \geq \tau$. As a direct consequence, for all $(y_0, \tau) \in \mathbb{I}^2$, we have

$$\{\mathbf{x} \in \mathbb{I}^{d-1} : Q_C^\tau(\mathbf{x}) \leq y_0\} = \{\mathbf{x} \in \mathbb{I}^{d-1} : K_C(\mathbf{x}, [0, y_0]) \geq \tau\},$$

so that measurability of $\mathbf{x} \mapsto Q_C^\tau(\mathbf{x})$ directly follows from measurability of $\mathbf{x} \mapsto K_C(\mathbf{x}, F)$ for every $F \in \mathcal{B}(\mathbb{I})$.

A handy and frequently used class of copulas are so-called checkerboard copulas (see [16, 6, 21] and the references therein). Roughly speaking, these

copulas are characterized by the property that they locally, on hypercubes of equal volume, resemble shrunk copies of d -dimensional copulas. For $N \in \mathbb{N}$ and $i \in \{1, \dots, n\}$ write $I_{N,i} := (\frac{i-1}{N}, \frac{i}{N})$. Then, the checkerboard Π -approximation of $C \in \mathcal{C}^d$ with resolution $N \in \mathbb{N}$, denoted by $\mathfrak{C}\mathfrak{b}_N(C)$, is the copula with density (see [6] and the references therein)

$$\mathfrak{c}\mathfrak{b}_N(C)(\mathbf{x}, y) := N^d \sum_{i_1, \dots, i_d=1}^N \mu_C(\times_{k=1}^d I_{N,i_k}) \mathbb{1}_{\times_{k=1}^d I_{N,i_k}}(\mathbf{x}, y) \quad ((\mathbf{x}, y) \in \mathbb{I}^d).$$

By construction, the conditional distribution functions $y \mapsto K_{\mathfrak{C}\mathfrak{b}_N(C)}(\mathbf{x}, [0, y])$ are piecewise linear and constant (in $\mathbf{x}$) on every hypercube $I_{N,i_1} \times \dots \times I_{N,i_{d-1}}$. As a direct consequence, the regression function $r_{\mathfrak{C}\mathfrak{b}_N(C)}$ as well as each quantile regression function $Q_{\mathfrak{C}\mathfrak{b}_N(C)}^T$ is piecewise constant. Specifically, if E_n denotes the empirical copula (i.e., the multilinear interpolation of the subcopula induced by the pseudo-ranks) of a sample $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ from $(\mathbf{X}, Y) \sim C$, we refer to $\mathfrak{C}\mathfrak{b}_N(E_n)$ as the empirical checkerboard approximation with resolution $N = N(n)$ of E_n , or, shortly as empirical N -checkerboard.

Our subsequent discussion is mostly focused on copula families with fixed $1 : (d-1)$ -marginal $A \in \mathcal{C}^{d-1}$. Formally, for $d \geq 3$ and a fixed $A \in \mathcal{C}^{d-1}$, we will consider families of the form

$$\mathcal{C}_A^d := \left\{ C \in \mathcal{C}^d : C_{1:(d-1)}(\mathbf{x}) = A(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{I}^{d-1} \right\},$$

and refer to A as the copula of the covariates. Considering that univariate marginals of copulas coincide with λ on $\mathbb{I}$, we will formulate all results concerning $\mathcal{C}_A^d$ for arbitrary $d \geq 2$ and interpret the case $d = 2$ accordingly. For fixed $A \in \mathcal{C}^{d-1}$ define the L^p -norm of a measurable function $f : \mathbb{I}^{d-1} \rightarrow \mathbb{R}$ as (notice the dependence on A through the integrating measure)

$$\|f\|_{A,p} := \left\{ \int_{\mathbb{I}^{d-1}} |f(\mathbf{x})|^p \, d\mu_A(\mathbf{x}) \right\}^{\frac{1}{p}} \quad (p \in [1, \infty)).$$

The family

$$\mathcal{C}_{\Pi}^d = \left\{ C \in \mathcal{C}^d : C_{1:(d-1)}(\mathbf{x}) = \prod_{i=1}^{d-1} x_i \text{ for all } \mathbf{x} \in \mathbb{I}^{d-1} \right\}$$

is known as the linkage class and has been used in [6] to construct a measure quantifying the extent of dependence of a random variable Y on a random

vector $\mathbf{X}$ (which was a direct extension of the original bivariate approach in [9]). For $(\mathbf{X}, Y) \sim C \in \mathcal{C}_{\mathbb{I}}^d$ obviously all covariates are independent, i.e., $\mu_{C_{1:(d-1)}} = \lambda_{d-1}$ holds, which is why we pay special attention to $\mathcal{C}_{\mathbb{I}}^d$. Following [6], setting

$$\Phi_{C_1, C_2; p}(y) := \int_{\mathbb{I}^{d-1}} |K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])|^p d\lambda_{d-1}(\mathbf{x}), \quad (6)$$

for every $y \in \mathbb{I}$, the metric D_p on $\mathcal{C}_{\mathbb{I}}^d$ is defined by

$$D_p(C_1, C_2) := \left\{ \int_{\mathbb{I}} \Phi_{C_1, C_2; p}(y) d\lambda(y) \right\}^{\frac{1}{p}} \quad (p \in [1, \infty)). \quad (7)$$

In the next section, we will extend this metric to families $\mathcal{C}_A^d$ and study its interrelation with the L^p -norm of regression functions.

For deriving sharp inequalities for the maximal deviation of r_C from the mean $\frac{1}{2}$, we will work with the Hardy-Littlewood-Pólya theorem (see [15, Ch. 1, Theorem D.2]) involving rearrangements. We therefore complete this section with some definitions concerning decreasing rearrangements and a simple lemma, which we will use in the sequel. For an arbitrary measurable function $f : \mathbb{I}^{d-1} \rightarrow \mathbb{R}$ and $a \in \mathbb{R}$, we define the a -superlevel set $[f]_a$ and the strict a -superlevel set $\langle f \rangle_a$ by

$$[f]_a := \left\{ \mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) \geq a \right\}$$

and the strict a -superlevel set by

$$\langle f \rangle_a := \left\{ \mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) > a \right\}.$$

For the remainder of this section let $C \in \mathcal{C}^d$ be arbitrary but fixed. Defining the functions $m_{f,C}, \bar{m}_{f,C} : \mathbb{R} \rightarrow \mathbb{I}$ by

$$\begin{aligned} m_{f,C}(v) &:= \mu_{C_{1:(d-1)}}(\langle f \rangle_v), \\ \bar{m}_{f,C}(v) &:= \mu_{C_{1:(d-1)}}([f]_v), \end{aligned} \quad (8)$$

obviously $m_{f,C}$ and $\bar{m}_{f,C}$ are non-increasing. Based on these two functions, the so-called decreasing rearrangements $f_{C,\downarrow}, \bar{f}_{C,\downarrow} : \mathbb{I} \rightarrow \mathbb{R}$ of f are given by

$$\begin{aligned} f_{C,\downarrow}(u) &:= \sup \{ v \in \mathbb{I} : m_{f,C}(v) > u \}, \\ \bar{f}_{C,\downarrow}(u) &:= \sup \{ v \in \mathbb{I} : \bar{m}_{f,C}(v) > u \}. \end{aligned}$$

Obviously $f_{C,\downarrow}(u)$ and $\bar{f}_{C,\downarrow}(u)$ are both non-increasing functions too. To the best of our knowledge, available literature is confined to the decreasing rearrangement $f_{C,\downarrow}$ in the situation $\mu_{C_{1:(d-1)}} = \lambda_{d-1}$. Since in the sequel we will work with non-strict superlevel sets, the following lemma, stating that there is no difference between these two versions, is useful (the proof can be found in Appendix A).

Lemma 2. *For each $d \geq 2$, $C \in \mathcal{C}^d$ and every measurable $f : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$, the two functions $f_{C,\downarrow}$ and $\bar{f}_{C,\downarrow}$ are identical.*

3. Mean Regression

We now focus on mean regression and start with the following simple observation, stating that, as a direct consequence of working with d -stochastic measures, the regression function r_C of C integrates to a constant not depending on the copula C .

Lemma 3. *For every $d \geq 2$ and every $C \in \mathcal{C}^d$ we have*

$$\int_{\mathbb{I}^{d-1}} r_C(\mathbf{x}) d\mu_{C_{1:(d-1)}}(\mathbf{x}) = \frac{1}{2} \equiv r_\Pi.$$

Proof. Using eq. (4), Fubini's theorem and disintegration, for any $C \in \mathcal{C}^d$, we directly get

$$\begin{aligned} \int_{\mathbb{I}^{d-1}} r_C(\mathbf{x}) d\mu_{C_{1:(d-1)}}(\mathbf{x}) &= \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} K_C(\mathbf{x}, (y, 1]) d\lambda(y) d\mu_{C_{1:(d-1)}}(\mathbf{x}) \\ &= \int_{\mathbb{I}} (1 - y) d\lambda(y) = \frac{1}{2}. \end{aligned}$$

The fact that $r_\Pi(\mathbf{x}) = \frac{1}{2}$ for every $\mathbf{x} \in \mathbb{I}^{d-1}$ is trivial. $\square$

In the sequel, given a fixed copula family $\mathcal{C}_A^d$, we consider $\|r_C - \frac{1}{2}\|_{A,p}$ as a measure for the average deviation of r_C from its mean, or equivalently, from the regression function of the copula corresponding to the d -stochastic product measure $\mu_A \otimes \lambda$, modeling independence of $\mathbf{X} \sim A$ and $Y \sim \lambda$. Although our focus is on the linkage class, we formulate and prove our main results for the general setting $\mathcal{C}_A^d$ with arbitrary $A \in \mathcal{C}^{d-1}$.

3.1. Bounds for the L^p -Deviation from the mean

In general, the magnitude of $\|r_C - \frac{1}{2}\|_{A,p}$ substantially depends on the copula $C \in \mathcal{C}_A^d$. An exception, however, occurs under complete dependence in the sense of [6, Lemma 5.4], i.e., in the case in which there exists some μ_A - λ -preserving transformation $h : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$ such that $K(\mathbf{x}, F) = \mathbb{1}_F(h(\mathbf{x}))$ is (a version of) the Markov kernel of C . Again, following [6], we will let $C_h \in \mathcal{C}_A^d$ denote the completely dependent copula induced by h .

Lemma 4 (complete dependence). *Suppose that $d \geq 2$ and let $C_h \in \mathcal{C}_A^d$ denote the completely dependent copula induced by the μ_A - λ -preserving transformation $h : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$. Then, $r_{C_h}(\mathbf{x}) = h(\mathbf{x})$ for μ_A -a.e. $\mathbf{x} \in \mathbb{I}^{d-1}$, and for every $p \in [1, \infty)$ we have*

$$\|r_{C_h} - \frac{1}{2}\|_{A,p} = \frac{1}{2}(p+1)^{-\frac{1}{p}}.$$

Proof. In the described setup we have $K_{C_h}(\mathbf{x}, [0, y]) = \mathbb{1}_{[0, y]}(h(\mathbf{x}))$, which, using eq. (4) directly yields

$$r_{C_h}(\mathbf{x}) = \int_{\mathbb{I}} K_{C_h}(\mathbf{x}, (y, 1]) \, d\lambda(y) = h(\mathbf{x}).$$

Moreover, applying change of coordinates and using the fact that the push-forward μ_A^h of μ_A via h coincides with λ it follows that

$$\begin{aligned} \|r_{C_h} - \frac{1}{2}\|_{A,p}^p &= \int_{\mathbb{I}^{d-1}} |h(\mathbf{x}) - \frac{1}{2}|^p \, d\mu_A(\mathbf{x}) = \int_{\mathbb{I}} |y - \frac{1}{2}|^p \, d\mu_A^h(y) \\ &= \int_{\mathbb{I}} |y - \frac{1}{2}|^p \, d\lambda(y). \end{aligned}$$

Elementary computations show $\int_{\mathbb{I}} |y - \frac{1}{2}|^p \, d\lambda(y) = \frac{1}{2}(p+1)^{-\frac{1}{p}}$, and the proof is complete. $\square$

It turns out that completely dependent copulas exhibit maximal average deviation from $\frac{1}{2}$:

Theorem 5 (upper bound). *For every $d \geq 2$ and every $p \in [1, \infty)$, we have*

$$\max_{C \in \mathcal{C}_A^d} \|r_C - \frac{1}{2}\|_{A,p} = \frac{1}{2}(p+1)^{-\frac{1}{p}}.$$

Proof. By definition of r_C and since $\int_{\mathbb{I}} K_C(\mathbf{x}, dy) = 1$, for each $\mathbf{x} \in \mathbb{I}^{d-1}$ we can write

$$r_C(\mathbf{x}) - \frac{1}{2} = \int_{\mathbb{I}} (y - \frac{1}{2}) K_C(\mathbf{x}, dy),$$

for every $\mathbf{x} \in \mathbb{I}^{d-1}$. Applying Jensen's inequality yields

$$|r_C(\mathbf{x}) - \frac{1}{2}|^p \leq \int_{\mathbb{I}} |y - \frac{1}{2}|^p K_C(\mathbf{x}, dy).$$

Hence, using disintegration, we altogether conclude that

$$\begin{aligned} \|r_C - \frac{1}{2}\|_{A,p}^p &\leq \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} |y - \frac{1}{2}|^p K_C(\mathbf{x}, dy) d\mu_A(\mathbf{x}) \\ &= \int_{\mathbb{I}} |y - \frac{1}{2}|^p d\lambda(y) = \frac{1}{p+1} 2^{-p}. \end{aligned}$$

This shows that for every copula $C \in \mathcal{C}_A^d$ the norm $\|r_C - \frac{1}{2}\|_{A,p}$ is bounded from above by $\frac{1}{2}(p+1)^{-\frac{1}{p}}$. Since Lemma 4 shows that this bound is attained, the proof of the theorem is complete. $\square$

The previous result allows to quantify the maximum distance between two regression functions associated with copulas from the $\mathcal{C}_A^d$ family.

Corollary 6. *For each $d \geq 2$ and every $p \in [1, \infty)$ the following identity holds:*

$$\max_{C_1, C_2 \in \mathcal{C}_A^d} \|r_{C_1} - r_{C_2}\|_{A,p} = (p+1)^{-\frac{1}{p}}.$$

Proof. The triangle inequality implies

$$\|r_{C_1} - r_{C_2}\|_{A,p} \leq \|r_{C_1} - \frac{1}{2}\|_{A,p} + \|r_{C_2} - \frac{1}{2}\|_{A,p},$$

so applying Theorem 5 we find that $\|r_{C_1} - r_{C_2}\|_{A,p}^p \leq (p+1)^{-1}$. Moreover, for $C \in \mathcal{C}_A^d$, obviously also $\overline{C} \in \mathcal{C}_A^d$, with $\|r_C - r_{\overline{C}}\|_{A,p} = 2\|r_C - \frac{1}{2}\|_{A,p}$. Hence, Lemma 4 confirms the sharpness of the asserted bound. $\square$

Our next example shows that completely dependent copulas are not the only copulas attaining the upper bound in Theorem 5.

Example 7. Assume $N \geq 2$, let σ^N be a permutation of $\{1, \dots, N\}$, and consider the (checkerboard) copula $C_N^\square$ with density $c_N^\square : \mathbb{I}^d \rightarrow \mathbb{I}$, given by

$$c_N^\square(\mathbf{x}, y) := N \sum_{i=1}^N \mathbb{1}_{I_{N,i} \times I_{N,\sigma^N(i)}}(x_1, y).$$

Notice that $c_N^\square(\mathbf{x}, y)$ does not depend on $x_2, \dots, x_{d-1}$ and that $C_N^\square$ is an element of the linkage family $\mathcal{C}_A^\Pi$. Considering that for $x_1 \in I_{N,i}$, the measure $K_{C_N^\square}(\mathbf{x}, \cdot)$ coincides with the uniform distribution on the interval $I_{N,\sigma^N(i)}$, it follows immediately that the regression function $r_{C_N^\square}$ is given by

$$r_{C_N^\square}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{I_{N,i}}(x_1) \left(\sigma^N(i) - \frac{1}{2} \right).$$

Thus, for every $p \in [1, \infty)$, we get

$$\|r_{C_N^\square} - \frac{1}{2}\|_{\Pi,p}^p = N^{-p-1} \sum_{i=1}^N |i - \frac{N+1}{2}|^p.$$

Specifically, for $p = 1$, we have

$$\|r_{C_N^\square} - \frac{1}{2}\|_{\Pi,1} = \frac{1}{4} \mathbb{1}_{\{N \in 2\mathbb{N}\}} + \frac{N^2-1}{4N^2} \mathbb{1}_{\{N \in 2\mathbb{N}_0+1\}},$$

so for even N the upper bound from Theorem 5 is attained. In general, using the Euler-Maclaurin summation formula shows that

$$\|r_{C_N^\square} - \frac{1}{2}\|_{\Pi,p} = \frac{1}{2}(p+1)^{-\frac{1}{p}} + o(1) \quad (N \rightarrow \infty).$$

Example 8 (cube copula). Consider the so-called cube copula $C^{\text{cube}} \in \mathcal{C}_\Pi^3$ introduced in [17, Example 3.4], i.e., the three-dimensional copula distributing its mass uniformly on the cubes $I_{2,1}^3$, $I_{2,2}^2 \times I_{2,1}$, $I_{2,2} \times I_{2,1} \times I_{2,2}$ and $I_{2,1} \times I_{2,2}^2$. For this copula each conditional distribution $K_{C^{\text{cube}}}(\mathbf{x}, \cdot)$ is either the uniform distribution on $I_{2,1}$ or on $I_{2,2}$. Therefore, it is straightforward to verify that the regression function $r_{C^{\text{cube}}}$ is given by

$$r_{C^{\text{cube}}}(\mathbf{x}) = \frac{1}{4} \mathbb{1}_{I_{2,1}^2 \cup I_{2,2}^2}(\mathbf{x}) + \frac{3}{4} \mathbb{1}_{(I_{2,2} \times I_{2,1}) \cup (I_{2,1} \times I_{2,2})}(\mathbf{x}) \quad (\mathbf{x} \in \mathbb{I}^2).$$

According to [17, Section 4.2] the partial vine copula $\psi(C^{\text{cube}})$ of C^{cube} is the independence/product copula Π_3 , whose regression function is $r_{\Pi_3} \equiv \frac{1}{2}$. As a direct consequence, we have

$$\|r_{C^{\text{cube}}} - r_{\psi(C^{\text{cube}})}\|_{\Pi,1} = \|r_{C^{\text{cube}}} - \frac{1}{2}\|_{\Pi,1} = \frac{1}{4}.$$

Considering that the according to Corollary 6 the maximal L^1 -distance of regression functions is at most $\frac{1}{2}$, the cube examples shows that the partial vine copula (which, in the context of pair copula constructions is commonly used as natural approximation of the original copula) can be a strikingly far off also from the regression perspective. In fact, for the cube example the error is 50% of the maximum possible distance. This simple observations underlines once more that working with partial vine copulas must be done with care, despite the frequently praised ‘flexibility’ (see [1, 17, 19] and the references therein).

We conclude this section with showing that the deviation from the mean can only decrease with reducing dimension.

Theorem 9 (dimension reduction). *For each $d \geq 2$, $C \in \mathcal{C}^d$ and $p \in [1, \infty)$ the following inequality holds:*

$$\left\| r_C - \frac{1}{2} \right\|_{C_{1:(d-1)}, p} \geq \left\| r_{C_{1:(d-2), d}} - \frac{1}{2} \right\|_{C_{1:(d-2)}, p}$$

Proof. Fix $p \in [1, \infty)$ and $d \geq 2$. Then, using disintegration, we obtain

$$\left\| r_C - \frac{1}{2} \right\|_{C_{1:(d-1)}, p}^p = \int_{\mathbb{I}^{d-2}} I(\mathbf{x}_{1:d-2}) d\mu_{C_{1:(d-2)}}(\mathbf{x}_{1:d-2})$$

with

$$I(\mathbf{x}_{1:d-2}) := \int_{\mathbb{I}} \left| r_C(\mathbf{x}_{1:d-2}, x_{d-1}) - \frac{1}{2} \right|^p K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}).$$

Since $x \mapsto |x|^p$ is a convex function on $[-1, 1]$, Jensen’s inequality yields

$$I(\mathbf{x}_{1:d-2}) \geq \left| \int_{\mathbb{I}} \left\{ r_C(\mathbf{x}_{1:d-2}, x_{d-1}) - \frac{1}{2} \right\} K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}) \right|^p.$$

On the one hand, we obviously have $\int_{\mathbb{I}} K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}) = 1$. On the other hand, representing the regression function through eq. (4) and using Fubini’s theorem, we obtain

$$I(\mathbf{x}_{1:d-2}) \geq \left| \int_{\mathbb{I}} \int_{\mathbb{I}} K_C(\mathbf{x}_{1:d-2}, x_{d-1}, (y, 1]) K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}) d\lambda(y) - \frac{1}{2} \right|^p.$$

Hence, from the disintegration identity (2), we infer that

$$\begin{aligned} \|r_C - \tfrac{1}{2}\|_{C_{1:(d-1)},p}^p &\geq \int_{\mathbb{I}} \left| \int_{\mathbb{I}} K_{C_{1:(d-2),d}}(\mathbf{x}_{1:d-2}, (y, 1]) \, d\lambda(y) - \tfrac{1}{2} \right|^p d\mu_{C_{1:(d-2)}}(\mathbf{x}_{1:d-2}) \\ &= \left\| r_{C_{1:(d-2),d}} - \tfrac{1}{2} \right\|_{C_{1:(d-2)},p}^p, \end{aligned}$$

and the proof is complete. $\square$

3.2. Best-possible bounds for the distribution function of the $\overline{m}_{|r_C - \frac{1}{2}|,C}$

Considering the function $f : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$, given by $f(\mathbf{x}) = |r_C(\mathbf{x}) - \frac{1}{2}|$, in this section we will study properties of the function

$$\begin{aligned} a \mapsto \overline{m}_{|r_C - \frac{1}{2}|,C}(a) &= \mu_{C_{1:(d-1)}}([|r_C - \tfrac{1}{2}|]_a) \\ &= \mu_{C_{1:(d-1)}}(\{\mathbf{x} \in \mathbb{I}^{d-1} : |r_C(\mathbf{x}) - \tfrac{1}{2}| \geq a\}), \end{aligned}$$

for $a \in [0, \frac{1}{2}]$. In other words, we study the (right-continuous version of the) survival function of the random variable $|r_C - \frac{1}{2}|$ on the probability space $(\mathbb{I}^{d-1}, \mathcal{B}(\mathbb{I}^{d-1}), \mu_{C_{1:(d-1)}})$. In the sequel we will simply write $m_{|r_C - \frac{1}{2}|}$ instead of $m_{|r_C - \frac{1}{2}|,C}$ since the dependence on C is already indicated by r_C and no confusion will arise. Considering

$$[|r_C - \tfrac{1}{2}|]_a = r_C^{-1}([0, \tfrac{1}{2} - a]) \cup r_C^{-1}([\tfrac{1}{2} + a, 1]),$$

and using $r_C^{-1}([\tfrac{1}{2} + a, 1]) = [r_C]_{a + \frac{1}{2}}$ as well as $r_C^{-1}([0, \tfrac{1}{2} - a]) = [-r_C]_{a - \frac{1}{2}}$ directly yields the following alternative expression for $\overline{m}_{|r_C - \frac{1}{2}|}(a)$, which we will use in some of the proofs:

$$\overline{m}_{|r_C - \frac{1}{2}|}(a) = \overline{m}_{r_C}(a + \tfrac{1}{2}) + \overline{m}_{-r_C}(a - \tfrac{1}{2}). \quad (9)$$

As for the L_p -norm we first have a look at the completely dependent case.

Example 10. Let $d \geq 2$ and suppose that $C_h \in \mathcal{C}_A^d$ is completely dependent with μ_A - λ -preserving transformation $h : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$. Then, according to Lemma 4, we have $r_C(\mathbf{x}) = h(\mathbf{x})$, so for every $a \in [0, \frac{1}{2}]$

$$\overline{m}_{|r_{C_h} - \frac{1}{2}|}(a) = 1 - 2a$$

follows immediately.

As we are going to show, in the bivariate case it turns out that a best-possible upper bound for $\overline{m}_{|r_{C_h}-\frac{1}{2}|}(a)$ can be obtained via a slight modification of the Hardy-Littlewood-Pólya theorem (see, e.g, [15, Ch. 1, Theorem D.2]). After having shown the result in the bivariate setting, we will then tackle its extension to arbitrary dimension $d \geq 3$, using a measure-isomorphism argument in the following sense: It is well known (see [23, Theorem 2.1]) that for arbitrary but fixed $A \in \mathcal{C}^{d-1}$, the probability spaces $(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda_A)$ and $(\mathbb{I}^{d-1}, \mathcal{B}(\mathbb{I}^{d-1}), \mu_A)$ are isomorphic (since the latter has no point masses). To be precise, there exist some Borel sets $\Lambda_1 \in \mathcal{B}(\mathbb{I})$ and $\Lambda_{d-1} \in \mathcal{B}(\mathbb{I}^{d-1})$ with $\lambda(\Lambda_1) = \mu_A(\Lambda_{d-1}) = 1$ and a measurable (with respect to the trace σ -fields) bijection

$$\iota_A : \Lambda_{d-1} \rightarrow \Lambda_1, \quad (10)$$

such that the push-forward $\mu_A^{\iota_A}$ of μ_A via ι_A coincides with λ . Extending ι_A to an $\mathbb{I}$ -valued measurable transformation on $\mathbb{I}^{d-1}$ (by, e.g., setting $\iota_A(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{I}^{d-1} \setminus \Lambda_{d-1}$) and defining the mapping $\Psi_A : \mathbb{I}^d \rightarrow \mathbb{I}^2$ by

$$\Psi_A(\mathbf{x}, y) = (\iota_A(\mathbf{x}), y), \quad (11)$$

the following simple lemma holds.

Lemma 11. *For every fixed $A \in \mathcal{C}^{d-1}$ and every $C \in \mathcal{C}_A^d$, the measure $\mu_C^{\Psi_A}$ is doubly stochastic. In particular, Ψ_A can be interpreted as mapping from $\mathcal{C}_A^d$ to $\mathcal{C}^2$.*

Proof. Since the push-forward $\mu_C^{\Psi_A}$ obviously is a probability measure on $\mathcal{B}(\mathbb{I}^2)$, it suffices to show that $\mu_C^{\Psi_A}$ is doubly stochastic. Letting $E \in \mathcal{B}(\mathbb{I})$ be fixed, we have

$$\mu_C^{\Psi_A}(E \times \mathbb{I}) = \mu_C(\iota_A^{-1}(E) \times \mathbb{I}) = \mu_A(\iota_A^{-1}(E)) = \mu_A^{\iota_A}(E) = \lambda(E).$$

Since the property $\mu_C^{\Psi_A}(\mathbb{I} \times E) = \lambda(E)$ is obvious, the proof is complete. $\square$

Building upon the previous lemma we now prove the following main result of this section.

Theorem 12. *For each $d \geq 2$, $A \in \mathcal{C}^{d-1}$ and $C \in \mathcal{C}_A^d$ the following inequality holds for every $a \in [0, \frac{1}{2}]$:*

$$\overline{m}_{|r_C-\frac{1}{2}|}(a) \leq \min\{1, 2 - 4a\}$$

Proof. We proceed in two steps, first derive the result for $d = 2$ and then extend to general dimension $d \geq 3$. (i) First, consider $C \in \mathcal{C}^2$ and set $q(y) := y$ for every $y \in \mathbb{I}$. According to Jensen's inequality for conditional expectation, for any convex $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ and $(X, Y) \sim C$, it holds that $\varphi(r_C(X)) = \varphi(\mathbb{E}[Y|X]) \leq \mathbb{E}[\varphi(Y)|X]$, implying $\mathbb{E}[\varphi(r_C(X))] \leq \mathbb{E}[\varphi(q(Y))]$, i.e., $r_C(X)$ is dominated by $q(Y) = Y$ in convex order. Consequently, referring to the Hardy-Littlewood-Polya theorem (see [15, Ch. 1, Theorem D.2]), we conclude that

$$\int_{[0,t]} r_{C,\downarrow}(u) d\lambda(u) \leq \int_{[0,t]} q_{\downarrow}(u) d\lambda(u)$$

for every $t \in \mathbb{I}$. Moreover, considering $m_q(v) = 1 - v$ and $q_{\downarrow}(u) = 1 - u$ yields

$$\int_{[0,t]} r_{C,\downarrow}(u) d\lambda(u) \leq t - \frac{t^2}{2} \quad (12)$$

for every $t \in \mathbb{I}$. At the same time, recalling that $r_{C,\downarrow} : \mathbb{I} \rightarrow \mathbb{I}$ is decreasing, obviously $r_{C,\downarrow}(u) \geq \frac{1}{2} + a$ for every $u \in [0, \bar{m}_{r_C}(a + \frac{1}{2})]$. Considering $t := \bar{m}_{r_C}(a + \frac{1}{2})$ in eq. (12) we thus arrive at

$$\left(\frac{1}{2} + a\right) \bar{m}_{r_C}\left(a + \frac{1}{2}\right) \leq \bar{m}_{r_C}\left(a + \frac{1}{2}\right) - \frac{1}{2} \left\{\bar{m}_{r_C}\left(a + \frac{1}{2}\right)\right\}^2,$$

which implies that $\bar{m}_{r_C}(a + \frac{1}{2}) \leq 1 - 2a$ holds for every $a \in [0, \frac{1}{2}]$. Regarding $\bar{m}_{-r_C}(a - \frac{1}{2})$, we can write

$$\bar{m}_{-r_C}\left(a - \frac{1}{2}\right) = \bar{m}_{1-r_C}\left(a + \frac{1}{2}\right).$$

In addition, $\mathbb{E}[\varphi(1 - r_C(X))] \leq \mathbb{E}[\varphi(1 - Y)]$ for every convex φ . Therefore, proceeding analogously to the first part, it follows that $\bar{m}_{-r_C}(a - \frac{1}{2}) \leq 1 - 2a$. Finally, considering eq. (9) it altogether follows that

$$\bar{m}_{|r_C - \frac{1}{2}|}(a) \leq \min\{1, 2 - 4a\},$$

which completes the proof for $d = 2$.

(ii) Suppose now that $d \geq 3$ and that $C \in \mathcal{C}_A^d$ is arbitrary but fixed. Interpreting Ψ_A as mapping from $\mathcal{C}_A^d$ to $\mathcal{C}^2$, Lemma 11 implies that for μ_A -almost every $\mathbf{x} \in \Lambda_{d-1}$

$$K_C(\mathbf{x}, \cdot) = K_{\Psi_A(C)}(\iota_A(\mathbf{x}), \cdot)$$

holds. This, however, yields that for every $E \in \mathcal{B}(\mathbb{I})$ we have

$$\mu_A\left(\{\mathbf{x} \in \mathbb{I}^{d-1} : r_C(\mathbf{x}) \in E\}\right) = \lambda\left(\{x \in \mathbb{I} : r_{\Psi_A(C)}(x) \in E\}\right). \quad (13)$$

As a direct consequence, the two functions $\overline{m}_{|r_C - \frac{1}{2}|}$ and $\overline{m}_{|r_{\Psi_A(C)} - \frac{1}{2}|}$ coincide, so the desired inequality follows from case (i). $\square$

Completely dependent copulas turned out to maximize the L^p -norm, considering Example 10, however, they apparently do not attain the upper bound according to Theorem 12. Nevertheless, as the following example shows, the upper bound can be attained.

Example 13. Let $A \in \mathcal{C}^{d-1}$ be arbitrary but fixed. For $b \in (0, \frac{1}{2})$, let $O_2 \in \mathcal{C}^2$ denote the ordinal sum (see [4, §3.8]) of the independence copula Π with respect to the segments $(0, b)$, $(b, 1 - b)$ and $(1 - b, 1)$ (see Figure 1). Setting

$$K(\mathbf{x}, [0, y]) := K_{O_2}(x_1, [0, y])$$

for all $(\mathbf{x}, y) \in \mathbb{I}^d$ obviously defines a Markov kernel of a copula $O_A \in \mathcal{C}_A^d$, given by

$$O_A(\mathbf{x}, y) = \int_{[0, \mathbf{x}]} K(\mathbf{t}, [0, y]) \, d\mu_A(\mathbf{t})$$

for every $(\mathbf{x}, y) \in \mathbb{I}^d$. By construction, the regression function r_C of C is given by

$$r_C(\mathbf{x}) = \frac{b}{2} \mathbb{1}_{(0, b)}(x_1) + \frac{1}{2} \mathbb{1}_{(b, 1-b)}(x_1) + (1 - \frac{b}{2}) \mathbb{1}_{(1-b, 1)}(x_1).$$

So, in particular we have

$$|r_C(\mathbf{x}) - \frac{1}{2}| = \frac{1-b}{2} \mathbb{1}_{\mathbb{I} \setminus (b, 1-b)}(x_1)$$

for all $\mathbf{x} \in \mathbb{I}^{d-1}$, which directly yields

$$\overline{m}_{|r_C(x) - \frac{1}{2}|}(a) = \mathbb{1}_{\{a=0\}} + 2b \mathbb{1}_{(0, \frac{1-b}{2}]}(a)$$

for $a \in [0, \frac{1}{2}]$. Considering $a \in (\frac{1}{4}, \frac{1}{2}]$ and setting $b = 1 - 2a$, the upper bound from Theorem 12 is attained. To show exactness for $a \in [0, \frac{1}{4}]$, instead of O_2 consider a checkerboard copula $C_2^\square$ according to Example 7. This copula fulfills $|r_{C_2^\square}(\mathbf{x}) - \frac{1}{2}| = \frac{1}{4} \geq a$ for all $\mathbf{x} \in \mathbb{I}^{d-1}$.

Combining Theorem 12 and Example 13 yields the following corollary:

Corollary 14. *For every $d \geq 2$, every $A \in \mathcal{C}^{d-1}$ and every $a \in [0, \frac{1}{2}]$ we have*

$$\max_{C \in \mathcal{C}_A^d} \overline{m}_{|r_C - \frac{1}{2}|}(a) = \min\{1, 2 - 4a\}.$$

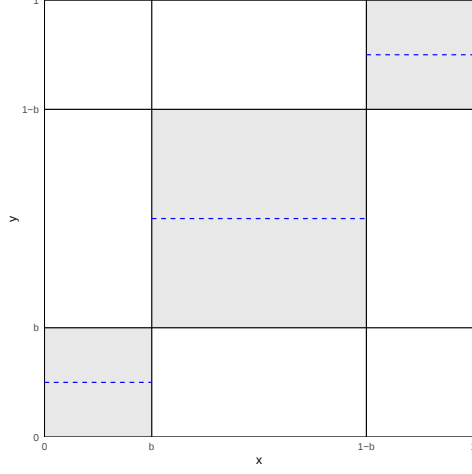


Figure 1: Support of the ordinal sum O_2 considered in Example 13 (gray) and regression function r_{O_2} (dashed blue line).

3.3. Relation to the metric D_p

We extend the D_p -metric according to eq. (7) to $\mathcal{C}_A^d$ by setting

$$D_{A,p}(C_1, C_2) := \left\{ \int_{[0,1]} \Phi_{C_1, C_2; A, p}(y) d\lambda(y) \right\}^{\frac{1}{p}} \quad (p \in [1, \infty)), \quad (14)$$

for all $C_1, C_2 \in \mathcal{C}_A^d$ where, analogously to (6), we define

$$\Phi_{C_1, C_2; A, p}(y) := \int_{\mathbb{I}^{d-1}} |K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])|^p d\mu_A(\mathbf{x}) \quad (15)$$

for every $y \in \mathbb{I}$. According to [6] $D_{\Pi, p}(C_1, C_2) = D_p(C_1, C_2)$ is a metric on $\mathcal{C}_{\Pi}^d$. The following more general result holds:

Lemma 15. *For each $d \geq 2$, every $p \in [1, \infty)$, and $A \in \mathcal{C}^{d-1}$, the mapping $D_{A,p} : \mathcal{C}_A^d \times \mathcal{C}_A^d \rightarrow [0, 1]$ establishes a metric on $\mathcal{C}_A^d$. This metric fulfills*

$$\|r_{C_1} - r_{C_2}\|_{A,p} \leq D_{A,p}(C_1, C_2) \quad (C_1, C_2 \in \mathcal{C}_A^d).$$

Proof. It is an easy exercise to verify the metric properties. For the proof of the asserted inequality, from (4) we directly get

$$\|r_{C_1} - r_{C_2}\|_{A,p}^p = \int_{\mathbb{I}^{d-1}} \left| \int_{\mathbb{I}} (K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])) \, d\lambda(y) \right|^p d\mu_A(\mathbf{x}).$$

Thus, an application of Jensen's inequality yields

$$\|r_{C_1} - r_{C_2}\|_{A,p}^p \leq \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} |K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])|^p \, d\lambda(y) \, d\mu_A(\mathbf{x}),$$

and the proof is complete. $\square$

An immediate consequence of Lemma 15 is the following result stating that for L_p -convergence of regression functions of elements in $\mathcal{C}_A^d$ weak convergence of μ_A -almost all conditional distributions is not necessary - convergence w.r.t. $D_{A,p}$ suffices.

Corollary 16. *$D_{A,p}$ -convergence of copulas in $\mathcal{C}_A^d$ implies L^p -convergence of the associated regression functions.*

The function $\Phi_{C_1, C_2} \equiv \Phi_{C_1, C_2; \Pi, 1}$ has already been studied in [21] and [6] in the bivariate and the multivariate linkage setting, respectively. The subsequent lemma summarizes the analogous properties for the function $\Phi_{C_1, C_2; A, p}$.

Lemma 17. *Suppose that $d \geq 2$, that $A \in \mathcal{C}^{d-1}$, and that $p \in [1, \infty)$. Then for arbitrary $C_1, C_2 \in \mathcal{C}_A^d$ the function $\Phi_{C_1, C_2; A, p}$ defined according to eq. (15) has the following properties:*

1. $\Phi_{C_1, C_2; A, p}(0) = \Phi_{C_1, C_2; A, p}(1) = 0$
2. $\Phi_{C_1, C_2; A, p}$ is continuous on $\mathbb{I}$, for $p = 1$ even 2-Lipschitz.
3. $\Phi_{C_1, C_2; A, p}(y) \leq 2 \min\{y, 1 - y\}$ for every $y \in \mathbb{I}$ and this bound is sharp.

Moreover the metric $D_{A,p}$ fulfills

$$\max_{C_1, C_2 \in \mathcal{C}_A^d} D_{A,p}(C_1, C_2) = 2^{-\frac{1}{p}}, \quad (16)$$

i.e., the diameter of the metric space $(\mathcal{C}_A^d, D_{A,p})$ is $2^{-\frac{1}{p}}$.

Proof. The boundary behavior of $\Phi_{C_1, C_2; A, p}$ is obvious. To prove the remaining assertions we proceed as follows. Recall that, according to Lemma 11, for every $C \in \mathcal{C}_A^d$ the measure $\tilde{C} := \mu_C^{\Psi_A}$ is doubly stochastic. Again using the isomorphism ι_A and the mapping Ψ_A according to eq. (11), for each $C \in \mathcal{C}_A^d$ and μ_A -almost every $\mathbf{x} \in \mathbb{I}^{d-1}$ we have $K_C(\mathbf{x}, [0, y]) = K_{\tilde{C}}(\iota_A(\mathbf{x}), [0, y])$. For $C_1, C_2 \in \mathcal{C}_A^d$ using change of coordinates we therefore get

$$\begin{aligned} \Phi_{C_1, C_2; A, p}(y) &= \int_{\mathbb{I}^{d-1}} |K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])|^p d\mu_A(\mathbf{x}) \\ &= \int_{\mathbb{I}^{d-1}} \left| K_{\tilde{C}_1}(\iota_A(\mathbf{x}), [0, y]) - K_{\tilde{C}_2}(\iota_A(\mathbf{x}), [0, y]) \right|^p d\mu_A(\mathbf{x}) \\ &= \int_{\mathbb{I}^{d-1}} \left| K_{\tilde{C}_1}(u, [0, y]) - K_{\tilde{C}_2}(u, [0, y]) \right|^p d\lambda(u) \\ &= \Phi_{\tilde{C}_1, \tilde{C}_2; \lambda, p}(y). \end{aligned}$$

Lipschitz-continuity of $y \mapsto \Phi_{C_1, C_2; A, 1}(y)$ is now a direct consequence of the bivariate setting analyzed in [21, Lemma 5]. For general $p > 1$, as a direct consequence of Dominated Convergence, $\Phi_{C_1, C_2; A, p}$ is right-continuous. Moreover, using the fact that we have $\mu_{C_1}(\mathbb{I}^{d-1} \times \{s\}) = 0 = \mu_{C_2}(\mathbb{I}^{d-1} \times \{s\})$ for every $s \in \mathbb{I}$, it follows that the function $y \mapsto \Phi_{C_1, C_2; A, 1}$ is also left-continuous, which completes the proof of the second assertion.

Considering $\Phi_{\tilde{C}_1, \tilde{C}_2; \lambda, p} \leq \Phi_{\tilde{C}_1, \tilde{C}_2; \lambda, 1}$ and again using [21, Lemma 5] we have $\Phi_{C_1, C_2; A, p}(y) \leq 2y\mathbb{I}_{[0, \frac{1}{2}]}(y) + 2(1-y)\mathbb{I}_{(\frac{1}{2}, 1]}(y)$, which yields the third assertion. Finally, by the very definition of $D_{A, p}$

$$\{D_{A, p}(C_1, C_2)\}^p \leq 2 \int_{[0, \frac{1}{2}]} y d\lambda(y) + 2 \int_{[\frac{1}{2}, 1]} (1-y) d\lambda(y) = \frac{1}{2}$$

follows and it only remains to show the existence of copulas $C_1, C_2 \in \mathcal{C}_A^d$ fulfilling $(D_{A, p}(C_1, C_2))^p = \frac{1}{2}$, which can be done as follows: Consider a completely dependent copula $C_h \in \mathcal{C}_A^d$ and its flipped counterpart $\bar{C}_h =$

$C_{1-h} \in \mathcal{C}_A^d$. Then, using change of coordinates, we get

$$\begin{aligned}\Phi_{C_h, C_{1-h}; A, p}(y) &= \int_{\mathbb{I}^{d-1}} |\mathbb{1}_{[0,y]}(h(\mathbf{x})) - \mathbb{1}_{[1-y,1]}(h(\mathbf{x}))|^p d\mu_A(\mathbf{x}) \\ &= \int_{\mathbb{I}} |\mathbb{1}_{[0,y]}(u) - \mathbb{1}_{[1-y,1]}(u)|^p d\lambda(u) \\ &= \int_{\mathbb{I}} |\mathbb{1}_{[0,y]}(u) - \mathbb{1}_{[1-y,1]}(u)|^p d\lambda(u) = 2 \min\{y, 1-y\},\end{aligned}$$

which completes the proof. $\square$

Combining Lemma 15 and Lemma 17 it follows that for all $d \geq 2$, $p \in [1, \infty)$ and $C_1, C_2 \in \mathcal{C}_A^d$

$$\|r_{C_1} - r_{C_2}\|_{A,p} \leq 2^{-\frac{1}{p}}$$

holds. This bound coincides with the bound from Corollary 6 for $p = 1$, while being too rough for $p > 1$.

4. Quantile Regression

After having derived various results on the regression function of copulas we now turn towards quantile regression and derive various analogous results. Doing so, we start with some first observations concerning the extreme cases of independence and complete dependence (in the linkage class).

Example 18 (Independence and complete dependence). Let $d \geq 2$ as well as $\tau \in (0, 1]$ be arbitrary but fixed. Then obviously for the product copula Π we have $Q_{\Pi}^{\tau}(\mathbf{x}) = \tau$ for all $\mathbf{x} \in \mathbb{I}^{d-1}$.

Considering the other extreme, let $h : \mathbb{I}^{d-1} \rightarrow \mathbb{I}$ denote a λ_{d-1} - λ -preserving transformation and C_h the corresponding induced completely dependent copula. In this case, $Q_{C_h}^{\tau}(\mathbf{x}) = h(\mathbf{x})$, for every $\mathbf{x} \in \mathbb{I}^{d-1}$, implying that

$$\int_{\mathbb{I}^{d-1}} Q_{C_h}^{\tau}(\mathbf{x}) d\mu_A(\mathbf{x}) = \int_{\mathbb{I}^{d-1}} h(\mathbf{x}) d\lambda_{d-1}(\mathbf{x}) = \int_{\mathbb{I}} z d\lambda_{d-1}^h(z) = \int_{\mathbb{I}} z d\lambda(z) = \frac{1}{2}.$$

In other words: the quantile function $Q_{C_h}^{\tau}$ integrates to $\frac{1}{2}$ for every $\tau \in (0, 1]$.

In the following we will derive (best-possible) inequalities for the average value

$$\int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) d\mu_A(\mathbf{x})$$

of the quantile function Q_C^τ for $C \in \mathcal{C}_A^d$. Doing so, we will use the following elementary identity (a generalization of which is usually referred to as the ‘layer cake representation’, see [14, Theorem 1.13]):

$$\int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) d\mu_A(\mathbf{x}) = \int_{\mathbb{I}} \overline{m}_{Q_C^\tau}(q) d\lambda(q) = \int_{\mathbb{I}} m_{Q_C^\tau}(q) d\lambda(q). \quad (17)$$

Theorem 19. *For each $d \geq 2$, $A \in \mathcal{C}^{d-1}$, $C \in \mathcal{C}_A^d$ and $\tau \in (0, 1]$ the following inequality holds:*

$$\frac{\tau}{2} \leq \int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) d\mu_A(\mathbf{x}) \leq \frac{\tau+1}{2}.$$

Thereby, the lower and the upper bound are best possible.

Proof. First, using disintegration we obviously have

$$\begin{aligned} q &= \int_{\mathbb{I}^{d-1}} K_C(\mathbf{x}, [0, q]) d\mu_A(\mathbf{x}) = \int_{\mathbb{I}^{d-1}} K_C(\mathbf{x}, [0, q]) d\mu_A(\mathbf{x}) \\ &= \int_{[Q_C^\tau]_q} K_C(\mathbf{x}, [0, q]) d\mu_A(\mathbf{x}) + \int_{\mathbb{I}^{d-1} \setminus [Q_C^\tau]_q} K_C(\mathbf{x}, [0, q]) d\mu_A(\mathbf{x}). \end{aligned}$$

For every $\mathbf{x} \in [Q_C^\tau]_q$ by construction of the τ -quantile function we have $K_C(\mathbf{x}, [0, y]) < \tau$ for every $y < q$, implying that $K_C(\mathbf{x}, [0, q]) \leq \tau$. The first integral is therefore bounded from above by $\tau \cdot \overline{m}_{Q_C^\tau}(q)$. Using the obvious upper bound $1 - \overline{m}_{Q_C^\tau}(q)$ for the second integral altogether yields

$$q \leq \tau \cdot \overline{m}_{Q_C^\tau}(q) + 1 - \overline{m}_{Q_C^\tau}(q),$$

from which it directly follows that

$$\overline{m}_{Q_C^\tau}(q) \leq \min \left\{ 1, \frac{1-q}{1-\tau} \right\}.$$

Applying eq. (17) and calculating the integral directly yields the upper bound of $\frac{\tau+1}{2}$.

To prove the lower point we proceed as follows: Using the fact that $Q_C^\tau(\mathbf{x}) < q$ implies $K_C(\mathbf{x}, [0, q]) \geq \tau$ it follows that

$$\begin{aligned} q &= \int_{\mathbb{I}^{d-1}} K_C(\mathbf{x}, [0, q]) \, d\mu_A(\mathbf{x}) \geq \int_{\mathbb{I}^{d-1} \setminus [Q_C^\tau]_q} K_C(\mathbf{x}, [0, q]) \, d\mu_A(\mathbf{x}) \\ &\geq \tau \mu_A(\mathbb{I}^{d-1} \setminus [Q_C^\tau]_q) = \tau(1 - \overline{m}_{Q_C^\tau}(q)). \end{aligned}$$

This directly yields

$$\overline{m}_{Q_C^\tau}(q) \geq \max\left\{0, 1 - \frac{q}{\tau}\right\}.$$

for every $q \in \mathbb{I}$. Again using eq. (17) and calculating the integral directly yields the lower bound $\frac{\tau}{2}$.

Finally, it remains to show that the established bounds are best-possible. For fixed $\tau \in (0, 1)$, defining $K : \mathbb{I}^{d-1} \times \mathcal{B}(\mathbb{I}) \rightarrow \mathbb{I}$ by

$$K(\mathbf{x}, E) = \tau \mathbb{1}_E(\tau x_1) + (1 - \tau) \mathbb{1}_E(\tau + (1 - \tau)x_1),$$

obviously K is the $(d-1)$ -Markov kernel of a unique copula $C \in \mathcal{C}_A^d$. For this very copula C the τ -quantile function Q_C^τ , however, is given by $Q_X^\tau(\mathbf{x}) = \tau x_1$, which yields

$$\int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) \, d\lambda_{d-1}(\mathbf{x}) = \int_{\mathbb{I}} \tau x_1 \, d\lambda(x_1) = \frac{\tau}{2},$$

so the lower bound is attainable. For showing that the upper bound is best-possible, consider $n \in \mathbb{N}$ sufficiently large, so that $\tau - \frac{1}{n} \in (0, 1)$ holds and set

$$K(\mathbf{x}, E) := (\tau - \frac{1}{n}) \mathbb{1}_E((\tau - \frac{1}{n})x_1) + (1 - \tau + \frac{1}{n}) \mathbb{1}_E(\tau - \frac{1}{n} + (1 - \tau + \frac{1}{n})x_1),$$

for every $\mathbf{x} \in \mathbb{I}^{d-1}$ and $E \in \mathcal{B}(\mathbb{I})$. Then, K is the $(d-1)$ -Markov kernel of a unique copula $C \in \mathcal{C}_A^d$, whose the τ -quantile function Q_C^τ obviously is given by $Q_C^\tau(\mathbf{x}) = \tau - \frac{1}{n} + (1 - \tau + \frac{1}{n})x_1 = Q_C^1(\mathbf{x})$. A straightforward calculation yields

$$\int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) \, d\lambda_{d-1}(\mathbf{x}) = \frac{1}{2} + \frac{1}{2}(\tau - \frac{1}{n}),$$

hence, considering $n \rightarrow \infty$ completes the proof. $\square$

Although the integral of the τ -quantile function Q_C^τ does not need to coincide with the value τ , as the following lemma shows, integrating over τ again yields the same constant for all copulas.

Lemma 20. For each $d \geq 2$ and $C \in \mathcal{C}_A^d$, we have

$$\int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) \, d\mu_A(\mathbf{x}) \, d\lambda(\tau) = \frac{1}{2}.$$

Proof. Considering eq. (17) and using the fact that $Q_C^\tau(\mathbf{x}) > q$ if, and only if $K_C(\mathbf{x}, [0, q]) < \tau$, directly yields

$$\begin{aligned} \int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) \, d\mu_A(\mathbf{x}) &= \int_{\mathbb{I}} \mu_A \left(\{\mathbf{x} \in \mathbb{I}^{d-1} : K_C(\mathbf{x}, [0, q]) < \tau\} \right) \, d\lambda(q) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \mathbb{1}_{[0, \tau)}(K_C(\mathbf{x}, [0, q])) \, d\mu_A(\mathbf{x}) \, d\lambda(q) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \mathbb{1}_{[K_C(\mathbf{x}, [0, q]), 1]}(\tau) \, d\mu_A(\mathbf{x}) \, d\lambda(q). \end{aligned}$$

Having this, applying Fubini's theorem and disintegration, we altogether get

$$\begin{aligned} \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} Q_C^\tau(\mathbf{x}) \, d\mu_A(\mathbf{x}) \, d\lambda(\tau) &= \int_{\mathbb{I}} \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \mathbb{1}_{[K_C(\mathbf{x}, [0, q]), 1]}(\tau) \, d\mu_A(\mathbf{x}) \, d\lambda(q) \, d\lambda(\tau) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} \mathbb{1}_{[K_C(\mathbf{x}, [0, q]), 1]}(\tau) \, d\lambda(\tau) \, d\mu_A(\mathbf{x}) \, d\lambda(q) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} K_C(\mathbf{x}, (q, 1]) \, d\mu_A(\mathbf{x}) \, d\lambda(q) \\ &= \int_{\mathbb{I}} (1 - q) \, d\lambda(q) = \frac{1}{2} \end{aligned}$$

and the proof is complete. $\square$

Proceeding analogous to the proof of the previous result, we conclude this section with a sharp upper bound for the average L^1 -distance of quantile functions for copulas in the family $\mathcal{C}_A^d$ and a direct consequence to the cube copula.

Theorem 21. For each $d \geq 2$, $A \in \mathcal{C}^{d-1}$, and arbitrary $C_1, C_2 \in \mathcal{C}_A^d$ the following inequality holds:

$$D_{A,1}(C_1, C_2) = \int_{\mathbb{I}} \|Q_{C_1}^\tau - Q_{C_2}^\tau\|_{A,1} \, d\lambda(\tau) \leq \frac{1}{2}. \quad (18)$$

This inequality is best-possible.

Proof. Our proof builds upon the facts that for arbitrary $a, b \in \mathbb{I}$ we have

$$|a - b| = \int_{\mathbb{I}} |\mathbb{1}_{[0,a]} - \mathbb{1}_{[0,b]}| d\lambda = \int_{\mathbb{I}} |\mathbb{1}_{(a,1]} - \mathbb{1}_{(b,1]}| d\lambda, \quad (19)$$

and that (by the definition of the quantile function) for every $C \in \mathcal{C}^d$, every $\mathbf{x} \in \mathbb{I}^{d-1}$, every $\tau \in (0, 1]$ and every $v \in [0, 1]$ the following equivalence holds: (i) $Q_C^\tau(\mathbf{x}) > v$ if, and only if (ii) $\tau > K_C(\mathbf{x}, [0, v])$.

Hence, setting

$$V := \int_{\mathbb{I}} \|Q_{C_1}^\tau - Q_{C_2}^\tau\|_{A,1} d\lambda(\tau)$$

and using Fubini's theorem, it follows that

$$\begin{aligned} V &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} \left| \mathbb{1}_{[0, Q_{C_1}^\tau(\mathbf{x})]}(v) - \mathbb{1}_{[0, Q_{C_2}^\tau(\mathbf{x})]}(v) \right| d\lambda(v) d\mu_A(\mathbf{x}) d\lambda(\tau) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} \left| \mathbb{1}_{(K_{C_1}(\mathbf{x}, [0, v]), 1]}(\tau) - \mathbb{1}_{(K_{C_2}(\mathbf{x}, [0, v]), 1]}(\tau) \right| d\lambda(v) d\mu_A(\mathbf{x}) d\lambda(\tau) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} \int_{\mathbb{I}} \left| \mathbb{1}_{(K_{C_1}(\mathbf{x}, [0, v]), 1]}(\tau) - \mathbb{1}_{(K_{C_2}(\mathbf{x}, [0, v]), 1]}(\tau) \right| d\lambda(\tau) d\mu_A(\mathbf{x}) d\lambda(v) \\ &= \int_{\mathbb{I}} \int_{\mathbb{I}^{d-1}} |K_{C_1}(\mathbf{x}, [0, y]) - K_{C_2}(\mathbf{x}, [0, y])| d\mu_A(\mathbf{x}) d\lambda(v) \\ &= D_{A,1}(C_1, C_2) \end{aligned}$$

Having this, using Lemma 17 with $p = 1$ completes the proof. $\square$

Example 22 (cube copula cont.). Again consider the three-dimensional copula $C^{\text{cube}} \in \mathcal{C}_{\Pi}^3$ from Example 8. In this case $K_{C^{\text{cube}}}(\mathbf{x}, \cdot)$ is either the uniform distribution on $I_{2,1}$ or on $I_{2,2}$ and it is straightforward to verify that the τ -quantile function $Q_{C^{\text{cube}}}^\tau$ is given by

$$Q_{C^{\text{cube}}}^\tau(\mathbf{x}) = \frac{\tau}{2} \mathbb{1}_{I_{2,1}^2 \cup I_{2,2}^2}(\mathbf{x}) + \frac{\tau+1}{2} \mathbb{1}_{(I_{2,2} \times I_{2,1}) \cup (I_{2,1} \times I_{2,2})}(\mathbf{x}).$$

Since the partial vine copula $\psi(C^{\text{cube}})$ coincides with the independence copula Π , we have

$$\left\| Q_{C^{\text{cube}}}^\tau - Q_{\psi(C^{\text{cube}})}^\tau \right\|_1 = \frac{1}{2} \left| \frac{\tau}{2} - \tau \right| + \frac{1}{2} \left| \frac{\tau+1}{2} - \tau \right| = \frac{1}{4},$$

for every $\tau \in (0, 1]$, so, having in mind Theorem 21 also from the perspective of quantile regression the approximation quality of the partial vine may be very poor.

5. Estimation in the bivariate setting

We conclude this paper with some results on estimating the mean and the quantile regression functions via the empirical checkerboard estimator in dimension $d = 2$. Suppose that $C \in \mathcal{C}^2$ is fixed and that $(U_1, V_1) \dots, (U_n, V_n)$ is a sample from $(U, V) \sim C$. As before let E_n denote the induced empirical copula and $\mathfrak{Cb}_{N(n)}(E_n)$ its checkerboard approximation with $N(n) = \lfloor n^s \rfloor$, for some fixed $s \in (0, \frac{1}{2})$. Then, as shown in [2] (also see [5, 6, 9]), the sequence $(\mathfrak{Cb}_{N(n)}(E_n))_{n \in \mathbb{N}}$ converges weakly conditional to C with probability 1, without any regularity/smoothness restrictions on C . In other words: With probability 1, for λ -almost every $x \in \mathbb{I}$, the conditional distributions $K_{\mathfrak{Cb}_{N(n)}(E_n)}(x, \cdot)$ converge weakly to $K_C(x, \cdot)$ for $n \rightarrow \infty$. Although weak conditional convergence has been proved in full generality in the afore-mentioned papers, to the best of our knowledge, asymptotics of the checkerboard estimator $\mathfrak{Cb}_{N(n)}(E_n)$ in its pure (unaggregated) as well as its aggregated form are still unknown.

Resturning to consistency in the regression context, the following result is a direct consequence of weak condition convergence.

Theorem 23. *Suppose that $C \in \mathcal{C}$ and that $(U_1, V_1) \dots, (U_n, V_n)$ is a sample from $(U, V) \sim C$ with empirical copula E_n . Furthermore set $N(n) := \lfloor n^s \rfloor$ for some fixed $s \in (0, \frac{1}{2})$. Then for λ -almost every $x \in \mathbb{I}$ we have*

$$\lim_{n \rightarrow \infty} r_{\mathfrak{Cb}_{N(n)}(E_n)}(x) = r_C(x),$$

so in particular

$$\lim_{n \rightarrow \infty} \|r_{\mathfrak{Cb}_{N(n)}(E_n)} - r_C\|_p = 0$$

holds for every $p \in [1, \infty)$.

Since weak convergence of a sequence $(F_n)_{n \in \mathbb{N}}$ of distribution functions to a distribution function F is equivalent to pointwise convergence of the corresponding quasi-inverses $(F_n^-)_{n \in \mathbb{N}}$ in every continuity point of F^- (see, e.g., [22]), the afore-mentioned property on weak conditional convergence implies the following: There exists some $\Lambda \in \mathcal{B}(\mathbb{I})$ with $\lambda(\Lambda) = 1$, such that for every $x \in \Lambda$ and every continuity point $\tau \in \mathbb{I}$ of $\tau \mapsto Q_C^\tau(x)$ we have

$$\lim_{n \rightarrow \infty} Q_{\mathfrak{Cb}_{N(n)}(E_n)}^\tau(x) = Q_C^\tau(x).$$

For proving our second main result of this section - consistency of the empirical checkerboard estimator for quantile function - we will use the following technical lemma, in which the set S_q for $q \in (0, 1)$ is defined by

$$S_q := \{x \in \mathbb{I} : q \text{ is a discontinuity point of } \tau \mapsto Q_C^\tau(x)\} \in \mathcal{B}(\mathbb{I}). \quad (20)$$

Lemma 24. *There are at most countably many $q \in (0, 1)$ with $\lambda(S_q) > 0$.*

Proof. Suppose that $q \in (0, 1)$ is a discontinuity point of $\tau \mapsto Q_C^\tau(x)$. Then, left-continuity implies that $Q_C^q(x) < Q_C^{q+}(x)$, where $Q_C^{q+}(x)$ denotes the right-hand limit of $\tau \mapsto Q_C^\tau(x)$ at q . Hence, by definition of the quantile function, there exists some $\Delta > 0$, such that $y \mapsto K_C(x, [0, y])$ is constant on the interval $[Q_C^q(x), Q_C^q(x) + \Delta]$. In case S_q fulfills $\lambda(S_q) > 0$, the previous observation directly implies that

$$\lambda_2(\{(x, y) \in \mathbb{I}^2 : K_C(x, [0, y]) = q\}) > 0.$$

Setting $\Psi_C(x, y) := K_C(x, [0, y])$ and letting $[\Psi_C]_z$ denote the upper z -level of Ψ_C , we obviously have that the function $\ell : \mathbb{I} \rightarrow \mathbb{I}$, defined by

$$\ell(z) = \lambda_2([\Psi_C]_z),$$

is non-increasing on $\mathbb{I}$. Every q fulfilling $\lambda(S_q) > 0$ obviously is a discontinuity point of ℓ . As non-increasing function, ℓ can have at most countably many discontinuity points, and the proof is complete. $\square$

As an immediate consequence of Lemma 24, the following statements hold.

Theorem 25. *Suppose that $C \in \mathcal{C}$, that $(U_1, V_1) \dots, (U_n, V_n)$ is a sample from $(U, V) \sim C$, and that E_n is the empirical copula. Furthermore set $N(n) := \lfloor n^s \rfloor$, for some fixed $s \in (0, \frac{1}{2})$. Then, for all but at most countably many $\tau \in (0, 1)$ and λ -almost every x , we have*

$$\lim_{n \rightarrow \infty} Q_{\mathfrak{C}_{N(n)}(E_n)}^\tau(x) = Q_C^\tau(x).$$

In particular, for all but at most countably many $\tau \in (0, 1)$ and every $p \in [1, \infty)$, it holds that

$$\lim_{n \rightarrow \infty} \|Q_{\mathfrak{C}_{N(n)}(E_n)}^\tau - Q_C^\tau\|_p = 0.$$

We conclude our discussion with two concrete examples - a Marshall-Olkin copula as well as a Clayton copula - and a small simulation study illustrating the speed of convergence.

Example 26 (Marshall-Olkin). It is well-known (see [4, §6.4] that the Marshall-Olkin copula $M_{\alpha, \beta} \in \mathcal{C}^2$, for $\alpha, \beta > 0$ and $a(x) := x^{\frac{\alpha}{\beta}}$, is given by

$$M_{\alpha, \beta}(x, y) := x^{1-\alpha} y \mathbb{1}_{[0, a(x)]}(y) + xy^{1-\beta} \mathbb{1}_{(a(x), 1]}(y) \quad ((x, y) \in \mathbb{I}^2).$$

As shown in [21], the associated Markov kernel for $x \in (0, 1)$ is given by

$$K_{M_{\alpha,\beta}}(x, [0, y]) = (1 - \alpha)x^{-\alpha}y\mathbb{1}_{[0,a(x))}(y) + y^{1-\beta}\mathbb{1}_{(a(x),1]}(y),$$

implying that $K_{M_{\alpha,\beta}}(x, \cdot)$ has a point mass at $y = a(x)$. In fact, we have $y^- < y^+$ with

$$\begin{aligned} y^- &:= K_{M_{\alpha,\beta}}(x, [0, a(x)-]) = (1 - \alpha)x^{\alpha(\frac{1}{\beta}-1)} \\ y^+ &= K_{M_{\alpha,\beta}}(x, [0, a(x)]) = x^{\alpha(\frac{1}{\beta}-1)} \end{aligned}$$

As a direct consequence the quantile function is of the form

$$Q_{M_{\alpha,\beta}}^\tau(x) = \frac{\tau x^\alpha}{1-\alpha}\mathbb{1}_{[0,y^-)}(\tau) + a(x)\mathbb{1}_{[y^-,y^+]}(\tau) + \tau^{\frac{1}{1-\beta}}\mathbb{1}_{(y^+,1]}(\tau).$$

Moreover, using eq. (4) it is straightforward to verify that the regression function is given by

$$r_{M_{\alpha,\beta}}(x) = 1 - \frac{1-\alpha}{2}x^{\alpha(\frac{2}{\beta}-1)} - \frac{1-x^{\alpha(\frac{2}{\beta}-1)}}{2-\beta} \quad (x \in \mathbb{I}).$$

Notice that for $M_{\alpha,\beta}$ there is no $q \in (0, 1)$ fulfilling $\lambda(S_q) > 0$, with S_q according to eq. (20).

As second example we consider a member of the Archimedean family. In this case the regression function then does not admit an elementary analytic form.

Example 27 (Clayton). The Clayton copula (see [7, Example 2.1.5]), with $\theta > 0$, is defined by

$$C_\theta(x, y) = (x^{-\theta} + y^{-\theta} - 1)^{-\frac{1}{\theta}} \quad ((x, y) \in \mathbb{I}^2).$$

By direct computation, one thus verifies that

$$K_{C_\theta}(x, [0, y]) = x^{-\theta-1}(x^{-\theta} + y^{-\theta} - 1)^{-\frac{1}{\theta}-1} \quad ((x, y) \in \mathbb{I}^2).$$

Obviously, the Markov kernel is a strictly increasing function of y , we have $K_{C_\theta}(x, [0, Q_{C_\theta}^\tau(x)]) = \tau$ as well as

$$Q_{C_\theta}^\tau(x) = \left(1 + x^{-\theta}(\tau^{-\frac{\theta}{\theta+1}} - 1)\right)^{-\frac{1}{\theta}}.$$

With this, using change of coordinates we easily obtain $r_{C_\theta}(x) = \int_0^1 Q_{C_\theta}^y(x) dy$ or

$$r_{C_\theta}(x) = \int_0^1 \left(1 + x^{-\theta} (y^{-\frac{\theta}{\theta+1}} - 1)\right)^{-\frac{1}{\theta}} d\lambda(y).$$

As in the previous example there is no $q \in (0, 1)$ fulfilling $\lambda(S_q) > 0$, so the empirical checkerboard estimator is strongly consistent for every quantile $\tau \in (0, 1]$.

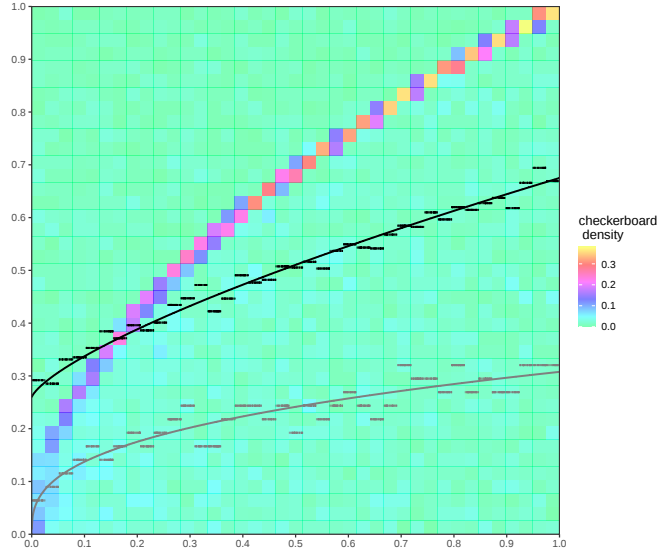


Figure 2: Empirical $N = 63$ checkerboard density for a sample of size $n = 10,000$ from the Marshall Olkin copula with parameters $(\alpha, \beta) = (0.35, 0.65)$, true mean and median regression functions (solid black and gray line, respectively), and corresponding estimators $r_{\mathfrak{C}_N(E_n)}$ and $Q_{\mathfrak{C}_N(E_n)}^{0.2}$ (black and gray step functions).

We close this section with a small simulation study illustrating the estimation procedure and the speed of convergence of the involved, checkerboard-based estimators and consider the Marshall Olkin copula with parameters $(\alpha, \beta) = (0.35, 0.65)$ and a Clayton copula with $\theta = 2$. Figures 2 and 3 depict the density of the empirical N -checkerboards for a sample of size $n = 10,000$ and resolution $N = \lfloor n^{0.4} \rfloor$. The black and the gray solid lines correspond to the true mean and quantile regression function, respectively, whose explicit

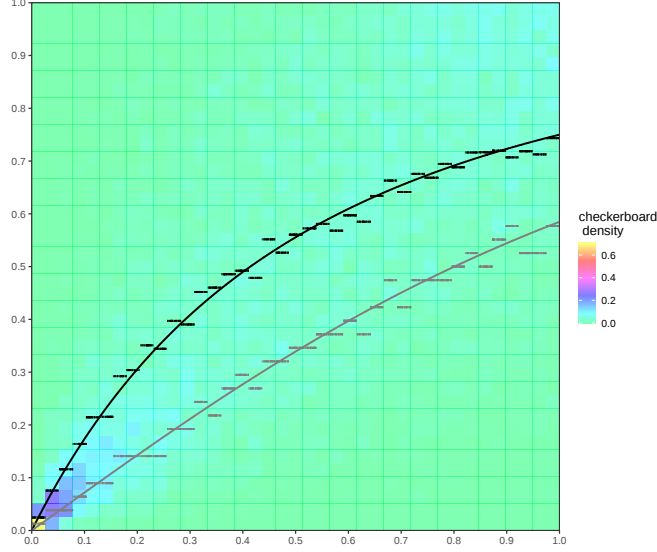


Figure 3: Empirical $N = 63$ checkerboard density for a sample of size $n = 10.000$ from the Clayton copula with parameter $\theta = 2$, true mean and median regression functions (solid black and gray line, respectively), and the corresponding estimators $r_{\mathfrak{C}_N(E_n)}$ and $Q_{\mathfrak{C}_N(E_n)}^{0.2}$ (black and gray step functions).

expressions were derived in Examples 26 and 27. The black and gray step functions correspond to $r_{\mathfrak{C}_N(E_n)}$ and $Q_{\mathfrak{C}_N(E_n)}^{0.2}$, respectively.

In addition, Figure 4 illustrates the speed of convergence of $r_{\mathfrak{C}_N(E_n)}$ and $Q_{\mathfrak{C}_N(E_n)}^{0.2}$. For each of the samples sizes n mentioned on the x -axis we drew a sample of n (from the considered copula), numerically calculated

$$\|r_{\mathfrak{C}_{N(n)}(E_n)} - r_C\|_1, \quad \|Q_{\mathfrak{C}_{N(n)}(E_n)}^T - Q_C^T\|_1,$$

repeated the procedure $R = 500$ times and summarized the obtained results as boxplots. All computations were performed in R using the packages `copula` and `qad`.

Acknowledgements

Both authors gratefully acknowledge the support of the WISS 2025 project ‘IDA-lab Salzburg’ (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

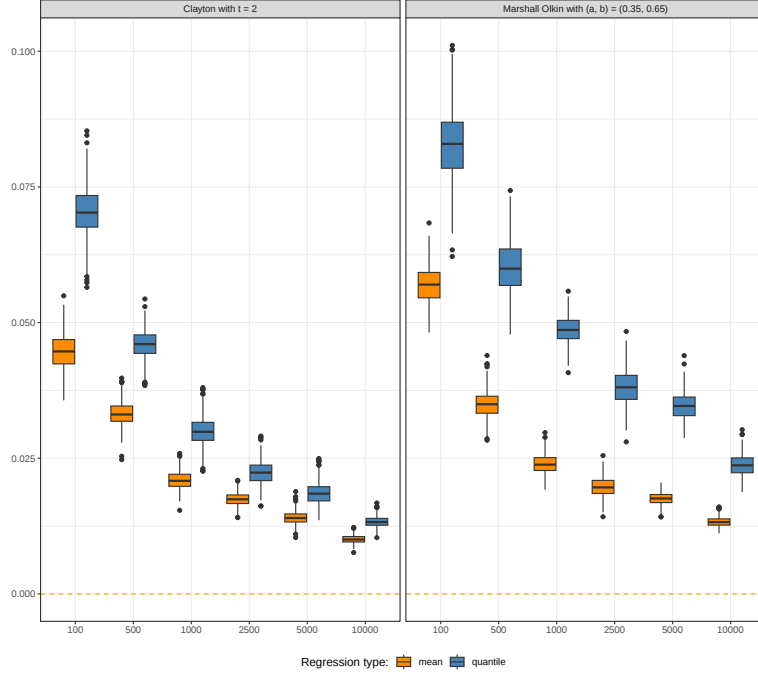


Figure 4: Boxplot summarizing the L^1 -distances between the estimated and the true mean regression as well as the estimated and the true median regression function, respectively. For each sample size n (on the x -axis) a total of $R = 500$ runs were performed.

Appendix A. Proofs to Section 2

Proof of Lemma 1. For brevity, we write

$$Q_d(\mathbf{x}_{1:d-2}, B) := \int_{\mathbb{I}} K_C(\mathbf{x}_{1:d-1}, B) K_{C_{1:(d-1)}}(\mathbf{x}_{1:d-2}, dx_{d-1}).$$

Clearly, for fixed $\mathbf{x}_{1:d-2} \in \mathbb{I}^{d-2}$, the assignment $B \mapsto Q_d(\mathbf{x}_{1:d-2}, B)$ fulfills the properties of a probability measure. Furthermore, by [11, Lemma 14.23], for fixed $B \in \mathcal{B}(\mathbb{I})$ the mapping $\mathbf{x}_{1:d-2} \mapsto Q_d(\mathbf{x}_{1:d-2}, B)$ is Borel measurable. In other words: $Q_d : \mathbb{I}^{d-2} \times \mathcal{B}(\mathbb{I}) \rightarrow \mathbb{I}$ is a Markov kernel and it remains to show that it is a Markov kernel of $C_{1:(d-2),d}$. For $E_1, E_2, \dots, E_{d-2}, E_d \in \mathcal{B}(\mathbb{I})$, setting $\mathbf{E} = E_1 \times E_2 \times \dots \times E_{d-2} \times \mathbb{I} \times E_d$ and using disintegration twice

(first for μ_C , then for $\mu_{C_{1:(d-1)}}$) yields

$$\begin{aligned}
\mu_C(\mathbf{E}) &= \int_{E_1 \times E_2 \times \dots \times E_{d-2} \times \mathbb{I}} K_C(\mathbf{x}_{1:d-1}, E_d) \, d\mu_{C_{1:d-1}}(\mathbf{x}_{1:d-1}) \\
&= \int_{E_1 \times E_2 \times \dots \times E_{d-2}} \int_{\mathbb{I}} K_C(\mathbf{x}_{1:d-1}, E_d) K_{C_{1:d-1}}(\mathbf{x}_{1:d-2}, dx_{d-1}) \, d\mu_{C_{1:d-2}}(\mathbf{x}_{1:d-2}) \\
&= \int_{E_1 \times E_2 \times \dots \times E_{d-2}} Q_d(\mathbf{x}_{1:d-2}, E_d) \, d\mu_{C_{1:d-2}}(\mathbf{x}_{1:d-2})
\end{aligned}$$

Considering $\mu_C(\mathbf{E}) = \mu_{C_{1:(d-2),d}}(E_1 \times \dots \times E_{d-2} \times E_d)$ and using the fact that the family of all rectangles of the form $E_1 \times \dots \times E_{d-2} \times E_d$ constitute a semiring generating $\mathcal{B}(\mathbb{I}^{d-1})$ this completes the proof. $\square$

Proof of Lemma 2. For every $v \in \mathbb{I}$ and $n \in \mathbb{N}$ we obviously have

$$\begin{aligned}
\{\mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) \geq v + \frac{1}{n}\} &\subseteq \{\mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) > v\} \subseteq \{\mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) \geq v\} \\
&\subseteq \{\mathbf{x} \in \mathbb{I}^{d-1} : f(\mathbf{x}) \geq v - \frac{1}{n}\},
\end{aligned}$$

which directly yields

$$\overline{m}_{f,C}(v + \frac{1}{n}) \leq m_{f,C}(v) \leq \overline{m}_{f,C}(v) \leq \overline{m}_{f,C}(v - \frac{1}{n}).$$

Considering $n \rightarrow \infty$ and using monotonicity of all involved functions we get

$$\overline{m}_{f,C}(v+) \leq m_{f,C}(v) \leq \overline{m}_{f,C}(v) \leq \overline{m}_{f,C}(v-).$$

As a direct consequence, we can only have $m_{f,C}(v) \neq \overline{m}_{f,C}(v)$ if v is a discontinuity point of $\overline{m}_{f,C}$. By monotonicity of $\overline{m}_{f,C}$, however, the set of discontinuity points of $\overline{m}_{f,C}$ is at most countably infinite, i.e.,

$$m_{f,C}(v) = \overline{m}_{f,C}(v)$$

holds outside an at most countably infinite set. In particular, it follows that $V := \{v \in \mathbb{I} : m_{f,C}(v) = \overline{m}_{f,C}(v)\}$ is dense in $\mathbb{I}$. Having this, it is straightforward to show that $f_{C,\downarrow}$ and $\overline{f}_{C,\downarrow}$ coincide on $\mathbb{I}$. $\square$

References

- [1] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insur. Math. Econ.*, 44(2):182–198.
- [2] Ansari, J., Langthaler, P. B., Fuchs, S., and Trutschnig, W. (2025). Quantifying and estimating dependence via sensitivity of conditional distributions. *Bernoulli*. to appear.
- [3] Dette, H., Van Hecke, R., and Volgushev, S. (2014). Some comments on copula-based regression. *J. Am. Stat. Assoc.*, 109(507):1319–1324.
- [4] Durante, F. and Sempi, C. (2016). *Principles of copula theory*. Boca Raton, FL: CRC Press.
- [5] Fuchs, S. and Trutschnig, W. (2020). On quantile based co-risk measures and their estimation. *Depend. Model.*, 8:396–416.
- [6] Griessenberger, F., Junker, R. R., and Trutschnig, W. (2022). On a multivariate copula-based dependence measure and its estimation. *Electron. J. Stat.*, 16(1):2206–2251.
- [7] Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018). *Elements of copula modeling with R*. Use R. Cham: Springer.
- [8] Joe, H. (2014). *Dependence Modeling with Copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.
- [9] Junker, R. R., Griessenberger, F., and Trutschnig, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Stat. Data Anal.*, 153:22. Id/No 107058.
- [10] Kallenberg, O. (2002). *Foundations of modern probability*. Probab. Appl. New York, NY: Springer, 2nd ed. edition.
- [11] Klenke, A. (2020). *Probability theory. A comprehensive course*. Universitext. Cham: Springer, 3rd revised and expanded edition edition.
- [12] Koenker, R. (2005). *Quantile regression.*, volume 38 of *Econom. Soc. Monogr.* Cambridge: Cambridge University Press.
- [13] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

- [14] Lieb, E. H. and Loss, M. (2001). *Analysis.*, volume 14 of *Grad. Stud. Math.* Providence, RI: American Mathematical Society (AMS), 2nd ed. edition.
- [15] Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: theory of majorization and its applications.* Springer Ser. Stat. New York, NY: Springer, 2nd edition edition.
- [16] Mikusiński, P. and Taylor, M. (2010). Some approximations of n-copulas. *Metrika*, 72:385–414.
- [17] Mroz, T., Fuchs, S., and Trutschnig, W. (2021). How simplifying and flexible is the simplifying assumption in pair-copula constructions – analytic answers in dimension three and a glimpse beyond. *Electron. J. Stat.*, 15(1):1951–1992.
- [18] Nagler, T. (2025). Simplified vine copula models: State of science and affairs. *Risk Sciences*, 1:100022.
- [19] Nagler, T. and Czado, C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *J. Multivariate Anal.*, 151:69–89.
- [20] Nelsen, R. B. (2006). *An introduction to copulas.* Springer Ser. Stat. New York, NY: Springer, 2nd ed. edition.
- [21] 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.
- [22] Vaart, A. W. v. d. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- [23] Walters, P. (1982). *An introduction to ergodic theory*, volume 79 of *Grad. Texts Math.* Springer, Cham.