On distributions with fixed marginals maximizing the joint or the prior default probability, estimation, and related results

Thomas Mroz[a], Juan Fernández Sánchez[b], Sebastian Fuchs[a], Wolfgang Trutschnig[a,*]

[a] *Department of Mathematics, University of Salzburg, Hellbrunner Strasse 34, 5020 Salzburg, Austria, Tel.: +43 662 8044-5312, Fax: +43 662 8044-137*

[b] *Grupo de Teoría de Cópulas y aplicaciones, Universidad de Almería, La Cañada de San Urbano, Almería, Spain*

Abstract

Motivated by (random) lifetimes of electronic components or financial institutions we study the problem of maximizing the probability that (i) a random variable X is not smaller than another random object Y and (ii) that X and Y coincide within the class of all random variables X, Y with given univariate continuous distribution functions F and G, respectively. We show that the maximization problems correspond to finding copulas maximizing the mass of the endograph $\Gamma^{\leq}(T) = \{(x,y) \in [0,1]^2 : y \leq T(x)\}$ and the graph $\Gamma(T) = \{(x,T(x)) : x \in [0,1]\}$ of $T = G \circ F^-$, respectively. After providing simple, copula-based proofs for the existence of copulas attaining the two maxima $\overline{m}_T$ and $\overline{w}_T$ we generalize the obtained results to the case of general (not necessarily monotonic) transformations $T : [0,1] \to [0,1]$ and derive simple and easily calculable formulas for $\overline{m}_T$ and $\overline{w}_T$ involving the distribution function F_T of T (interpreted as random variable on $[0,1]$). The latter are then used to characterize all non-decreasing transformations $T : [0,1] \to [0,1]$ for which $\overline{m}_T$ and $\overline{w}_T$ coincide. A strongly consistent estimator for $\overline{m}_T$ is derived and proven to be asymptotically normal under very mild regularity conditions. Several examples and graphics illustrate the main results and falsify some seemingly natural conjectures, an application of some of the obtained results to the seemingly unrelated topic of relative effects indicates the importance of the tackled questions.

Keywords: Copula, Dependence, Estimator, Graph, Endograph, Markov Kernel

2010 MSC: 60E05, 28A50, 91G70

1. Introduction

Suppose that F and G are (continuous) distribution functions of two random variables X and Y modeling, e.g., (i) the default times of financial institutions (see, e.g., [3, 22]) or (ii) the lifetimes of electronic components (see, e.g., [21]). Especially in the context of (i) the

*Corresponding author

Email addresses: thomasmroz@a1.net (Thomas Mroz), juanfernandez@ual.es (Juan Fernández Sánchez), sebastian.fuchs@plus.ac.at (Sebastian Fuchs), wolfgang@trutschnig.net (Wolfgang Trutschnig)

marginal distributions might be known or at least be estimated in standard ways, whereas the joint distribution is often unknown and harder to estimate. In such situations (particularly in the context of so-called credit default swaps) is seems natural to consider the worst-case scenario and study bivariate distribution functions H in the Fréchet class $\mathcal{H}_{F,G}$ of F, G (the family of all bivariate distribution functions having marginals F and G) with the following property: In case (X, Y) has distribution function H the joint or prior default probability (i.e. the probability of the events $\{X = Y\}$ and $\{X \geq Y\}$, respectively) is maximal within $\mathcal{H}_{F,G}$.

Translating to the class of copulas (see [23] and Section 2), maximizing the afore-mentioned probabilities means calculating

$$\overline{w}_T := \sup_{A \in \mathcal{C}} \mu_A(\Gamma(T)), \quad \overline{m}_T := \sup_{A \in \mathcal{C}} \mu_A(\Gamma^{\leq}(T)) \tag{1}$$

where $T : [0,1] \to [0,1]$ is defined by $T = G \circ F^-$, F^- denotes the quasi-inverse of F, $\Gamma(T) = \{(x, T(x)) : x \in [0,1]\}$ the graph of T, $\Gamma^{\leq}(T) = \{(x,y) \in [0,1]^2 : y \leq T(x)\}$ the so-called endograph of T, $\mathcal{C}$ the family of all two-dimensional copulas and μ_A the doubly stochastic measure corresponding to the copula $A \in \mathcal{C}$.

It has been brought to our attention that formulas for the suprema in eq. (1) also follow from deep and much heavier machinery going back to Rüschendorf in [28]. In the current paper we provide (a) independent alternative simple, copula-based proofs and show the existence of copulas $B \in \mathcal{C}$ attaining the right-hand suprema in (1) (including the fact that it is possible to choose B completely dependent) and the existence of copulas attaining the left-hand suprema in (1). Complementing these results, (b) we calculate $\overline{w}_T$ and $\overline{m}_T$ also for general measurable, not necessarily monotonic transformations $T : [0,1] \to [0,1]$, (c) characterize for which non-decreasing T we even have $\overline{w}_T = \overline{m}_T$ and, (d) derive a strongly consistent estimator for $\overline{m}_T$ and show that the latter is asymptotically normal under mild regularity conditions.

The rest of the paper is organized as follows: Section 2 gathers some preliminaries and notations, and proves the afore-mentioned translation of the problem of maximizing the joint or prior default probability to the copula setting. The main results concerning the calculation of the maximum probabilities and various related questions are gathered in Sections 3 and 4, whereas in Section 5 we characterize the case $\overline{w}_T = \overline{m}_T$ for non-decreasing T. Section 6 introduces an estimator for $\overline{m}_T$, shows consistency and studies its asymptotic distribution. Finally, in Section 7 a real data example is presented that demonstrates the potential of the obtained results for estimating the relative effect in the presence of dependent data and small sample sizes. Several examples and graphics illustrate our findings and the chosen approach.

2. Notation and Preliminaries

For every d-dimensional random vector $\mathbf{X}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ we will write $\mathbf{X} \sim F$ if $\mathbf{X}$ has distribution function (d.f., for short) F and let $\mu_F = \mathbb{P}^{\mathbf{X}}$ denote the corresponding distribution on the Borel σ-field $\mathcal{B}(\mathbb{R}^d)$ of $\mathbb{R}^d$. For every univariate distribution function F we will let F^- denote the quasi-inverse of F, i.e. $F^-(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$. Note that for every $q \in (0,1)$ we have $F^-(q) \leq x$ if and only if $q \leq F(x)$, that for $X \sim F$ and F continuous we have $F \circ X \sim \mathcal{U}(0,1)$, i.e., $F \circ X$ is uniformly distributed on $[0,1]$, and that the random variable $F^- \circ F \circ X$ coincides with X with probability one. For further properties of F^- we refer, for instance, to [12]. Given univariate distribution functions F and G, we will let $\mathcal{H}_{F,G}$ denote the Fréchet class of F and G, i.e. the family of all two-dimensional distribution functions having F and G as marginals; $\mathcal{P}_{F,G}$ will denote the corresponding class of probability measures on $\mathcal{B}(\mathbb{R}^2)$. $\mathcal{B}([0,1])$ and $\mathcal{B}([0,1]^2)$ denote the Borel σ-fields on $[0,1]$ and $[0,1]^2$, λ and λ_2 the Lebesgue measure on $\mathcal{B}([0,1])$ and $\mathcal{B}([0,1]^2)$ respectively. For every measurable transformation $T : [0,1] \to [0,1]$ the push-forward of λ via T will be denoted by λ^T, i.e., $\lambda^T(E) = \lambda(T^{-1}(E))$ for every $E \in \mathcal{B}([0,1])$.

As already mentioned before, $\mathcal{C}$ will denote the family of all two-dimensional *copulas*. For background on copulas we refer to [8, 26]. M and W will denote the upper and lower Fréchet-Hoeffding bounds, Π the product copula. d_∞ will denote the uniform distance on $\mathcal{C}$; it is well known that $(\mathcal{C}, d_\infty)$ is a compact metric space and that d_∞ is a metrization of weak convergence in $\mathcal{C}$. For every $A \in \mathcal{C}$ μ_A will denote the corresponding *doubly stochastic measure* defined via $\mu_A([0,x] \times [0,y]) = A(x,y)$ for all $x, y \in [0,1]$ (and extended in the standard way to $\mathcal{B}([0,1]^2)$), $\mathcal{P}_\mathcal{C}$ the class of all these doubly stochastic measures.

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

$$K(X(\omega), B) = \mathbb{E}(\mathbf{1}_B \circ Y | X)(\omega) \tag{2}$$

holds $\mathbb{P}$-a.s. It is well known that for each pair (X,Y) of real-valued random variables a regular conditional distribution $K(\cdot,\cdot)$ of Y given X exists, that $K(\cdot,\cdot)$ is unique $\mathbb{P}^X$-a.s. (i.e. unique for $\mathbb{P}^X$-almost every $x \in \mathbb{R}$) and that $K(\cdot,\cdot)$ only depends on the distribution $\mathbb{P}^{(X,Y)}$. Hence, given $(X,Y) \sim H$, we will denote (a version of) the regular conditional distribution of Y given X by $K_H(\cdot,\cdot)$ and refer to $K_H(\cdot,\cdot)$ simply as *Markov kernel of* H or *Markov kernel of* (X,Y). Note that for every two-dimensional distribution function H, its Markov kernel $K_H(\cdot,\cdot)$, and every Borel set $G \in \mathcal{B}(\mathbb{R}^2)$ the following *disintegration* formula holds ($G_x = \{y \in \mathbb{R} : (x,y) \in G\}$ denoting the x-section of G for every $x \in \mathbb{R}$)

$$\int_{\mathbb{R}} K_H(x, G_x)\, d\lambda(x) = \mu_H(G). \tag{3}$$

For $A \in \mathcal{C}$ we will directly consider the corresponding Markov kernel $K_A(\cdot,\cdot)$ to be defined on $[0,1] \times \mathcal{B}([0,1])$. Considering that in this case eq. (3) implies that

$$\int_{[0,1]} K_A(x, F)\, d\lambda(x) = \lambda(F) \tag{4}$$

holds for every $F \in \mathcal{B}([0,1])$, and that, additionally, every Markov kernel $K : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ fulfilling eq. (4) obviously induces a unique element $\mu \in \mathcal{P}_\mathcal{C}$, it follows that there is a one-to-one correspondence between $\mathcal{C}$ and the family of all Markov kernels $K : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ fulfilling eq. (4). Notice that for $A \in \mathcal{C}$ eq. (4) also implies that $K_A(x, \{0,1\}) = 0$ holds for λ-almost every $x \in [0,1]$, so it is always possible to choose a (version of the) kernel fulfilling $K_A(x, \{0,1\}) = 0$ for every $x \in [0,1]$. For more details and properties of conditional expectation, regular conditional distributions, and disintegration see [17] and [18], various results underlining the usefulness of the Markov kernel perspective can be found in [8] and the references therein.

In the sequel $\mathcal{T}$ will denote the class of all λ-preserving transformations $h : [0,1] \to [0,1]$, i.e., the class of all h fulfilling $\lambda^h = \lambda$, $\mathcal{T}_b$ the subset of all bijective $h \in \mathcal{T}$, and $\mathcal{T}_l$ the subset of all piecewise linear, bijective $h \in \mathcal{T}$. A copula $A \in \mathcal{C}$ will be called *completely dependent* if and only if there exists $h \in \mathcal{T}$ such that $K(x, E) = \mathbf{1}_E(h(x))$ is a regular conditional distribution of A (see [19, 31] for equivalent definitions and main properties). For every $h \in \mathcal{T}$ the induced completely dependent copula will be denoted by A_h throughout the rest of the paper, $\mathcal{C}_d$ will denote the family of all completely dependent copulas.

Following [8, 32], for every $h \in \mathcal{T}$ and every copula $A \in \mathcal{C}$ we will let $\mathcal{S}_h(A) \in \mathcal{C}$ denote the (generalized) *h-shuffle of A*, defined implicitly via the corresponding doubly stochastic measures by

$$\mu_{\mathcal{S}_h(A)}(E \times F) = \mu_A(h^{-1}(E) \times F) \tag{5}$$

for all $E, F \in \mathcal{B}([0,1])$. Notice that $\mathcal{S}_h(A)$ is a shuffle in the sense of [6] if $h \in \mathcal{T}_b$, and that for $A = M$ it is a shuffle in the sense of [24] (to which we will refer as classical shuffle in the sequel) if $h \in \mathcal{T}_l$.

We conclude this section with the afore-mentioned translation of the maximization problems to the copula setting and start with the following lemma which is straightforward to prove via disintegration and a Dynkin system argument.

Lemma 1. *Suppose that F, G are continuous distribution functions, that (X, Y) has d.f. $H \in \mathcal{H}_{F,G}$ and copula A, and let $K_A(\cdot, \cdot)$ denote a Markov kernel of A fulfilling $K_A(x, \{0,1\}) = 0$ for all $x \in [0,1]$. Then setting*

$$K\big(x, (-\infty, y]\big) := K_A\big(F(x), [0, G(y)]\big) \tag{6}$$

for all $x, y \in \mathbb{R}$ defines a Markov kernel $K(\cdot, \cdot)$ of $(X, Y) \sim H$.

Suppose now that $S : \mathbb{R} \to \mathbb{R}$ is an arbitrary Borel-measurable mapping. In the sequel we will let $\Gamma(S)$ and $\Gamma^{\leq}(S)$ denote the graph and the endograph of S respectively, i.e.

$$\Gamma(S) = \{(x, S(x)) : x \in \mathbb{R}\}, \quad \Gamma^{\leq}(S) = \{(x, y) \in \mathbb{R}^2 : y \leq S(x)\}. \tag{7}$$

Lemma 1 allows to express $\mathbb{P}(Y \leq X)$ as well as $\mathbb{P}(Y = X)$ in terms of F, G and the underlying copula A. In order to prove a more general result and to simplify notation, given (continuous) F, G and (measurable) S we will write

$$T := G \circ S \circ F^{-} \tag{8}$$

in the sequel. In general, T is only well-defined on $(0,1)$ - we will however, directly consider it as function on $[0,1]$ by setting $T(0) := 0$ and $T(1) := T(1-) = \lim_{x \to 1-} T(x)$.

Theorem 2. *Suppose that X, Y are random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ with joint distribution function H, continuous marginals F and G and copula A. Furthermore let $S : \mathbb{R} \to \mathbb{R}$ be an arbitrary Borel-measurable mapping and define T according to eq. (8). Then the following identities hold for $T := G \circ S \circ F^{-}$:*

$$\mathbb{P}^{(X,Y)}\big(\Gamma(S)\big) = \mu_A(\Gamma(T)), \quad \mathbb{P}^{(X,Y)}\big(\Gamma^{\leq}(S)\big) = \mu_A(\Gamma^{\leq}(T)) \tag{9}$$

Proof. Using the fact that $\mathbb{P}(F^- \circ F \circ X = X) = 1$, change of coordinates, disintegration and Lemma 1 the second identity can be proved as follows:

$$\begin{aligned}
\mathbb{P}^{(X,Y)}\big(\Gamma^{\leq}(S)\big) &= \int_\Omega K_H\big(X(\omega), (-\infty, S \circ X(\omega)]\big)\, d\mathbb{P}(\omega) \\
&= \int_\Omega K_A\big(F \circ X(\omega), [0, G \circ S \circ F^- \circ F \circ X(\omega)]\big)\, d\mathbb{P}(\omega) \\
&= \int_{[0,1]} K_A\big(z, [0, G \circ S \circ F^-(z)]\big)\, d\mathbb{P}^{F \circ X}(z) \\
&= \int_{[0,1]} K_A\big(z, [0, T(z)]\big)\, d\lambda(z) = \mu_A(\Gamma^{\leq}(T)).
\end{aligned}$$

Working with $K\big(X(\omega), \{S \circ X(\omega)\}\big)$ instead of $K\big(X(\omega), (-\infty, S \circ X(\omega)]\big)$ the first identity $\mathbb{P}^{(X,Y)}\big(\Gamma(S)\big) = \mu_A(\Gamma(T))$ follows in the same manner. □

3. Maximizing the mass of the endograph and the prior default probability

Suppose that $X \sim F$ and $Y \sim G$ model default times and that F, G are continuous. Considering $S = id_{\mathbb{R}}$ then calculating $\sup_{\mu \in \mathcal{P}_{F,G}} \mu(\Gamma^{\leq}(S))$ obviously corresponds to finding (joint) distributions of (X, Y) maximizing the probability of a prior or joint default. To simplify notation in the sequel we will simply refer to the event $\{Y \leq X\}$ as 'prior default' (of Y) although $\{Y \leq X\}$ corresponds to the prior and joint default. Notice that, setting $\psi(x, y) = x + y$ and considering the pair $(-X, Y)$ the afore-mentioned maximization problem can be considered a special case of the more general situation studied in [10, 11]. Theorem 2 implies

$$\overline{m}_{F,G} := \sup_{\mu \in \mathcal{P}(F,G)} \mu(\Gamma^{\leq}(id_{\mathbb{R}})) = \sup_{A \in \mathcal{C}} \mu_A(\Gamma^{\leq}(T)) =: \overline{m}_T \tag{10}$$

as well as

$$\underline{m}_{F,G} := \inf_{\mu \in \mathcal{P}(F,G)} \mu(\Gamma^{\leq}(id_{\mathbb{R}})) = \inf_{A \in \mathcal{C}} \mu_A(\Gamma^{\leq}(T)) =: \underline{m}_T \tag{11}$$

whereby $T = G \circ S \circ F^- = G \circ F^-$. Since $G \circ F^-$ is non-decreasing it is possible to derive a simple formula for $\underline{m}_T$ and even construct a dependence structure for which $\mathbb{P}(Y \leq X)$ coincides with $\overline{m}_T$. The following result holds:

Theorem 3. *Suppose that $T : [0,1] \to [0,1]$ is non-decreasing. Then we have*

$$\overline{m}_T = \sup_{A \in \mathcal{C}} \mu_A\big(\Gamma^{\leq}(T)\big) = 1 + \inf_{x \in [0,1]} (T(x) - x). \tag{12}$$

Moreover, defining $R \in \mathcal{T}$ by $R(x) = x + \overline{m}_T \,(mod\, 1)$, we have $\mu_{A_R}(\Gamma^{\leq}(T)) = \overline{m}_T$.

Proof. Considering $\Gamma^{\leq}(T) \subseteq [0, x] \times [0, T(x)] \cup [x, 1] \times [0, 1]$ it follows that $\mu_A(\Gamma^{\leq}(T)) \leq T(x) + 1 - x$ holds for every $x \in [0,1]$ and every $A \in \mathcal{C}$, which implies that the left-hand side of (12) is smaller than or equal to the right-hand side.
To prove the reverse inequality set $z = \inf_{x \in [0,1]} \big(T(x) + 1 - x\big)$. For $z = 1$ we have $T(x) \geq x$ for every x, so taking into account $\mu_M(\Gamma^{\leq}(T)) = 1$ we are done, and it suffices to consider $z < 1$. Compactness of $[0,1]$ implies the existence of a sequence $(x_n)_{n \in \mathbb{N}}$ and a point $x^\star \in [0,1]$

such that $\lim_{n\to\infty} x_n = x^\star$ and $\lim_{n\to\infty}(T(x_n)+1-x_n) = z$. Using $z<1$ we get $x^\star > 0$ and, using monotonicity of T it follows that $T(x^\star-)+1-x^\star = z$. Letting $R:[0,1]\to[0,1]$ denote the rotation defined by $R(x) = x + z\,(mod\,1)$, obviously $R \in \mathcal{T}$ holds. Considering that for every $x \in [x^\star - T(x^\star-), 1]$ we have (see Figure 1)

$$\begin{aligned} R(x) &= T(x^\star-) - x^\star + x = T(x^\star-) + 1 - x^\star - 1 + x \\ &\leq T(x) + 1 - x - 1 + x = T(x) \end{aligned}$$

it follows immediately that

$$\mu_{A_R}(\Gamma^{\leq}(T)) \;\geq\; 1 - (x^\star - T(x^\star-)) = z = \inf_{x\in[0,1]} \big(T(x)+1-x\big),$$

which completes the proof. □

Remark 4. Considering that continuity of T plays no role in Theorem 3, that T has (as non-decreasing function) at most countably many discontinuities, and that $\mu_A(E\times[0,1]) = 0$ for every countable set E and $A \in \mathcal{C}$ we may, w.l.o.g., assume that T is left continuous, in which case the infimum in eq. (12) is a minimum.

Corollary 5. *Suppose that X, Y are random variables with continuous distribution functions F and G respectively, set $T = G\circ F^-$ and $z := 1+\inf_{x\in[0,1]}(T(x)-x)$, define $R:[0,1]\to[0,1]$ by $R(x) = z + x\,(mod\,1)$, and let A_R denote the completely dependent copula induced by R. Then for $(X,Y)\sim H \in \mathcal{H}(F,G)$ with $H(x,y) = A_R(F(x),G(y))$ we have $\mathbb{P}(Y\leq X) = \overline{m}_{F,G}$.*

Example 6. Suppose that the default times X and Y are exponentially distributed with parameters θ_1 and θ_2, respectively. It is straightforward to verify that in this case $T = G\circ F^-$ is given by $T_\theta(x) = 1-(1-x)^\theta$, where $\theta = \frac{\theta_2}{\theta_1}$. For the case of $\theta \geq 1$ we have $T_\theta(x) \geq x$ for every $x\in[0,1]$, so $\sup_{A\in\mathcal{C}} \mu_A(\Gamma^{\leq}(T_\theta)) = 1$. Remarkably, for the case of $\theta < 1$ the maximal mass of the endograph of T_θ and the maximal mass of the graph of T_θ coincide. In fact, applying Theorem 3, on the one hand we get

$$\sup_{A\in\mathcal{C}} \mu_A\big(\Gamma^{\leq}(T_\theta)\big) = 1 + \theta^{\frac{1}{1-\theta}} - \theta^{\frac{\theta}{1-\theta}}.$$

And on the other hand, according to Theorem 3 and Theorem 4 in [4] (also see [23, 30]) we have

$$\sup_{A\in\mathcal{C}} \mu_A(\Gamma(T_\theta)) = \int_{[0,1]} \Big(\mathbf{1}_{[0,1]}(f\circ T_\theta) + \frac{1}{f\circ T_\theta}\mathbf{1}_{(1,\infty)}(f\circ T_\theta)\Big)\, d\lambda \tag{13}$$

where f denotes the density of λ^{T_θ}. Since the latter is given by $f(x) = \frac{1}{\theta}(1-x)^{\frac{1-\theta}{\theta}}$ we get $f\circ T_\theta(x) = \frac{1}{\theta}(1-x)^{1-\theta}$ and eq. (13) calculates to

$$\begin{aligned} \sup_{A\in\mathcal{C}} \mu_A(\Gamma(T_\theta)) &= \int_{\left[0,1-\theta^{\frac{1}{1-\theta}}\right]} \frac{1}{\frac{1}{\theta}(1-x)^{1-\theta}} d\lambda(x) + 1 - \big(1-\theta^{\frac{1}{1-\theta}}\big) = 1 - \theta^{\frac{\theta}{1-\theta}} + \theta^{\frac{1}{1-\theta}} \\ &= \sup_{A\in\mathcal{C}} \mu_A\big(\Gamma^{\leq}(T_\theta)\big). \end{aligned}$$

For the special case of $\theta = \frac{1}{2}$, which is depicted in Figure 1, we get

$$\sup_{A\in\mathcal{C}} \mu_A(\Gamma(T)) = \sup_{A\in\mathcal{C}} \mu_A\big(\Gamma^{\leq}(T)\big) = \frac{3}{4}.$$

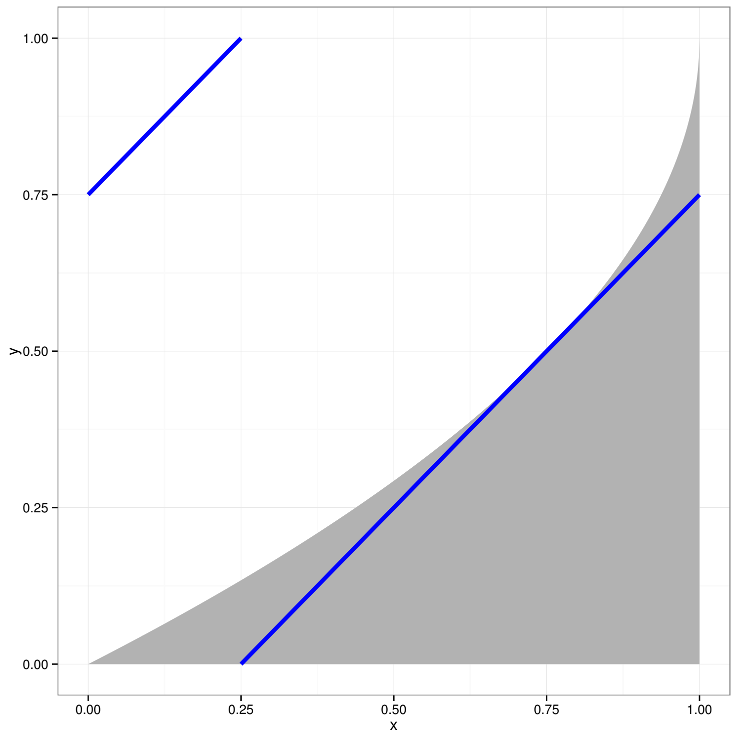

Figure 1: The endograph $\Gamma^{\leq}(T)$ of the transformation $T(x) = 1-(1-x)^{\frac{1}{2}}$ (shaded region) and the support of the mutually completely dependent copula A_R constructed in the proof of Theorem 3 assigning maximum mass to $\Gamma^{\leq}(T)$ (blue).

Example 7. Based on Example 6 it might seem natural to conjecture that the equality $\sup_{A\in\mathcal{C}} \mu_A(\Gamma(T)) = \sup_{A\in\mathcal{C}} \mu_A\big(\Gamma^{\leq}(T)\big)$ holds for a much bigger class of non-decreasing transformations T fulfilling $T(x) \leq x$ for every $x \in [0,1]$. Since counterexamples are easily constructed for the case where T is singular ($\lambda^T(E) > 0$ for some $E \in \mathcal{B}([0,1])$ with $\lambda(E) = 0$) and the case where T has discontinuities, the conjecture reduces to strictly increasing, continuous transformations T. For every $n \in \mathbb{N}$ the transformation $T_n : [0,1] \to [0,1]$, defined by

$$T_n(x) = \begin{cases} \frac{x}{2} & \text{if } x \in [0, \frac{1}{2}] \\ \frac{x}{2} + \frac{x}{2}\sqrt[n]{4x-2} & \text{if } x \in (\frac{1}{2}, \frac{3}{4}) \\ -1+2x & \text{if } x \in [\frac{3}{4}, 1] \end{cases}$$

is easily verified to be homeomorphism with $T_n(x) \leq x$ for every $x \in [0,1]$ (see Figure 2 for the case $n = 10$). Applying Theorem 3 we get $\sup_{A\in\mathcal{C}} \mu_A(\Gamma^{\leq}(T)) = \frac{3}{4}$, however, either by graphical arguments or by using Theorem 3 and Theorem 4 in [4] it is straightforward to verify that $\lim_{n\to\infty} \sup_{A\in\mathcal{C}} \mu_A(\Gamma(T_n)) = \frac{1}{2} < \frac{3}{4}$, so the conjecture is wrong.

Although monotonicity is crucial in the proof of Theorem 3 it is even possible to calculate

$$\overline{m} := \sup_{\mu\in\mathcal{P}(F,G)} \mu(\Gamma^{\leq}(S)) = \sup_{A\in\mathcal{C}} \mu_A(\Gamma^{\leq}(T))$$

for the case of arbitrary measurable (not necessarily monotonic) transformations $S : \mathbb{R} \to \mathbb{R}$ (as before $T := G \circ S \circ F^-$). Letting $T : [0,1] \to [0,1]$ denote an arbitrary measurable

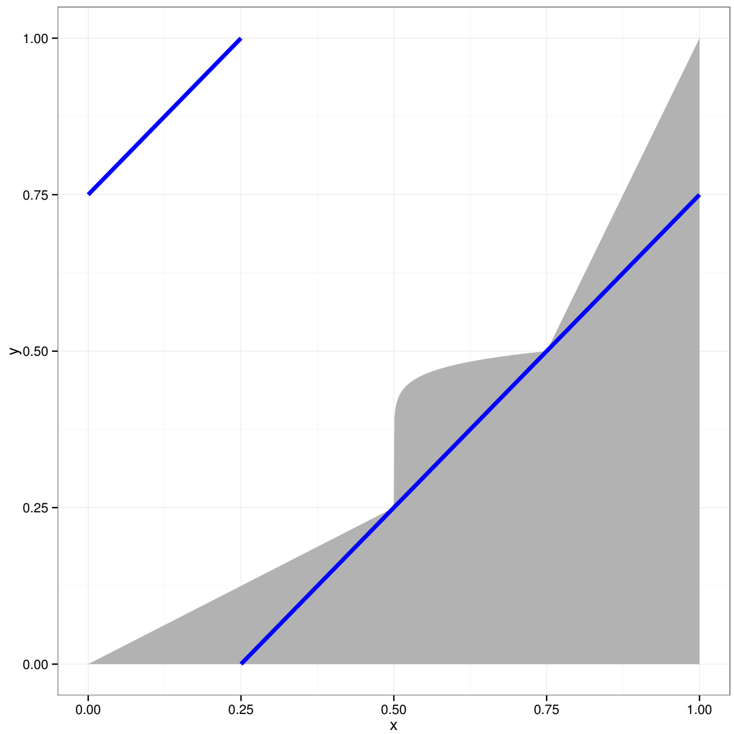

Figure 2: The endograph $\Gamma^{\leq}(T_{10})$ of the transformation T_{10} from Example 7 (shaded region) and the support of the mutually completely dependent copula A_R constructed in the proof of Theorem 3 assigning maximum mass to $\Gamma^{\leq}(T_{10})$ (blue).

transformation, we will now directly concentrate on the quantity

$$\overline{m}_T := \sup_{A \in \mathcal{C}} \mu_A(\Gamma^{\leq}(T)) \tag{14}$$

and prove a simple formula for $\overline{m}_T$ only involving the d.f. $F_T : [0,1] \to [0,1]$ of T, defined by

$$F_T(x) = \lambda^T([0,x]) = \lambda(T^{-1}([0,x])). \tag{15}$$

We start with two simple lemmata that will be used in the proof of the main results.

Lemma 8. *Suppose that $T : [0,1] \to [0,1]$ is measurable. Then we have*

$$\overline{m}_T \leq 1 + \inf_{y \in [0,1]} \big(y - F_T(y)\big) = 1 + \min_{y \in [0,1]} \big(y - F_T(y)\big) \tag{16}$$

If T is non-decreasing then we have equality in (16).

Proof. Considering $\Gamma^{\leq}(T) \subseteq [0,1] \times [0,y] \cup T^{-1}((y,1]) \times [0,1]$ and using $\lambda^T((y,1]) = 1 - F_T(y)$ we get

$$\mu_A(\Gamma^{\leq}(T)) \leq y + 1 - F_T(y)$$

for every $y \in [0,1]$ and every $A \in \mathcal{C}$, from which the first inequality follows immediately. Proving the existence of $y^\star \in [0,1]$ fulfilling $I := \inf_{y \in [0,1]}(y - F_T(y)) = y^\star - F_T(y^\star)$ can be done

as follows: For every $n \in \mathbb{N}$ we can find $y_n \in [0,1]$ with $y_n - F_T(y_n) < I + \frac{1}{2^n}$. Compactness of $[0,1]$ implies the existence of a subsequence $(y_{n_j})_{j\in\mathbb{N}}$ and some $y^\star \in [0,1]$ with $\lim_{j\to\infty} y_{n_j} = y^\star$. If $y^\star = 1$ we are done since $I = \lim_{j\to\infty}(y_{n_j} - F_T(y_{n_j})) = y^\star - \lim_{j\to\infty} F_T(y_{n_j}) \geq y^\star - 1 = y^\star - F_T(y^\star)$. Suppose therefore that $y^\star < 1$ and let $\delta \in (0, 1-y^\star]$ be arbitrary. Then there exists an index $j_0 \in \mathbb{N}$ such that $y_{n_j} < y^\star + \delta$, hence $y_{n_j} - F_T(y_{n_j}) \geq y_{n_j} - F_T(y^\star + \delta)$, holds for all $j \geq j_0$. Considering $j \to \infty$ yields $I \geq y^\star - F_T(y^\star + \delta)$, hence, using right-continuity of F_T we get $I \geq y^\star - F_T(y^\star)$.
Finally, suppose that T is non-decreasing. We want to show that

$$\inf_{y\in[0,1]}(y - F_T(y)) = \inf_{x\in[0,1]}(T(x) - x) \tag{17}$$

It follows directly from the construction that $F_T \circ T(x) \geq x$ holds for every $x \in [0,1]$ implying

$$\inf_{y\in[0,1]}(y - F_T(y)) \leq T(x) - F_T(T(x)) \leq T(x) - x$$

for every $x \in [0,1]$ and hence

$$\overline{m}_T \leq 1 + \inf_{y\in[0,1]}\big(y - F_T(y)\big) \leq 1 + \inf_{x\in[0,1]}(T(x) - x) = \overline{m}_T$$

which completes the proof. □

Lemma 9. *Suppose that $T, T' : [0,1] \to [0,1]$ are measurable transformations. Then the following two assertions hold:*

1. *For $D := \{x \in [0,1] : T(x) \neq T'(x)\}$ we have $|\overline{m}_{T'} - \overline{m}_T| \leq \lambda(D)$.*
2. *If $\Delta \in [0,1)$ and $T' \geq T - \Delta$, then $\overline{m}_{T'} \geq \overline{m}_T - \Delta$ holds.*

Proof. To prove the first assertion set $L := T\,\mathbf{1}_{D^c}$ and $U := T\,\mathbf{1}_{D^c} + \mathbf{1}_D$. Considering that obviously

$$\mu_A(\Gamma^{\leq}(L)) \leq \min\left\{\mu_A(\Gamma^{\leq}(T)), \mu_A(\Gamma^{\leq}(T'))\right\} \leq \max\left\{\mu_A(\Gamma^{\leq}(T)), \mu_A(\Gamma^{\leq}(T'))\right\} \leq \mu_A(\Gamma^{\leq}(U))$$

as well as $0 \leq \mu_A(\Gamma^{\leq}(U)) - \mu_A(\Gamma^{\leq}(L)) = \mu_A(D \times [0,1]) = \lambda(D)$ holds for every $A \in \mathcal{C}$, the desired inequality follows immediately.
To prove the second assertion let $R_\Delta : [0,1] \to [0,1]$ be defined by $R_\Delta(x) = x + \Delta (mod\, 1)$ and fix $A \in \mathcal{C}$. Since obviously $R_\Delta \in \mathcal{T}$, defining $\mu(E \times F) = \mu_A(E \times R_\Delta(F))$ yields a doubly stochastic measure μ which corresponds to a copula A_Δ (which, in turn, is easily seen to be the transpose of the R_Δ-shuffle $\mathcal{S}_{R_\Delta}(A)$ of A). Defining $\tilde{T} : [0,1] \to [0,1]$ by $\tilde{T}(x) = \max\{T(x) - \Delta, 0\}$, $\tilde{T} \leq T'$ follows and, using disintegration, we get

$$\begin{aligned}
\mu_{A_\Delta}(\Gamma^{\leq}(T')) &\geq \mu_{A_\Delta}(\Gamma^{\leq}(\tilde{T})) = \int_{T^{-1}([\Delta,1])} K_{A_\Delta}\big(x, [0, T(x) - \Delta]\big) d\lambda(x) \\
&= \int_{T^{-1}([\Delta,1])} K_A\big(x, [\Delta, T(x)]\big) d\lambda(x) \\
&= \int_{[0,1]} K_A\big(x, [0, T(x)]\big) d\lambda(x) - \int_{T^{-1}([0,\Delta))} K_A\big(x, [0, T(x)]\big) d\lambda(x)
\end{aligned}$$

$$- \int_{T^{-1}([\Delta,1])} K_A\big(x,[0,\Delta)\big)d\lambda(x)$$
$$\geq \mu_A(\Gamma^{\leq}(T)) - \int_{[0,1]} K_A\big(x,[0,\Delta)\big)d\lambda(x) = \mu_A(\Gamma^{\leq}(T)) - \Delta.$$

Since $A \in \mathcal{C}$ was arbitrary it follows immediately that $\overline{m}_{T'} \geq \overline{m}_T - \Delta$. □

Slightly modifying the ideas in the first Section of [29] it can be shown that for each measurable $T : [0,1] \to [0,1]$ there exists a non-decreasing function $T^\star : [0,1] \to [0,1]$ (called the non-decreasing rearrangement of T) and a λ-preserving transformation $\varphi : [0,1] \to [0,1]$ such that

$$T^\star \circ \varphi = T \tag{18}$$

holds. Based on Lemma 8 we can now prove the following main result of this section:

Theorem 10. *Suppose that $T : [0,1] \to [0,1]$ is measurable. Then we have*

$$\overline{m}_T = 1 + \min_{x\in[0,1]} (x - F_T(x)) = \overline{m}_{T^\star}. \tag{19}$$

Proof. Letting $\mathcal{U}_\varphi : \mathcal{C} \to \mathcal{C}$ denote the operator studied in [32] and implicitly defined via

$$K_{\mathcal{U}_\varphi(A)}(x,E) = K_A(\varphi(x),E),$$

and using disintegration as well as change of coordinates we get that

$$\begin{aligned} \mu_{\mathcal{U}_\varphi(A)}(\Gamma^{\leq}(T)) &= \int_{[0,1]} K_{\mathcal{U}_\varphi(A)}(x,[0,T(x)])d\lambda(x) = \int_{[0,1]} K_A\big(\varphi(x),[0,T^\star\circ\varphi(x)]\big)d\lambda(x) \\ &= \int_{[0,1]} K_A\big(z,[0,T^\star(z)]\big)d\lambda(z) = \mu_A(\Gamma^{\leq}(T^\star)) \end{aligned} \tag{20}$$

holds for every $A \in \mathcal{C}$, implying $\overline{m}_T \geq \overline{m}_{T^\star}$. Again using $T^\star \circ \varphi = T$ and the fact that φ is λ-preserving, it is straightforward to verify that T and $T^\star$ have the same d.f., i.e. $F_{T^\star} = F_T$ holds. Therefore, applying Lemma 8 yields

$$1 + \min_{x\in[0,1]} (x - F_T(x)) = 1 + \min_{x\in[0,1]} (x - F_{T^\star}(x)) = \overline{m}_{T^\star} \leq \overline{m}_T \leq 1 + \min_{x\in[0,1]} (x - F_T(x)), \tag{21}$$

from which the desired equality $\overline{m}_{T^\star} = \overline{m}_T$ follows immediately. □

According to Theorem 3 the completely dependent copula $A_R \in \mathcal{C}_d$ fulfills $\overline{m}_{T^\star} = \mu_{A_R}(\Gamma^{\leq}(T^\star))$, so eq. (20) implies $\mu_{\mathcal{U}_\varphi(A_R)}(\Gamma^{\leq}(T)) = \mu_{A_R}(\Gamma^{\leq}(T^\star)) = \overline{m}_{T^\star} = \overline{m}_T$. By definition of $\mathcal{U}_\varphi(C)$ we have

$$K_{\mathcal{U}_\varphi(A_R)}(x,F) = K_{A_R}(\varphi(x),F) = \mathbf{1}_F(R\circ\varphi(x)) = K_{A_{R\circ\varphi}}(x,F), \tag{22}$$

so $\mathcal{U}_\varphi(A_R)$ coincides with the completely dependent copula $A_{R\circ\varphi}$ and the following corollary holds:

Corollary 11. *Suppose that $T : [0,1] \to [0,1]$ is measurable. Then there exists a completely dependent copula $A_h \in \mathcal{C}_d$ such that $\mu_{A_h}(\Gamma^{\leq}(T)) = \overline{m}_T$.*

Having found a simple analytic formula for the maximal mass of $\Gamma^{\leq}(T)$ we now derive the analogous result for the minimal mass and set

$$\underline{m}_T = \inf_{A \in \mathcal{C}} \mu_A(\Gamma^{\leq}(T)). \tag{23}$$

Given the aforementioned results, the subsequent corollary does not come as a surprise:

Corollary 12. *For every measurable transformation $T : [0,1] \to [0,1]$ the following equality holds:*

$$\underline{m}_T = 1 - \overline{m}_{1-T} = \max_{x \in [0,1]} (x - F_T(x-)) = \underline{m}_{T^\star} \tag{24}$$

Proof. We first concentrate on the *strict endograph* $\Gamma^{<}(T)$, defined by

$$\Gamma^{<}(T) = \big\{(x,y) \in [0,1]^2 : y < T(x)\big\}.$$

Defining $T_n : [0,1] \to [0,1]$ by $T_n(x) = \max\{T(x) - 2^{-n}, 0\}$ for every $x \in [0,1]$ and $n \in \mathbb{N}$ we obviously have that $(\Gamma^{\leq}(T_n))_{n\in\mathbb{N}}$ is monotonically increasing and that $\Gamma^{<}(T) = \bigcup_{n=1}^{\infty} \Gamma^{\leq}(T_n)$. Lemma 9 yields $\overline{m}_{T_n} \geq \overline{m}_T - 2^{-n}$ and Corollary 11 implies the existence of a copula $A_n \in \mathcal{C}_d$ with $\mu_{A_n}(\Gamma^{\leq}(T_n)) = \overline{m}_{T_n}$. Altogether we get

$$\overline{m}_{T_n} = \mu_{A_n}(\Gamma^{\leq}(T_n)) \leq \mu_{A_n}(\Gamma^{<}(T)) \leq \sup_{A \in \mathcal{C}} \mu_A(\Gamma^{<}(T)) \leq \overline{m}_T,$$

so considering $n \to \infty$ shows that $\sup_{A\in\mathcal{C}} \mu_A(\Gamma^{<}(T)) = \overline{m}_T$. Having this, considering

$$\underline{m}_T \;=\; 1 - \sup_{A\in\mathcal{C}} \mu_A\big(\Gamma^{<}(1-T)\big) = 1 - \overline{m}_{1-T} = - \min_{x\in[0,1]} (x - F_{1-T}(x)) = \max_{x\in[0,1]} (x - F_T(x-)).$$

yields eq. (24). □

We close this section with two examples - the first one shows that $\underline{m}_T$ is not necessarily attained whereas the second one focuses on a non-monotonic transformation for which copulas attaining $\underline{m}_T$ and $\overline{m}_T$ can easily be constructed.

Example 13. For $T(x) = x$ Corollary 12 yields $\underline{m}_T = 0$. There is, however, no copula A fulfilling $\mu_A(\Gamma^{\leq}(T)) = 0$, i.e. contrary to $\overline{m}_T$, there are situations, in which $\underline{m}_T$ is not attained for any copula. Suppose, on the contrary, that $A \in \mathcal{C}$ fulfills $\mu_A(\Gamma^{\leq}(T)) = 0$. Then, defining $h \in \mathcal{T}_b$ by $h(x) = 1 - x$ and setting $B = \mathcal{U}_h(A)$, we have $\mu_B(\Gamma^{\leq}(1-T)) = 0$, so, $B(x, 1-x) = 0$ holds for every $x \in [0,1]$. The latter implies $B = W$, which is a contradiction since $\mu_W(\Gamma^{\leq}(1-T)) = 1$.

Example 14. For $T(x) = 4(x - \frac{1}{2})^2$ it is straightforward to find a non-decreasing mapping $T^\star$ and a λ-preserving transformation φ such that eq. (18) holds. In fact, defining $\varphi : [0,1] \to [0,1]$ by

$$\varphi(x) = \begin{cases} \;\;\,1 - 2x & \text{if } x \in \big[0, \frac{1}{2}\big] \\ -1 + 2x & \text{if } x \in \big(\frac{1}{2}, 1\big] \end{cases}$$

and considering $T^\star(x) = x^2$ we immediately get $T^\star \circ \varphi = T$. Using eq. (22), and setting $R(x) = x + \frac{3}{4}\,(mod\,1)$, it follows that $h = R \circ \varphi$ is λ-preserving and that $A_h \in \mathcal{C}_d$ fulfills $\mu_{A_h}(\Gamma^{\leq}(T)) = \overline{m}_T = \overline{m}_{T^\star} = \frac{3}{4}$. Considering that for A_φ we obviously have $\mu_{A_\varphi}(\Gamma^{\leq}(T)) = 1$, we get $\underline{m}_T = 0$ which coincides with $\max_{x\in[0,1]}(x - F_T(x))$. Figure 3 depicts the supports of the copulas A_h and A_φ as well as the endograph of T.

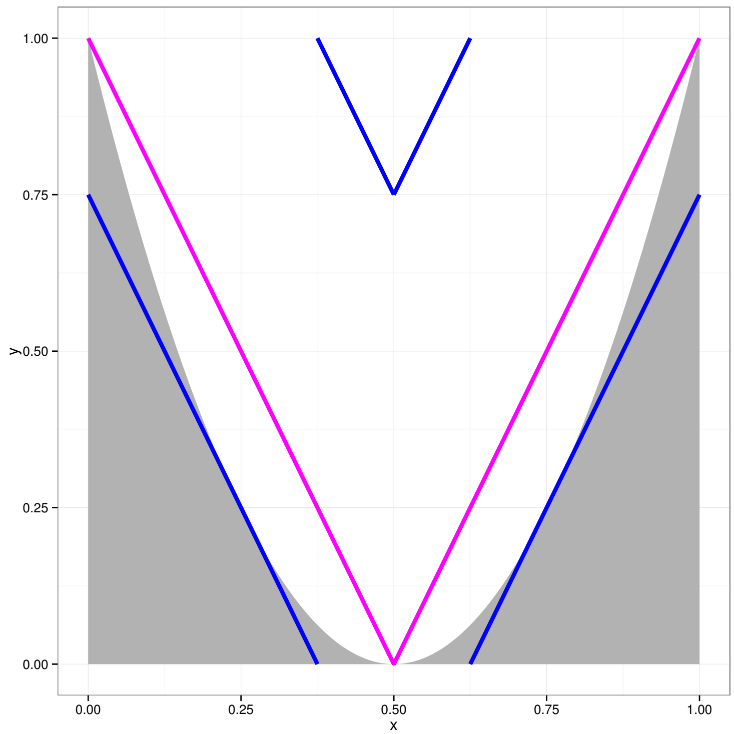

Figure 3: The endograph $\Gamma^{\leq}(T)$ of the transformation T from Example 14 (shaded region) as well as the support of the copulas A_h and A_φ maximizing/minimizing the mass of $\Gamma^{\leq}(T)$ (blue and magenta lines, respectively).

4. Maximizing the mass of the graph and the joint default probability

In what follows $T : [0,1] \to [0,1]$ will denote a general non-decreasing transformation. Since the values at the (at most countably many) discontinuity points of T are irrelevant for the maximization problem we will, however, assume that the non-decreasing transformation T is right-continuous (or left-continuous if this simplifies technical arguments). For every such T there exists a set $\Lambda_T \in \mathcal{B}([0,1])$ with $\lambda(\Lambda_T) = 1$ such that T is differentiable at every $x \in \Lambda_T$ (see, e.g., [27]). For its derivative T', in the sequel we will set $T'(x) = 0$ for every $x \in \Lambda_T^c$ and directly consider T' as integrable function on $[0,1]$ without explicit mentioning. Letting ν_T denote the measure on $\mathcal{B}([0,1])$ generated by T via $\nu_T((a,b]) = T(b) - T(a)$, it follows that T' is (a version of) the Radon-Nikodym derivative of the absolutely continuous component of ν_T w.r.t. λ (see [27, Chapter 7]). Consequently, for every interval $[a,b] \subseteq [0,1]$ we have

$$\int_{[a,b]} T' d\lambda = \int_{(a,b]} T' d\lambda \leq T(b-) - T(a) = \nu_T((a,b)) \leq \nu_T([a,b]). \tag{25}$$

Inequality (25) becomes a chain of equalities for all intervals $(a,b] \subseteq [0,1]$ if and only if T is absolutely continuous. Define a new measure ϑ on $\mathcal{B}([0,1])$ by setting

$$\vartheta(E) = \int_{T^{-1}(E)} T' d\lambda. \tag{26}$$

For a given interval $[a,b] \subseteq [0,1]$ we distinguish the following two cases: (i) If the preimage $T^{-1}([a,b])$ is of the form $[x_1,x_2]$ then using ineq. (25) it follows that

$$\vartheta([a,b]) = \int_{T^{-1}([a,b])} T' d\lambda = \int_{[x_1,x_2]} T' d\lambda \leq T(x_2) - T(x_1) \leq b - a = \lambda([a,b])$$

(ii) If $T^{-1}([a,b])$ is of the form $[x_1,x_2)$ then again by ineq. (25) we get

$$\vartheta([a,b]) = \int_{T^{-1}([a,b])} T' d\lambda = \int_{[x_1,x_2)} T' d\lambda \leq T(x_2-) - T(x_1) \leq b - a = \lambda([a,b]).$$

Having this, the following simple lemma (which will be used in the proof of the main result of this section) is straightforward to prove:

Lemma 15. *Suppose that $T : [0,1] \to [0,1]$ is right-continuous and non-decreasing and let ϑ be defined according to eq. (26). Then $\vartheta(E) \leq \lambda(E)$ holds for every $E \in \mathcal{B}([0,1])$. In particular, ϑ is absolutely continuous w.r.t. λ and the corresponding Radon-Nikodym derivative $f = \frac{d\vartheta}{d\lambda}$ fulfills $f \leq 1$ λ-a.e.*

Proof. Fix $E \in \mathcal{B}([0,1])$ and $\Delta > 0$. By construction of the Lebesgue measure λ there exists a family $(I_i)_{i\in\mathbb{N}}$ of compact intervals fulfilling $E \subseteq \bigcup_{i=1}^{\infty} I_i$ as well as $\sum_{i=1}^{\infty} \lambda(I_i) \leq \lambda(E) + \Delta$. Using $\vartheta([a,b]) \leq \lambda([a,b])$ it follows that

$$\vartheta(E) \quad \leq \quad \int_{T^{-1}(\bigcup_{i=1}^{\infty} I_i)} T' d\lambda \leq \sum_{i=1}^{\infty} \int_{T^{-1}(I_i)} T' d\lambda \leq \sum_{i=1}^{\infty} \lambda(I_i) \leq \lambda(E) + \Delta,$$

from which, considering that $\Delta > 0$ was arbitrary, we immediately get $\vartheta(E) \leq \lambda(E)$. The remaining assertions are straightforward consequences of Radon-Nikodym theorem ([27]). □

As by-product of the results in [4] we know that for the case of non-singular T (i.e. λ^T absolutely continuous w.r.t. λ) there exists a copula $A \in \mathcal{C}$ such that, firstly, $K_A(x, \{Tx\}) > 0$ for every $x \in [0,1]$ and, secondly,

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) = \mu_A(\Gamma(T))$$

holds. If T is not non-singular, there is no copula fulfilling $K_A(x, \{Tx\}) > 0$ for every $x \in [0,1]$ - nevertheless it is possible to find a copula A^T (we write A^T instead of A_T to avoid confusion with completely dependent copulas) assigning maximal mass to $\Gamma(T)$ and it is possible to derive a very simple formula for the maximal mass:

Theorem 16. *Suppose that $T : [0,1] \to [0,1]$ is non-decreasing. Then there exists a copula $A^T \in \mathcal{C}$ such that the following equality holds:*

$$\overline{w}_T = \sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) = \mu_{A^T}(\Gamma(T)) = \int_{[0,1]} \min\{T'(x), 1\} d\lambda(x) \tag{27}$$

Proof. We proceed in several steps and set $a(x) = \min\{T'(x), 1\}$ for every $x \in [0,1]$. As first step we show that for every copula A the mapping $m_A : [0,1] \to [0,1]$, defined by $m_A(x) = K_A(x, \{T(x)\})$, fulfills $m_A \leq a$ λ-a.e. Letting $L(m_A)$ denote the set of all Lebesgue points of m_A (see [27]) and setting $\Lambda := \Lambda_T \cap L(m_A) \cap (0,1)$ it follows that $\lambda(\Lambda) = 1$. For every $x \in \Lambda$ and $h > 0$ sufficiently small, using disintegration and monotonicity of T we get

$$\begin{aligned} \frac{1}{2h} \int_{[x-h,x+h]} m_A \, d\lambda &\leq \frac{1}{2h} \int_{[x-h,x+h]} K_A\big(t, [T(x-h), T(x+h)]\big) \, d\lambda(t) \\ &\leq \frac{1}{2h} \mu_A\big([0,1] \times [T(x-h), T(x+h)]\big) = \frac{T(x+h) - T(x-h)}{2h}, \end{aligned} \tag{28}$$

from which, considering $h \to 0+$ we directly get $0 \leq m_A(x) \leq T'(x)$. Since $x \in \Lambda$ was arbitrary and $m_A(x) \leq 1$ by construction, the desired inequality $m_A(x) \leq a(x)$ holds for every $x \in \Lambda$. As direct consequence we get

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) \leq \int_{[0,1]} a d\lambda$$

and the theorem is proved if we can show that there exists a copula A^T fulfilling $\mu_{A^T}(\Gamma(T)) = \int_{[0,1]} a d\lambda$. We distinguish three cases (and, as before, assume w.l.o.g. that T is right-continuous):

(i) If $\int_{[0,1]} a \, d\lambda = 1$ we get $T' \geq 1$ a.e. Since T' is the Radon-Nikodym derivative of the absolutely continuous component of the measure ν_T mentioned at the beginning of this section, considering $\nu_T([0,1]) \leq 1$ it follows that ν_T is absolutely continuous with density T' and that $T' = 1$ a.e. Hence $\nu_T = \lambda$ and $T = id$, and setting $A^T = M$ yields the desired result $\mu_{A^T}(\Gamma(T)) = 1$. **(ii)** The case $\int_{[0,1]} a \, d\lambda = 0$ is trivial since every absolutely continuous copula A fulfills $\mu_A(\Gamma(T)) = 0$. **(iii)** In the remaining case of $\int_{[0,1]} a \, d\lambda \in (0,1)$ we can proceed as follows: Define a measure μ on $\mathcal{B}([0,1]^2)$ by setting

$$\mu(E \times F) = \int_E a(x) \mathbf{1}_F(T(x)) d\lambda(x)$$

and extending in the standard way ([17, 18, 27]) to full $\mathcal{B}([0,1]^2)$. Letting $\pi_1, \pi_2 : [0,1]^2 \to [0,1]$ denote the projections onto the first and second coordinate, respectively, for every $E \in \mathcal{B}([0,1])$ we get

$$\mu^{\pi_1}(E) = \mu(E \times [0,1]) = \int_E a \, d\lambda$$

as well as

$$\mu^{\pi_2}(E) = \mu([0,1] \times E) = \int_{[0,1]} a(x) \mathbf{1}_E(T(x)) \, d\lambda(x) = \int_{T^{-1}(E)} a \, d\lambda \leq \lambda(E),$$

whereby the last inequality follows from Lemma 15. As direct consequence both μ^{π_1} and μ^{π_2} are absolutely continuous measures whose densities f_1, f_2 fulfill $f_1(x), f_2(x) \in [0,1]$ a.e. and we have $\mu^{\pi_1}([0,1]) = \mu^{\pi_2}([0,1]) = \mu([0,1]^2) = \int_{[0,1]} a\, d\lambda \in (0,1)$. Defining $F_1, F_2 : [0,1] \to [0,1]$ by

$$F_1(x) = \frac{x - \mu^{\pi_1}([0,x])}{1 - \mu^{\pi_1}([0,1])}, \quad F_2(x) = \frac{x - \mu^{\pi_2}([0,x])}{1 - \mu^{\pi_2}([0,1])}$$

yields absolutely continuous distribution functions F_1 and F_2 fulfilling $F_1(0) = F_2(0) = 0$. Finally, let $R, S : [0,1]^2 \to [0,1]$ be defined by

$$\begin{aligned} R(x_1, x_2) &= \big(1 - \mu^{\pi_1}([0,1])\big) F_1(x_1) F_2(x_2) \\ S(x_1, x_2) &= \mu\big([0,x_1] \times [0,x_2]\big) \end{aligned} \tag{29}$$

and set $A^T = R + S$. Considering $A^T(1,1) = 1$ and the fact that R and S are two-dimensional measure-generating functions by construction, it is now straightforward to show that A^T is a copula. In fact, the property $A^T(x_1, 0) = 0$ follows via

$$A^T(x_1, 0) = S(x_1, 0) = \mu([0,x_1] \times [0,0]) \le \mu([0,1] \times [0,0]) = 0$$

and the remaining boundary conditions are easily verified too. Since we obviously have $\mu_{A^T}(\Gamma(T)) = \mu(\Gamma(T)) = \int_{[0,1]} a\, d\lambda$ this completes the proof. □

Notice that in the case of $\int_{[0,1]} \min\{T'(x), 1\} d\lambda(x) \in (0,1)$ we could have also defined R by

$$R_C(x_1, x_2) = \big(1 - \mu^{\pi_1}([0,1])\big) C(F_1(x_1), F_2(x_2)),$$

whereby C is an arbitrary (not necessarily absolutely continuous) copula, worked with $A_C^T = R_C + S$ and used the fact that in this case $\mu_{A_C^T}(\Gamma(T)) \ge \mu(\Gamma(T)) = \int_{[0,1]} a\, d\lambda$ holds. As a consequence we get the following corollary:

Corollary 17. *If $T : [0,1] \to [0,1]$ is non-decreasing and $\int_{[0,1]} \min\{T'(x), 1\} d\lambda(x) \in (0,1)$, then for every copula $C \in \mathcal{C}$ there exists a copula $A_C^T \in \mathcal{C}$ such that*

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) = \mu_{A_C^T}(\Gamma(T)) = \int_{[0,1]} \min\{T'(x), 1\} d\lambda(x). \tag{30}$$

We now turn to the general problem of calculating

$$\overline{w}_T = \sup_{B \in \mathcal{C}} \mu_B(\Gamma(T))$$

for general measurable, not necessarily monotonic $T : [0,1] \to [0,1]$. Analogous to the case of $\overline{m}_T$ we first show that rearranging T non decreasingly as $T = T^* \circ \varphi$ does not change the maximum mass, i.e., $\overline{w}_T = \overline{w}_{T^*}$ holds. Doing so, we will work with the so-called $\star$-operator (see [8, Definition 5.4.6]) $\star : \mathcal{C}_2 \times \mathcal{C}_2 \to \mathcal{C}_3$, defined by

$$A \star B(x, y, z) = \int_{[0,y]} \partial_2 A(x, s) \partial_1 B(s, z) d\lambda(s) \tag{31}$$

for all $x, y, z \in [0,1]$. It is straightforward to verify (see [7, 8]) that $\star$ is well-defined, that for all $x, y, z \in [0,1]$ we have $A \star B(x, y, 1) = A(x, y)$, $A \star B(1, y, z) = B(x, y)$ as well as

$A\star B(x,1,z) = A*B(x,z)$, where $*$ denotes the star-product going back to [5]. Furthermore, considering $\partial_2 A(x,s) = \partial_1 A^t(s,x)$ it follows immediately that setting

$$K^{13|2}_{A\star B}(y, E\times G) := K_{A^t}(y,E)K_B(y,G)$$

for all $y\in[0,1]$ and $E,G\in\mathcal{B}([0,1])$ and extending in the standard way to $\mathcal{B}([0,1]^2)$ defines a Markov kernel of $A\star B$ w.r.t. the second coordinate y (see [13] and [25] for Markov kernels of multivariate copulas). Since $K^{13|2}_{A\star B}(y,\cdot)$ is the product measure of $K_{A^t}(y,\cdot)$ and $K_B(y,\cdot)$, applying Fubini's theorem we get that

$$K^{13|2}_{A\star B}(y,\Omega) = \int_{[0,1]} K_B(y,\Omega_x)K_{A^t}(y,dx) \tag{32}$$

holds for every $\Omega\in\mathcal{B}([0,1]^2)$.

Theorem 18. *Suppose that $T:[0,1]\to[0,1]$ is measurable and, as before, let T^* with $T^*\circ\varphi = T$ denote the non-decreasing rearrangement of T. Then $\overline{w}_T = \overline{w}_{T^*}$ holds.*

Proof. (i) For arbitrary $B\in\mathcal{C}$, working with $\mathcal{U}_\varphi$, using disintegration and change of coordinates we get

$$\begin{aligned}\mu_{\mathcal{U}_\varphi(B)}(\Gamma(T)) &= \int_{[0,1]} K_B\big(\varphi(x),\{T(x)\}\big)d\lambda(x) = \int_{[0,1]} K_B\big(\varphi(x),\{T^*\circ\varphi(x)\}\big)d\lambda(x)\\ &= \int_{[0,1]} K_B\big(z,\{T^*(z)\}\big)d\lambda(z) = \mu_B(\Gamma(T^*)),\end{aligned}$$

from which the inequality $\overline{w}_{T^*}\le\overline{w}_T$ follows immediately.
(ii) To prove $\overline{w}_{T^*}\ge\overline{w}_T$ we use the $\star$-operator and proceed as follows. Letting A_φ denote the completely dependent copula induced by φ, eq. (32) simplifies to

$$K^{13|2}_{A^t_\varphi\star B}(y,\Omega) = \int_{[0,1]} K_B(y,\Omega_x)K_{A_\varphi}(y,dx) = \int_{[0,1]} K_B(y,\Omega_x)d\delta_{\varphi(y)}(x) = K_B(y,\Omega_{\varphi(y)}). \tag{33}$$

Considering $\Omega=\Gamma(T^*)$ we obviously have $\Omega_{\varphi(y)} = \{T^\star\circ\varphi(y)\} = \{T(y)\}$, so it follows that

$$K^{13|2}_{A^t_\varphi\star B}(y,\Gamma(T^\star)) = K_B(y,\{T(y)\}).$$

Having this, using disintegration and the fact that $A^t_\varphi\star B(x,1,z) = A^t_\varphi * B(x,z)$ altogether we get

$$\begin{aligned}\mu_{A^t_\varphi * B}\big(\Gamma(T^\star)\big) &= \mu_{A^t_\varphi\star B}\big(\{(x,y,T^\star(x)) : x,y\in[0,1]\}\big) = \int_{[0,1]} K^{13|2}_{A^t_\varphi\star B}(y,\Gamma(T^\star))d\lambda(y)\\ &= \int_{[0,1]} K_B(y,\{T(y)\})d\lambda(y) = \mu_B(\Gamma(T)).\end{aligned}$$

Considering the fact that $B\in\mathcal{C}$ was arbitrary the desired inequality $\overline{w}_{T^\star}\ge\overline{w}_T$ follows and the theorem is proved. $\square$

Remark 19. In the proof of Theorem 18 the only properties needed were that φ is λ-preserving and that we have $T^* \circ \varphi = T$ - the fact that T^* is non-decreasing was not used. Consequently, for arbitrary measurable $S : [0,1] \to [0,1]$ and arbitrary λ-preserving $\varphi : [0,1] \to [0,1]$, setting $T := S \circ \varphi$ we have $\overline{w}_S = \overline{w}_T$.

Choosing $C = \mathcal{U}_\varphi(A^{T^*})$, where A^{T^*} denotes the copula maximizing the mass of the graph of the non-decreasing rearrangement T^* of T directly yields the following result.

Corollary 20. *For every measurable transformation $T : [0,1] \to [0,1]$ there exists a copula $C \in \mathcal{C}$ fulfilling $\overline{w}_T = \mu_C(\Gamma(T))$.*

Combining Theorem 18 and Corollary 17 shows that the identity

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) = \overline{w}_T = \overline{w}_{T^*} = \int_{[0,1]} \min\left\{\big(T^*\big)'(x), 1\right\} d\lambda(x) \tag{34}$$

holds for every measurable T. In most situations, however, the integral in eq. (34) is intractable, in particular since calculating the rearrangement T^* itself is a nontrivial endeavor. Calculating $\overline{m}_T$ for general measurable transformations T in the last section, the cumulative distribution function F_T of T plays an important role - we will show now that the same is true for $\overline{w}_T$ and derive a very simple formula only involving F_T.

Lemma 21. *Suppose that $T : [0,1] \to [0,1]$ is non-decreasing and let F_T denote the distribution function of T. Then $\overline{w}_T = \overline{w}_{F_T}$ holds.*

Proof. Let J_T denote the set of all discontinuities of T and set $I_T = \{y \in [0,1] : \lambda^T(\{y\}) > 0\}$. Then I_T and J_T are at most countably infinite and for every copula A and $y \in I_T$ we have $\mu_A(T^{-1}(\{y\}) \times \{y\})) \leq \mu_A([0,1] \times \{y\})) = 0$. Setting $N_T := J_T \cap T^{-1}(I_T)$ obviously T is injective on N_T and for every $A \in \mathcal{C}$ we have $\mu_A\big((N_T \times [0,1]) \cap \Gamma(T)\big) = 0$, implying

$$\mu_A(\Gamma(T)) = \mu_A\big(\underbrace{(N_T^c \times [0,1]) \cap \Gamma(T)}_{=:\Omega_T}\big).$$

Letting $I_{F_T}, J_{F_T}, N_{F_T}$ and Ω_{F_T} denote the corresponding sets for F_T it is straightforward to verify that for every $(x,y) \in [0,1]^2$ we have $(x,y) \in \Omega_T$ if, and only if $(y,x) \in \Omega_{F_T}$. Having this the desired result follows easily: In fact, letting A denote a copula with $\mu_A(\Gamma(T)) = \overline{w}_T$ and considering the transpose A^t we immediately get

$$\overline{w}_T = \mu_A(\Gamma(T)) = \mu_A(\Omega_T) = \mu_{A^t}(Q_{F_T}) = \mu_{A^t}(\Gamma(F_T)) \leq \overline{w}_{F_T}.$$

Since the other inequality follows in the same manner the desired equality is proved. □

Considering that T and T^* have the same distribution function and applying Lemma 21 yields a handier version of eq. (34):

Corollary 22. *For every measurable $T : [0,1] \to [0,1]$ the following equality holds:*

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma(T)) = \overline{w}_T = \overline{w}_{F_T} = \int_{[0,1]} \min\left\{F_T'(x), 1\right\} d\lambda(x) \tag{35}$$

Notice that in case $T : [0,1] \to [0,1]$ is non-decreasing and continuous and fulfills $T' = 0$ λ-almost everywhere according to eq. (17) $\overline{w}_T = 0$, i.e., no copula assigns mass to $\Gamma(T)$. This result is not surprising - considering the fact, however, that in the language of Baire categories a 'typical' monotonic function is singular (as established in [34]) we could infer that copulas assign no mass to 'typical' monotonic functions, which seems quite counterintuitive. In [4, Theorem 3] it was shown that for every non-singular $T : [0,1] \to [0,1]$ there exists a copula A such that the singular component of A is concentrated on $\Gamma(T)$ and that we have $K_A(x, \{T(x)\}) > 0$ for λ-almost every $x \in [0,1]$. Based on Theorem 16 and Theorem 18 we can give a necessary and sufficient condition for the existence of a copula A fulfilling $K_A(x, \{T(x)\}) > 0$ for every $x \in [0,1]$ in terms of the non-decreasing rearrangement T^* of T and in terms of absolute continuity of F_T.

Corollary 23. *Suppose that $T : [0,1] \to [0,1]$ is measurable, let T^* denote its non-decreasing rearrangement and F_T its distribution function. Then the following conditions are equivalent:*

- *(a) There exists a copula C fulfilling $K_C(x, \{T(x)\}) > 0$ for λ-almost every $x \in [0,1]$.*
- *(b) $(T^*)'(x) > 0$ for λ-almost every $x \in [0,1]$.*
- *(c) λ^T is absolutely continuous.*
- *(d) F_T is absolutely continuous.*

Proof. It is clear that (c) and (d) are equivalent so it suffices to prove $(b) \Rightarrow (a) \Rightarrow (c) \Rightarrow (b)$, which can be done as follows. (i) Let A be a copula assigning maximum mass to $\Gamma(T^*)$ as constructed in the proof of Theorem 16. Letting $K_A(\cdot,\cdot)$ denote a version of the Markov kernel of A fulfilling $K_A(x, \{T^*(x)\}) > 0$ for λ-almost every $x \in [0,1]$ and considering $C = \mathcal{U}_\varphi(A)$ we get

$$K_C(x, \{T(x)\}) = K_A(\varphi(x), \{T(x)\}) = K_A(\varphi(x), \{T^* \circ \varphi(x)\}) > 0, \tag{36}$$

so (b) implies (a).

The implication $(a) \Rightarrow (c)$ is a direct consequence of the fact that for every $N \in \mathcal{B}([0,1])$ with $\lambda(N) = 0$ and an arbitrary copula C fulfilling (a) we have

$$0 \;=\; \lambda(N) = \mu_C\big([0,1] \times N\big) \geq \mu_C\big(T^{-1}(N) \times N\big) \geq \int_{T^{-1}(N)} \underbrace{K_C(x, \{T(x)\})}_{>0} d\lambda(x),$$

from which $\lambda^T(N) = 0$ follows immediately.

(iii) Simplifying notation set $S := T^*$ and suppose now that $\lambda^S = \lambda^T$ is absolutely continuous. We want to show that $S'(x) > 0$ for λ-almost every $x \in [0,1]$. Since S is non-decreasing, considering that $\lambda^S(\{y\}) = 0$ and that $S^{-1}(\{y\})$ is an interval for every $y \in [0,1]$, it follows that $S^{-1}(\{y\})$ is either empty or a degenerated interval consisting of one single point, so S is necessarily strictly increasing on $[0,1]$. Additionally, for every $E \subseteq [0,1]$ we obviously have $S^{-1}(S(E)) = E$. Assume that $S(0) = 0$ (if $S(0) > 0$ holds proceed with the function $\tilde{S}$ that coincides with S on $(0,1]$ and fulfills $\tilde{S}(0) = 0$). Letting f denote the Radon-Nikodym derivative of λ^S w.r.t. λ we may w.l.o.g. assume $0 \leq f(z) < \infty$ for every $z \in [0,1]$. The function $g : [0,1] \to [0,1]$, defined by $y \mapsto \int_{[0,y]} f d\lambda$ is non-decreasing and fulfills

$$g(S(x)) = \lambda^S([S(0), S(x)]) = \lambda^S(S([0,x]) = x. \tag{37}$$

Choose $\Psi, \Lambda \in \mathcal{B}([0,1])$ with $\lambda(\Lambda) = 1 = \lambda(\Psi)$ in such a way that g is differentiable at every $y \in \Lambda$ and fulfills $g'(y) = f(y)$ and that S is differentiable at every $z \in \Psi$. For every $x \in S^{-1}(\Lambda) \cap \Psi$ applying the chain rule together with equ. (37) yields

$$1 = f(S(x))S'(x),$$

hence $S'(x) > 0$. This completes the proof since $\lambda(S^{-1}(\Lambda) \cap \Psi) = 1$. □

Remark 24. Again using the $\star$-operator allows for a direct proof of the implication $(a) \Rightarrow (b)$ of Corollary 23: Suppose that $E \in \mathcal{B}([0,1])$ is arbitrary but fixed. Applying eq. (33) to the set $\Gamma_E(T^*) := \{(x, T^*(x)) : x \in E\} = \Gamma(T^*) \cap (E \times [0,1])$ for every $B \in \mathcal{C}$ we get

$$K^{13|2}_{A^t_\varphi \star B}(y, \Gamma_E(T^*)) = \mathbf{1}_E(\varphi(y))\, K_B(y, \{T(y)\}),$$

so, using disintegration

$$\begin{aligned} \mu_{A^t_\varphi * B}\big(\Gamma_E(T^*)\big) &= \mu_{A^t_\varphi \star B}\big(\{(x, y, T^*(x)) : x \in E, y \in [0,1]\}\big) = \int_{[0,1]} K^{13|2}_{A^t_\varphi \star B}(y, \Gamma_E(T^*))d\lambda(y) \\ &= \int_{[0,1]} \mathbf{1}_E(\varphi(y))\, K_B(y, \{T(y)\})d\lambda(y) = \int_{\varphi^{-1}(E)} K_B(y, \{T(y)\})d\lambda(y) \end{aligned}$$

follows. Suppose now that $B \in \mathcal{C}$ fulfills $K_B(y, \{T(y)\}) > 0$ for every $y \in [0,1]$. Using the fact that φ is λ-preserving we get that $\mu_{A^t_\varphi * B}\big(\Gamma_E(T^*)\big) > 0$ if, and only if $\lambda(E) > 0$. Since for $E := \{x \in [0,1] : K_{A^t_\varphi * B}(x, \{T^*(x)\}) = 0\} \in \mathcal{B}([0,1])$ obviously $\mu_{A^t_\varphi * B}\big(\Gamma_E(T^*)\big) = 0$ holds, the latter implies $\lambda(E) = 0$, so we can find a version of the kernel $K_C(\cdot, \cdot)$ of the copula $C = A^t_\varphi * B$ such that $K_C(x, \{T^\star(x)\}) > 0$ holds for λ-almost every $x \in [0,1]$. Having this, proceeding analogously to (36) directly yields condition (a). ■

Remark 25. Interestingly, the maximum joint default probability for X and Y equals the value of their total variation. Recall that the total variation metric TV is defined by

$$\mathrm{TV}(\mathbb{P}^X, \mathbb{P}^Y) \;:=\; \sup_{B \in \mathcal{B}(\mathbb{R})} |\mathbb{P}^X(B) - \mathbb{P}^Y(B)|$$

and fulfills

$$\mathrm{TV}(\mathbb{P}^X, \mathbb{P}^Y) = \inf_{\mu \in \mathcal{P}(F,G)} \mu(\{(x,y) \in \mathbb{R}^2 \,:\, x \neq y\})$$

(see, e.g., [15, 20]). Thus, $\mathrm{TV}(\mathbb{P}^X, \mathbb{P}^Y) = 1 - \sup_{\mu \in \mathcal{P}(F,G)} \mu(\Gamma(id_{\mathbb{R}})) = 1 - \overline{w}_T$, i.e., maximizing the joint default probability means calculating the total variation.

5. When $\overline{m}_T$ and $\overline{w}_T$ coincide

The results in the previous two sections allows to characterize all non-decreasing functions $T : [0,1] \to [0,1]$ for which the maximum mass of the graph $\Gamma(T)$ and the maximum mass of the endograph $\Gamma^{\leq}(T)$ coincide:

Theorem 26. *Suppose that $T : [0,1] \to [0,1]$ is non-decreasing and let Λ_T denote the set of all points at which T is differentiable. Then the following two assertions are equivalent.*

(a) $\overline{m}_T = \overline{w}_T$.

(b) $T(0) = 0$ *and there exists a point* $x_0 \in [0,1]$ *such that the following conditions hold:*

(i) T *is absolutely continuous on* $[0, x_0]$,

(ii) $\Omega_0 := \{x \in [0, x_0] \cap \Lambda_T : T'(x) \leq 1\}$ *fulfills* $\lambda(\Omega_0) = x_0$,

(iii) $\Omega_1 := \{x \in [x_0, 1] \cap \Lambda_T : T'(x) \geq 1\}$ *fulfills* $\lambda(\Omega_1) = 1 - x_0$.

Proof. We may, w.l.o.g., assume that T is left-continuous.
(I) Suppose that T fulfills the second assertion. It follows immediately from condition (i) that for every $z \in [0, x_0]$ we have $T(z) = \int_{[0,z]} T' d\lambda$, hence, by condition (ii), the mapping $z \mapsto T(z) - z = \int_{[0,z]} (T' - 1) d\lambda$ is non-increasing on $[0, x_0]$ and we have $\inf_{z \in [0,x_0]} (T(z) - z) = T(x_0) - x_0$. Additionally, condition (iii) implies that $z \mapsto T(z) - z$ is non-decreasing on $[x_0, 1]$, from which, using Theorem 3 we altogether get

$$\sup_{B \in \mathcal{C}} \mu_B(\Gamma^{\leq}(T)) = \overline{m}_T = 1 + \inf_{x \in [0,1]} (T(x) - x) = 1 + T(x_0) - x_0.$$

Taking into account that (i)-(iii) also imply

$$\int_{[0,1]} \min\{T'(x), 1\} d\lambda(x) = \int_{[0,x_0]} T' d\lambda + \int_{[x_0,1]} 1 d\lambda = T(x_0) + 1 - x_0$$

the desired equality $\overline{m}_T = \overline{w}_T$ follows.
(II) On the other hand, if $\overline{m}_T = \overline{w}_T$ holds, then again by Theorem 3, left-continuity of T and Theorem 16, there exists some $x_0 \in [0,1]$ such that

$$1 + T(x_0) - x_0 = \overline{m}_T = \overline{w}_T = \int_{[0,1]} \min\{T'(x), 1\} d\lambda(x)$$

holds. Considering $\int_{[0,x_0]} \min\{T'(x), 1\} d\lambda(x) \leq T(x_0) - T(0) \leq T(x_0)$ together with the fact that $\int_{[x_0,1]} \min\{T'(x), 1\} d\lambda(x) \leq 1 - x_0$ it follows immediately that T has to fulfill $T(0) = 0$ as well as

$$\int_{[0,x_0]} \min\{T'(x), 1\} d\lambda(x) \quad = \quad T(x_0), \quad \int_{[x_0,1]} \min\{T'(x), 1\} d\lambda(x) = 1 - x_0.$$

The latter, however, implies that T is absolutely continuous on $[0, x_0]$ and that T fulfills (ii) and (iii). □

Considering that every convex function $T : [0,1] \to [0,1]$ with $T(0) = 0$ fulfills the properties listed in condition (b) of Theorem 26 we immediately get the following result:

Corollary 27. *If* $T : [0,1] \to [0,1]$ *is convex and fulfills* $T(0) = 0$ *then* $\overline{m}_T = \overline{w}_T$ *holds.*

According to Corollary 27, given a non-decreasing transformation $T : [0,1] \to [0,1]$, convexity and $T(0) = 0$ is sufficient for $\overline{m}_T = \overline{w}_T$. The two conditions are, however, far from being necessary - the following example shows that equality can also hold for non-decreasing transformations that are not even locally convex.

Example 28. Let $\Omega \in \mathcal{B}([0,1])$ denote a set with $\lambda(\Omega) = \frac{1}{2}$ such that $\lambda((a,b) \cap \Omega) > 0$ and $\lambda((a,b) \cap \Omega^c) > 0$ hold for every non-empty open interval $(a,b) \subseteq [0,1]$ (for a possible construction see [14, Lemma 3.1]). Define the function $S : [0,1] \to [0,1]$ by $S(x) = \int_{[0,x]} \mathbf{1}_\Omega(y)\, d\lambda(y)$ for every $x \in [0,1]$. Then S is strictly increasing, $S(0) = 0$, $S(1) = \frac{1}{2}$, S is absolutely continuous and $S'(x) = \mathbf{1}_\Omega(x) \in \{0,1\}$ holds for λ-almost every $x \in [0,1]$ (see [27]). There exists a unique $x_0 \in (\frac{1}{2}, 1)$ fulfilling $S(x_0) = 1 - x_0 < \frac{1}{2}$ and the properties of Ω imply that S is not convex on any non-degenerated subinterval of $[0,1]$. Based on S define a new transformation $T : [0,1] \to [0,1]$ by

$$T(x) = \begin{cases} S(x) & \text{if } x \in [0, x_0], \\ \frac{x - x_0}{2(1-x_0)} + S(x) & \text{if } x \in [x_0, 1]. \end{cases}$$

It is straightforward to verify that $T : [0,1] \to [0,1]$ is a strictly increasing homeomorphism of $[0,1]$, which fulfills all properties stated in assertion (b) of Theorem 26. T is, however, obviously not convex on any non-degenerated subinterval of $[0,1]$ (since its derivative is not non-decreasing on any non-empty open interval).

Remark 29. Let $T : [0,1] \to [0,1]$ be non-decreasing and right-continuous, and assume that $T(0) = 0$ holds. According to Theorem 26 in order to have $\overline{m}_T = \overline{w}_T$ the transformation T needs to be absolutely continuous on the interval $[0, x_0]$ - on the interval $[x_0, 1]$, however, T (interpreted as univariate measure-generating function) may be also have a non-degenerated discrete and/or singular component on $[x_0, 1]$ as long as $T' \geq 1$ holds λ-almost everywhere on $[x_0, 1]$.

Remark 30. The proof of Theorem 26 also shows that for non-decreasing $T : [0,1] \to [0,1]$ all copulas A with $\mu_A(\Gamma(T)) = \overline{m}_T = \overline{w}_T$ fulfill the following three conditions:

$$\begin{aligned} \mu_A\big([0,x_0] \times [0, T(x_0)]\big) &= \mu_A\big(([0,x_0] \times [0,1]) \cap \Gamma(T)\big) = T(x_0) \\ \mu_A\big([x_0,1] \times [T(x_0), 1]\big) &= \mu_A\big(([x_0,1] \times [T(x_0),1]) \cap \Gamma(T)\big) = 1 - x_0 \\ \mu_A\big(([x_0,1] \times [0, T(x_0)]\big) &= 0 \end{aligned}$$

Again working with non-decreasing rearrangements and using the previous results yields the following corollary:

Corollary 31. *For every measurable $T : [0,1] \to [0,1]$ the following two conditions are equivalent (as before T^* denotes the non-decreasing rearrangement and $\Lambda_{T^*} \in \mathcal{B}([0,1])$ the set of all points at which T^* is differentiable):*

(a) $\overline{m}_T = \overline{w}_T$.

(b) $T^*(0) = 0$ *and there exists some* $x_0 \in [0,1]$ *such that the following two properties hold:*

- *(i)* T^* *is absolutely continuous on* $[0, x_0]$,
- *(ii)* $\Omega_0 := \{x \in [0, x_0] \cap \Lambda_{T^*} : (T^\star)'(x) \leq 1\}$ *fulfills* $\lambda(\Omega_0) = x_0$,
- *(iii)* $\Omega_1 := \{x \in [x_0, 1] \cap \Lambda_{T^*} : (T^\star)'(x) \geq 1\}$ *fulfills* $\lambda(\Omega_1) = 1 - x_0$.

6. Estimating the maximum probability of a prior default

Throughout this section we assume that F and G are univariate continuous distribution functions, let T be defined by $T := G \circ F^-$ on $(0,1)$ and set $T(0) = 0$ and $T(1) = T(1-)$, which implies that T is left-continuous on $[0,1]$. Notice that for such T there exists some (not necessarily unique) $x \in (0,1]$ with $\overline{m}_T = 1 + T(x) - x$. If $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are independent samples of F and G, respectively, then it seems natural to estimate $\overline{m}_T$ by $\overline{m}_{T_n}$ where $T_n = G_n \circ F_n^-$ and F_n, G_n are the empirical distribution functions corresponding to $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. We are now going to show that $\overline{m}_{T_n}$ is a strongly consistent estimator for $\overline{m}_T$ and start with the following simple lemma.

Lemma 32. *Suppose that F and G are continuous univariate distribution functions. Then with probability one $\lim_{n\to\infty} |T_n(u) - T(u)| = 0$ holds for every continuity point $u \in (0,1)$ of F^-. In particular $(T_n)_{n\to\infty}$ converges to T λ-almost everywhere.*

Proof. Glivenko-Cantelli theorem implies that with probability we have uniform convergence of $(F_n)_{n\in\mathbb{N}}$ to F and of $(G_n)_{n\in\mathbb{N}}$ to G. Applying Lemma 21.2 in [33] it follows that for every continuity point $u \in (0,1)$ of F^- we have $\lim_{n\to\infty} F_n^-(u) = F^-(u)$ from which the desired result follows by a straightforward application of the triangle inequality. □

Theorem 33. *Suppose that F and G are continuous distribution functions and let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent samples of F and G, respectively. Then with probability one we have $\lim_{n\to\infty} \overline{m}_{T_n} = \overline{m}_T$, i.e. $\overline{m}_{T_n}$ is a strongly consistent estimator of $\overline{m}_T$.*

Proof. According to Lemma 32 we may assume that $(T_n)_{n\in\mathbb{N}}$ converges to T λ-almost everywhere. (i) Fix $\varepsilon > 0$ and suppose that $x \in (0,1]$ fulfills $\overline{m}_T = 1 + T(x) - x$. Then there exists some $z \in (x - \varepsilon, x)$ such that z is a continuity point of F^- and according to Lemma 32 we can find an index $n_0 \in \mathbb{N}$ such that $|T_n(z) - T(z)| < \varepsilon$, hence

$$\begin{aligned}\overline{m}_{T_n} &\leq 1 + T_n(z) - z \leq 1 + T(z) + \varepsilon - z \leq 1 + T(x) + \varepsilon - z < 1 + T(x) + \varepsilon - x + \varepsilon \\ &= 1 + T(x) - x + 2\varepsilon = \overline{m}_T - 2\varepsilon\end{aligned}$$

for every $n \geq n_0$. Considering that $\varepsilon > 0$ was arbitrary $\limsup_{n\to\infty} \overline{m}_{T_n} \leq \overline{m}_T$ follows.
(ii) Suppose now that $\liminf_{n\to\infty} \overline{m}_{T_n} = \overline{m}_T - 2\Delta$ holds for some $\Delta > 0$. Without loss of generality (choose an appropriate subsequence if necessary) we may assume that

$$\lim_{n\to\infty} \overline{m}_{T_n} = \overline{m}_T - 2\Delta.$$

Then for every $n \in \mathbb{N}$ there exists some $x_n \in (0,1]$ with $1 + T_n(x_n) - x_n < \overline{m}_{T_n} + \frac{\Delta}{2}$ and we can find an index n_0 such that for every $n \geq n_0$ we have $\overline{m}_{T_n} < \overline{m}_T - \frac{3\Delta}{2}$ and

$$1 + T_n(x_n) - x_n < \overline{m}_{T_n} + \tfrac{\Delta}{2} < \overline{m}_T - \Delta.$$

Compactness of $[0,1]$ implies the existence of a subsequence $(x_{n_j})_{j\in\mathbb{N}}$ with limit $x \in [0,1]$. Now, choose $\delta \in (0, \frac{\Delta}{4})$ so that $x - \delta$ is a continuity point of F^- and $T(x - \delta) \geq T(x) - \frac{\Delta}{4}$ holds. Choose $j_0 \in \mathbb{N}$ in such a way that $n_{j_0} \geq n_0$ and that $|x_{n_j} - x| \leq \delta$ for every $j \geq j_0$.

According to Lemma 32 we can find another index $j_1 \in \mathbb{N}$ in such a way that $n_{j_1} \geq n_{j_0}$ and that $|T_{n_j}(x-\delta) - T(x-\delta)| \leq \frac{\Delta}{4}$ for every $j \geq j_1$. Then for $j \geq j_1$ we altogether get

$$\begin{aligned} 1 + T(x) - x &\leq 1 + T(x-\delta) + \tfrac{\Delta}{4} - x \leq 1 + T_{n_j}(x-\delta) + \tfrac{\Delta}{2} - x \\ &\leq 1 + T_{n_j}(x_{n_j}) + \tfrac{\Delta}{2} - x \leq 1 + T_{n_j}(x_{n_j}) + \tfrac{\Delta}{2} - x_{n_j} + \tfrac{\Delta}{4} \\ &< \overline{m}_T - \Delta + \tfrac{3\Delta}{4} = \overline{m}_T - \tfrac{\Delta}{4}, \end{aligned}$$

a contradiction to the definition of $\overline{m}_T$. This shows $\liminf_{n\to\infty} \overline{m}_{T_n} \geq \overline{m}_T$ and the proof is complete. □

As final step we will now show that under mild regularity conditions on T (or, equivalently on F and G) the estimator $\overline{m}_{T_n}$ is asymptotically normal. To derive asymptotic normality we will apply the functional Delta method (see [33]) and build upon the following two lemmata, whereby as in [33] $[a,b] \subseteq [-\infty,\infty]$ and $\mathbb{D}([a,b])$ will denote the family of all cadlag functions endowed with the uniform distance $\|\cdot\|_\infty$:

Lemma 34. *Define $\phi\colon [a,b]\times\mathbb{D}[a,b] \to \mathbb{R}$ by $\phi(x,G) = G(x)$ and suppose that $G \in \mathbb{D}[a,b]$ is differentiable at $x \in (a,b)$. Then ϕ is Hadamard differentiable at (x,G) tangentially to the set of tuples $(h_1,h_2) \in \mathbb{R}\times\mathbb{D}[a,b]$ where h_2 is continuous at x, with derivative $\phi' : \mathbb{R}\times\mathbb{D}[a,b] \to \mathbb{R}$ fulfilling $\phi'(h_1,h_2) = h_1 G'(x) + h_2(x)$.*

Proof. Let $h_{1,t} \to h_1$, $h_{2,t} \to h_2$ and $t \to 0$ such that $x + th_{1,t} \in [a,b]$ for sufficiently small t. Then using Taylor's formula we get

$$\begin{aligned} \frac{\phi(x+th_{1,t}, G+th_{2,t}) - \phi(x,G)}{t} &= \frac{(G+th_{2,t})(x+th_{1,t}) - G(x)}{t} \\ &= \frac{G(x+th_{1,t}) + th_{2,t}(x+th_{1,t}) - G(x)}{t} \\ &= \frac{G(x) + G'(x)th_{1,t} + o(t) - G(x)}{t} + h_{2,t}(x+th_{1,t}) \\ &\to G'(x)h_1 + h_2(x), \end{aligned}$$

where in the last step we used continuity of h_2 at x. □

Lemma 35. *Define $\phi\colon \mathbb{D}[a,b]\times\mathbb{D}[a,b] \to \mathbb{R}$ by $\phi(F,G) = G \circ F^-(p)$, consider $p \in (0,1)$ and set $x_p = F^-(p) \in (a,b)$. Furthermore let $F,G \in \mathbb{D}[a,b]$ be differentiable at x_p with $F'(x_p) > 0$. Then ϕ is Hadamard differentiable at (F,G) tangentially to $(h_1,h_2) \in \mathbb{D}[a,b]\times\mathbb{D}[a,b]$ where h_1 and h_2 are continuous at x_p, with derivative $\phi' : \mathbb{D}[a,b]\times\mathbb{D}[a,b] \mapsto \mathbb{R}$ fulfilling*

$$\phi'(h_1,h_2) = -\,h_1(x_p)\,\frac{G'(x_p)}{F'(x_p)} + h_2(x_p).$$

Proof. As a consequence of [33, Lemma 21.3], the map $\phi_1\colon \mathbb{D}[a,b]\times\mathbb{D}[a,b] \to \mathbb{R}\times\mathbb{D}[a,b]$ defined by $\phi_1(F,G) = (F^-(p), G)$ is Hadamard differentiable at (F,G) tangentially to the set of functions $(h_1,h_2) \in \mathbb{D}[a,b]\times\mathbb{D}[a,b]$ where h_1 is continuous at x_p, with derivative $\phi_1' : \mathbb{D}[a,b]\times\mathbb{D}[a,b] \to \mathbb{R}\times\mathbb{D}[a,b]$ fulfilling $\phi_1'(h_1,h_2) = (-h_1(x_p)/F'(x_p), h_2)$. According to Lemma 34 the map $\phi_2\colon [a,b]\times\mathbb{D}[a,b] \to \mathbb{R}$ defined by $\phi_2(x,G) = G(x)$ is Hadamard differentiable at $(F^-(p), G)$ tangentially to the set of tuples $(h_1, h_2) \in \mathbb{R} \times \mathbb{D}[a, b]$ where h_2 is continuous at x_p, with derivative $\phi' : \mathbb{R} \times \mathbb{D}[a, b] \to \mathbb{R}$ fulfilling $\phi'(h_1, h_2) = h_1 G'(x) + h_2(x)$. It hence follows from the Chain rule for Hadamard derivatives (see [33, Theorem 20.9]) that the transformation $\phi_2 \circ \phi_1$ is Hadamard differentiable as well, which completes the proof. $\square$

Given an interval $[a, b] \subseteq \mathbb{R}$, let $\mathbb{D}_1$ denote the set of all restrictions of distribution functions on $\mathbb{R}$ to $[a, b]$, and $\mathbb{D}_2$ the subset of $\mathbb{D}_1$ consisting of all distribution functions of probability measures assigning mass 1 to $(a, b]$. Furthermore let $\mathbb{C}[a, b]$ denote the family of all continuous functions on $[a, b]$. The following corollary works analogously to Lemma 21.4 in [33].

Corollary 36.
1. *Let $0 < p_1 < p_2 < 1$ and let F, G be continuously differentiable on the interval $[a, b] = [F^-(p_1) - \epsilon, F^-(p_2) + \epsilon]$ for some $\epsilon > 0$, with the derivative of F being strictly positive. Then $\phi \colon \mathbb{D}_1 \times \mathbb{D}[a, b] \to \mathbb{D}[0, 1]$ defined by $\phi(\overline{F}, \overline{G}) = \overline{G} \circ \overline{F}^-$ is Hadamard differentiable at (F, G) tangentially to $\mathbb{C}[a, b] \times \mathbb{C}[a, b]$.*
2. *Let F have compact support $[a, b]$ and let F, G be continuously differentiable on $[a, b]$ with the derivative of F being strictly positive. Then $\phi \colon \mathbb{D}_2 \times \mathbb{D}[a, b] \to \mathbb{D}[0, 1]$ defined by $\phi(\overline{F}, \overline{G}) = \overline{G} \circ \overline{F}^-$ is Hadamard differentiable at (F, G) tangentially to $\mathbb{C}[a, b] \times \mathbb{C}[a, b]$.*

In both cases the derivative is the map

$$(h_1, h_2) \mapsto \left(- h_1 \frac{G'}{F'} + h_2 \right) \circ F^-.$$

The next result is immediate from [2]:

Lemma 37. *Define $\phi \colon \mathbb{D}[0, 1] \to \mathbb{R}$ as $\phi(T) = 1 + \inf_{x \in [0,1]}(T(x) - x)$. Let $T \in \mathbb{D}[0, 1]$ be such that there exists a unique $x^* \in (0, 1)$ with $1 + T(x^*-) - x^* = 1 + \inf_{x \in [0,1]} T(x) - x$. Then ϕ is Hadamard differentiable at T tangentially to the set of functions $h \in \mathbb{C}[0, 1]$ with derivative $\phi' : \mathbb{C}[0, 1] \to \mathbb{R}$ given by $\phi'(h) = h(x^*)$.*

We now show that under mild regularity conditions on T (or, equivalently on F and G) the estimator $\overline{m}_{T_n}$ is asymptotically normal.:

Theorem 38. *Let F_n and G_n be the empirical distribution functions of two independent random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ from (absolutely continuous) distribution functions F and G, respectively and let $T = G \circ F^-$. If T is such that there exists a unique $p^* \in [0, 1]$ with $T(p^*-) - p^* = \inf_{x \in [0,1]} T(x) - x$ and F, G are such as in Corollary 36, then for $T_n = G_n \circ F_n^-$,*

$$\sqrt{n}\big(\inf_x (T_n(x) - x) - \inf_x (T(x) - x) \big)$$

is asymptotically normal with mean 0 and variance

$$\frac{(G')^2(x_{p^*})}{(F')^2(x_{p^*})}\, p^*(1 - p^*) + G(x_{p^*})(1 - G(x_{p^*}))$$

where $x_{p^*} = F^-(p^*)$.

Proof. According to Donsker's Theorem (see [33]) $(\mathbb{G}_{n,F}, \mathbb{G}_{n,G}) = \sqrt{n}(F_n - F, G_n - G)$ converges in distribution to $(\mathbb{G}_F, \mathbb{G}_G)$ in the space $\mathbb{D}[-\infty, \infty] \times \mathbb{D}[-\infty, \infty]$, for a pair of independent Brownian Bridges $\mathbb{G}_F$ and $\mathbb{G}_G$. The sample paths of the two limit processes are continuous, since both, F and G, are continuous. By Corollary 36, Lemma 37 and the Chain Rule for Hadamard derivatives $\phi(F, G) = 1 + \inf_{x\in[0,1]}(G \circ F^{-}(x) - x)$ is Hadamard differentiable tangentially to the range of the limit processes. Applying the functional delta method yields that the sequence $\sqrt{n}(\inf_x(T_n(x)-x)-\inf_x(T(x)-x))$ is asymptotically equivalent to the derivative of ϕ evaluated at $(\mathbb{G}_{n,F}, \mathbb{G}_{n,G})$, i.e., to $-\frac{G'(x_{p^*})}{F'(x_{p^*})}\mathbb{G}_{n,F}(x_{p^*}) + \mathbb{G}_{n,G}(x_{p^*})$. Asymptotic normality now follows from the central limit theorem. $\square$

Theorem 38 considered uniqueness of the point attaining the infimum, the following final result focuses the other extreme case where each point is a minimizer:

Theorem 39. *Let F_n and G_n be the empirical distribution functions of two independent random samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ and let $T = G \circ F^{-1}$. If F, G are both $U(0,1)$ then $\sqrt{n}(\inf_x(T_n(x)-x)-\inf_x(T(x)-x))$ converges to $\inf_{t\in(0,1)} \sqrt{2}B_t$ (with B_t being a standard Brownian Bridge) and thus has density $f(x) = -2x\exp(-x^2)\mathbf{1}_{(-\infty,0]}(x)$.*

The following final example illustrates Theorem 38.

Example 40 (Example 6 continued). Consider the setting from Example 6 for the case $\theta_1 = 2$ and $\theta_2 = 1$. Then it is straightforward to verify that all assumptions of Theorem 38 are fulfilled, that $p^* = \frac{3}{4}$ is the unique minimizer, that $x_p = F^{-}(p^*) = \ln 2$, and that the asymptotic variance σ^2 is given by $\sigma^2 = \frac{7}{16}$. The right panel in Figure 4 depicts a histogram of $R = 1.000$ samples of the random variable $Z_n := \sqrt{n}(\overline{m}_{T_n} - \overline{m}_T)$ calculated by randomly drawing independent samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ from $X \sim Ex(\theta_1)$ and $Y \sim Ex(\theta_2)$ of size $n = 100.000$, respectively.

Remark 41. Based on simulations we conjecture that working with Bernstein approximations or splines it might be possible to derive strongly consistent estimators for $\overline{w}_T$ too. We plan to tackle this question in the near future.

7. Application: Estimation of the relative effect

In this section we relate our results to the estimation of the relative effect of a random variable Y on another random variable X. Recall that, for X and Y with distribution functions F and G, respectively, the quantity

$$p_{XY} := \int_\Omega G \circ X \, \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} G(x) \, \mathrm{d}\mathbb{P}^X(x)$$

is commonly referred to as the *relative effect* and appears, e.g., in the rank test of Wilcoxon, Mann and Whitney for the hypothesis $H_0 : F = G$. If X and Y are independent, then

$$p_{XY} = \int_\mathbb{R}\int_\mathbb{R} \mathbb{1}_{(-\infty,x]}(y) \, \mathrm{d}\mathbb{P}^Y(y)\mathrm{d}\mathbb{P}^X(x) = P(\{Y \leq X\})$$

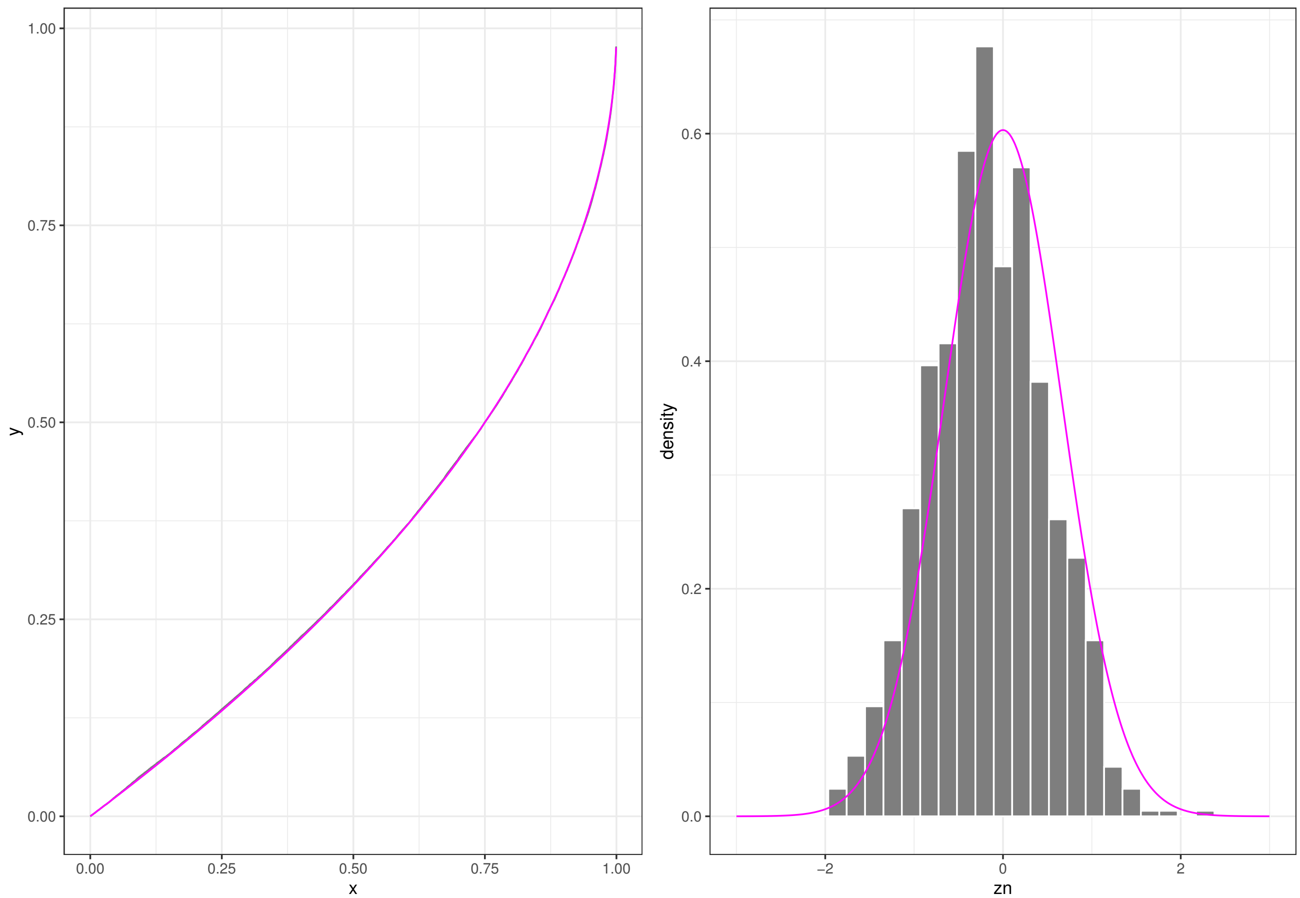

Figure 4: Left panel: T (magenta) and T_n with $n = 100.000$ as considered in Example 6; Right panel: Histogram of $R = 1.000$ values of Z_n with $n = 100.000$ and density of $\mathcal{N}(0, \frac{7}{16})$.

According to [1], X is called tendentiously larger than Y if $p_{XY} > \frac{1}{2}$, X is called tendentiously smaller than Y if $p_{XY} < \frac{1}{2}$, and, if $p_{XY} = \frac{1}{2}$, no tendency exists for the values of Y to be either larger or smaller that those of X.

Considering an i.i.d sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y) it is natural to estimate p_{XY} by the plug-in estimator

$$\hat{p}_{XY} = \frac{1}{n} \sum_{k=1}^{n} G_n(X_k) \tag{38}$$

where G_n denotes the empirical distribution function corresponding to $Y_1, \ldots, Y_n$. Since X_k and Y_l are independent whenever $k \neq l$ we obtain

$$\begin{aligned} E(\hat{p}_{XY}) &= \frac{1}{n} \sum_{k=1}^{n} E(G_n(X_k)) \\ &= \frac{1}{n} \sum_{k=1}^{n} \frac{1}{n} \sum_{l=1}^{n} P(\{Y_l \leq X_k\}) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1,l\neq k}^{n} P(\{Y_l \le X_k\}) + \frac{1}{n^2}\sum_{k=1}^{n} P(\{Y_k \le X_k\}) \\
&= \frac{n-1}{n}\, p_{XY} + \frac{1}{n}\, P(\{Y \le X\}) \\
&= p_{XY} + \frac{1}{n}\, \big(P(\{Y \le X\}) - p_{XY}\big)
\end{aligned}
$$

In case X and Y are independent, it follows from $E(\hat{p}_{XY}) = \frac{n-1}{n}\, p_{XY} + \frac{1}{n}\, P(\{Y \le X\}) = p_{XY}$ that $\hat{p}_{XY}$ is unbiased. In presence of dependence between X and Y, however, the estimator can be biased. In such a situation, the above equation allows to calculate lower and upper bounds for the relative effect p_{XY}: Since

$$
\frac{n-1}{n}\, p_{XY} + \frac{1}{n}\, \underline{m}_T \le E(\hat{p}_{XY}) \le \frac{n-1}{n}\, p_{XY} + \frac{1}{n}\, \overline{m}_T
$$

we obtain

$$
\frac{n}{n-1}\, E(\hat{p}_{XY}) - \frac{1}{n-1}\, \overline{m}_T \le p_{XY} \le \frac{n}{n-1}\, E(\hat{p}_{XY}) - \frac{1}{n-1}\, \underline{m}_T \tag{39}
$$

In manifold situations independence between the two samples is not realistic, in particular in the case when data is collected from a group of individuals over time. In this case, Inequality (39) can be used to obtain lower and upper bounds for p_{XY}.

Example 42. For an illustration, let us consider the dataset depression provided in the R package datarium. The dataset contains the depression score of patients from two groups (*control*: 12 patients; *treatment*: 12 patients) at different points in time (0: pre-test, 1: one month post-test, 3: 3 months follow-up and 6: 6 months follow-up). We denote by X_{ik} the observation of patient $i \in \{1, \dots, 12\}$ from group *treatment* at time $k \in \{0, 1, 3, 6\}$. Since the different time points are close to each other, we cannot assume independence of the samples $(X_{ik})_{i\in\{1,\dots,12\}}$, $k \in \{0, 1, 3, 6\}$ (Pearson correlation between $X_{.1}$ and $X_{.6}$ equals 0.49), such that the estimator for the relative effect may be biased.
Estimating the maximum probability of a prior default as described in Section 6, the minimum probability of a prior default according to Corollary 12, and the relative effect following (38), we obtain

$\underline{m}_T$	$\hat{p}_{X_{.1}X_{.6}}$	$\overline{m}_T$
0.1667	0.4167	0.7500

Now, Equation (39) yields $p_{X_{.1}X_{.6}} \in [0.3864, 0.4394]$. Therefore, we may conclude that the depression score at time $k = 1$ is tendentiously smaller than the depression score at time $k = 6$.

In presence of dependent data the plug-in estimator (38) may be biased, a fact that should be taken into account especially in the case of small sample sizes, which are common in medicine.

Acknowledgments

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

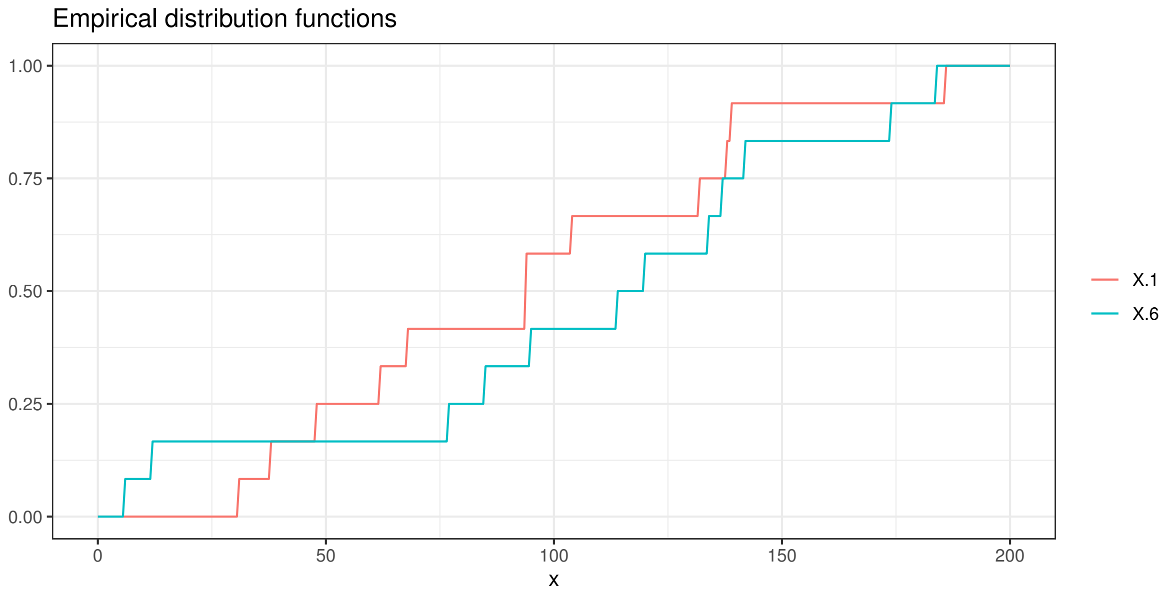

Figure 5: Empirical distributions functions of samples $(X_{1,k})_{i\in\{1,\dots,12\}}$ (red line) and $(X_{3,k})_{i\in\{1,\dots,12\}}$ (blue line) discussed in Example 42.

References

[1] E. Brunner, S. Domhof, F. Langer: *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, John Wiley & Sons, New York, (2002).

[2] J. Cárcamo, A. Cuevaz, L.-A. Rodríguez: Directional differentiability for supremum-type functionals: statistical applications, *Bernoulli* **26**, 2143-2175 (2020). (see `https://arxiv.org/abs/1902.01136`)

[3] U. Cherubini, E. Luciano, W. Vecchiato: *Copula Methods in Finance*, John Wiley & Sons, New York, (2004).

[4] F. Durante, J. Fernández Sánchez, W. Trutschnig: On the singular components of a copula, *J. Appl. Probab.* **52**, 1175-1182 (2015).

[5] W.F. Darsow, B. Nguyen, E.T. Olsen: Copulas and Markov processes, *Illinois J. Math.* **36**, 600-642 (1992).

[6] F. Durante, P. Sarkoci, C. Sempi: Shuffles of copulas, *J. Math. Anal. Appl.* **352**, 914-921 (2009).

[7] F. Durante, E.P. Klement, J. Quesada-Molina, P. Sarkoci: Remarks on Two Product-like Constructions for Copulas, *Kybernetika* **43**, 235–244 (2007).

[8] F. Durante, C. Sempi: *Principles of Copula Theory*, Chapman and Hall/CRC, 2015.

[9] J. Elstrodt: *Mass- und Integrationstheorie*, Springer, (1999).

[10] P. Embrechts, G. Puccetti: Bounds for functions of dependent risks, *Finance and Stoch* **10**, 341-352 (2006).

[11] P. Embrechts, G. Puccetti: Bounds for functions of multivariate risks, *J. Multivariate Anal.* **97**, 526-547 (2006).

[12] P. Embrechts, M. Hofert: A note on generalized inverses, *Math. Method. Oper. Res.* **77**, 423-432 (2013).

[13] J. Fernández Sánchez, W. Trutschnig: Conditioning based metrics on the space of multivariate copulas and their interrelation with uniform and levelwise convergence and Iterated Function Systems, *Journal of Theoretical Probability* **28**, 1311-1336 (2015).

[14] J. Fernández Sánchez, W. Trutschnig: Some members of the class of (quasi-) copulas with given diagonal from the Markov kernel perspective, Comm. Stat. A–Theor. **45**, 1508-1526 (2016).

[15] A.L. Gibbs, F.E. Su: On choosing and bounding probability metrics, *Int. Stat. Rev.* **70**(3), 419-435 (2002).

[16] E. Hewitt, K. Stromberg: *Real and Abstract Analysis*, Springer Verlag, Berlin Heidelberg, (1965).

[17] O. Kallenberg: *Foundations of modern probability*, Springer Verlag, New York Berlin Heidelberg, (1997).

[18] A. Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, Berlin Heidelberg, (2007).

[19] H.O. Lancaster: Correlation and complete dependence of random variables, *Ann. Math. Stat.* **34**, 1315-1321 (1963).

[20] L. Lindvall: *Lectures on the Coupling Method*, John Wiley & Sons, New York, (1992).

[21] A.W. Marshall, I. Olkin: *Life Distributions*, Springer, New York (2007).

[22] A.J. McNeil, R. Frey, P. Embrechts: *Quantitative Risk Management*, Princeton University Press (2005).

[23] J.F. Mai, M. Scherer: Simulating from the copula that generates the maximal probability for a joint default under given (inhomogeneous) marginals, in *Topics from the 7th International Workshop on Statistical Simulation* ed. V. Melas, S. Mignani, P. Monari, and L. Salmaso, Springer Proceedings in Mathematics & Statistics **114**, pp. 333-341, 2014.

[24] P. Mikusínski, H. Sherwood, M.D. Taylor: Shuffles of Min, *Stochastica* **290**, 61-74 (1992).

[25] T. Mroz, S. Fuchs, W. Trutschnig: How simplifying and flexible is the simplifying assumption in pair-copula constructions - analytic answers in dimension three and a glimpse beyond, *Electronic Journal of Statistics* **15**, 1951-1992 (2021).

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

[27] W. Rudin: *Real and Complex Analysis*, McGraw-Hill International Editions, Singapore, (1987).

[28] L. Rüschendorf: Random Variables with Maximum Sums, Advances in Applied Probability, Vol. 14, No. 3, 623-632 (1982).

[29] J.V. Ryff: Measure Preserving Transformations and Rearrangements, *J. Math. Anal. Appl.* **31**, 449-458 (1970).

[30] H. Thorisson: *Coupling, stationarity, and regeneration*, Probability and its Applications, Springer-Verlag, New York, (2000).

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

[32] W. Trutschnig, J. Fernández Sánchez: Some results on shuffes of two-dimensional copulas, *J. Stat. Plan. Infer.* **143**, 251-260 (2013).

[33] A.W. van der Vaart: *Asymptotic Statistics*, Cambridge University Press, (2000).

[34] T. Zamfirescu: Most Monotone Functions are Singular, *The American Mathematical Monthly* **88**(1), 47-49 (1981)