Extremal and functional dependence between continuous random variables

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik der
Technischen Universität Dortmund
vorgelegt von

Christopher Strothmann

im November 2021

Dissertation

Extremal and functional dependence between continuous random variables

Fakultät für Mathematik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Karl Friedrich Siburg

Zweitgutachter: Prof. Dr. Wolfgang Trutschnig

Tag der mündlichen Prüfung: 2. März 2022

Acknowledgements

This dissertation would not have been possible without the support of the people around me.

First and foremost, I want to express my sincerest gratitude to my advisor Karl Friedrich Siburg for his excellent guidance throughout both my undergraduate and my doctoral studies. From the contents of this thesis over approaches to teaching to professional advice, our discussions were always thought-provoking and have helped me become the person I am today. Secondly, I am very grateful to Wolfgang Trutschnig, whose article on measures of complete dependence first started me on this journey and who has, at first unknowingly, answered a multitude of questions. I greatly appreciate his and his research groups' hospitality during my visit to Salzburg and the insightful comments he shared regarding some of the research questions considered in this thesis. I am moreover grateful to Holger Dette, Sebastian Fuchs, Piotr Jaworski and Gregor N. F. Weiß for many interesting and fruitful discussions.

I highly appreciate the material as well as non-material support of the German National Academic Foundation, both during my undergraduate and my doctoral studies.

For valuable mathematical discussions, relaxing coffee breaks, entertaining game nights and rather bad yet amusing puns, I owe special thanks to my current and former colleagues, in particular to Alexander Dicke, Max Kämper and Albrecht Seelmann. I thank Sara Schmidt for plenty of reasons, none of which can be taken for granted. At this place, I thank her both for innumerable hours spent discussing and proofreading all parts of this Ph.D. project, and for her unwavering support. Last but not least, I would like to thank my parents and my sister for their unconditional support.

Contents

1 Introduction

Imagine the task of assembling a portfolio from different stocks or coordinating an appropriate response to changes in the water level of multiple water reservoirs. Understanding the stochastic relationship between the stock returns or water levels can mean the difference between a high-risk investment and a hedged portfolio or an adequate water supply and a shortage. Capturing these interactions and extracting relevant information lies at the heart of dependence modelling.

Naturally, investigating the joint behaviour of d random variables requires us to shift focus from their individual behaviour towards their interaction. Sklar's famous theorem sheds light on how this interaction is encoded in the joint distribution function: Given d random variables $X_1, \ldots, X_d$, their joint distribution function F can be decomposed into the univariate marginal distribution functions F_i of X_i for $i = 1, \ldots, d$ and a linking function $C : [0,1]^d \to [0,1]$ such that

$$F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{1.1}$$

holds for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Since Equation (1.1) couples d marginal distribution functions into a d-variate cumulative distribution function, C is called a *d-copula*.

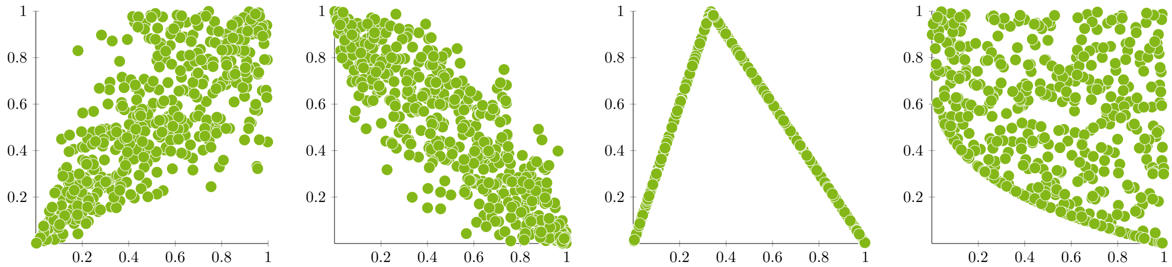
The appeal of copulas in dependence modelling stems from the fact that in case all margins F_i are continuous, C is *the unique* copula of $\boldsymbol{X} = (X_1, \ldots, X_d)$, which we will denote by $C_{\boldsymbol{X}}$. Intuitively, $C_{\boldsymbol{X}}$ then encodes all (scale independent) information concerning the interaction between the random variables. But the mere existence of such a unique copula leaves several important aspects unanswered, which will be central to this thesis:

- What are suitable dependence concepts?
- How can we compare dependence?
- How can we quantify dependence?

Solutions to these questions might be conflicting: Consider for example the broadest possible definition of stochastic dependence, where X and Y are called *stochastically dependent* if they are not stochastically independent. This classification includes vastly different stochastic behaviours as illustrated by Figure 1.1, hindering a precise analysis and meaningful comparison. Thus, to allow for a comparison of dependence strength, it is important to investigate general but concise concepts of dependence.

Figure 1.1 already suggests *one* possible distinction between dependence concepts. While (a) to (c) appear to follow a global underlying dependence structure (namely conical, elliptical and functional, respectively), the behaviour depicted in (d) varies locally.

A well-known local type of dependence between random variables, which graphically manifests in the lower left corner of the unit cube $[0,1]^d$, is the *tail dependence*. For a random



(a) (b) (c) (d)

Figure 1.1: Plots of samples of size $n = 500$ generated from different bivariate distributions.

vector[1] $\boldsymbol{X} = (X_1, X_2)$ with continuous univariate marginal distributions F_1 and F_2, the (lower) tail dependence coefficient is defined as

$$\lambda(\boldsymbol{X}) := \lim_{s \searrow 0} \mathbb{P}(X_2 \leq F_2^{-1}(s) \mid X_1 \leq F_1^{-1}(s)) = \lim_{s \searrow 0} \frac{C_{\boldsymbol{X}}(s,s)}{s} \,. \tag{1.2}$$

$\lambda(\boldsymbol{X})$ was first introduced by Sibuya (1960) and has since then found wide-spread use in many areas of science, ranging from economics to environmental studies. Revisiting the examples from the beginning, $\lambda(\boldsymbol{X})$ describes the probability of extremely large simultaneous losses in a portfolio of stock returns, or the probability of dangerously low water levels at all water reservoirs. $\lambda(\boldsymbol{X})$ is thus an important indicator for informed investment or policy decisions regarding *worst-case* scenarios. Nevertheless, the tail dependence coefficient has some severe drawbacks, which have partly been overcome by using the tail dependence function

$$\Lambda\,(\boldsymbol{w}\,; \boldsymbol{X}) := \Lambda\,(\boldsymbol{w}\,; C_{\boldsymbol{X}}) := \lim_{s \searrow 0} \frac{C_{\boldsymbol{X}}(s\boldsymbol{w})}{s} \quad \text{for } \boldsymbol{w} \in \mathbb{R}_+^2$$

as the natural generalization of $\lambda(\boldsymbol{X})$. From an analytical perspective, $\Lambda\,(\,\cdot\,; C_{\boldsymbol{X}})$ governs the behaviour of $C_{\boldsymbol{X}}$ around zero as it constitutes a Taylor-like expansion of $C_{\boldsymbol{X}}$ in $\mathbf{0}$, that is, $C_{\boldsymbol{X}}(\boldsymbol{u}) = \Lambda\,(\boldsymbol{u}\,; C_{\boldsymbol{X}}) + R(\boldsymbol{u})\,\|\boldsymbol{u}\|_1$ with a function $R(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\|_1 \to 0$ (see Jaworski (2006)).

Fundamentally different from this extremal dependence, where a high tail dependence gives no indication about the joint behaviour outside a neighbourhood of zero, functional dependence describes the global relationship of two random variables. Given a random vector $\boldsymbol{X} = (X_1, X_2)$, X_2 is called functionally dependent on X_1 if there exists a function f such that $X_2 = f(X_1)$ holds almost surely. This corresponds to the well-known concept of *linear dependence* in case f is affine linear and, in its most general form, to the concept of *complete dependence* in case f is merely measurable. Complete dependence captures the perfect predictability of X_2 given X_1 and, in some sense, constitutes *a* counterpart to the independence of random variables. Considering that the precise knowledge of the function f with $X_2 = f(X_1)$ is often unattainable and may even be unnecessary, measures of complete dependence have garnered considerable attention. In recent years, the literature on these measures has expanded significantly and now ranges from theoretical works (see, e.g., Siburg and Stoimenov

[1]For the sake of a concise notation and a better overview, we present the central ideas in this introduction for dimension $d = 2$ only.

(2010) and Trutschnig (2011)) over statistical analyses (see, e.g., Dette, Siburg and Stoimenov (2013), Chatterjee (2021) and Junker, Griessenberger and Trutschnig (2021)) to practical applications in various fields, such as climate studies and biology (see, e.g., Fruciano, Colangelo, Castiglia and Franchini (2020) and the examples in the previous references).

1.1 Main results of this thesis

Chapter 3: Comparing the extremal behaviour of copulas

Its simplicity and seemingly straightforward interpretation have established the tail dependence coefficient λ as a widely used tool in areas such as finance and hydrology. Despite its popularity, λ has some apparent deficiencies: Equation (1.2) shows that λ is solely determined by the behaviour of the copula $C_{\boldsymbol{X}}$ along the diagonal, thereby ignoring any additional information located in a neighbourhood around zero. Put plainly, one may wonder why the direction (s, s) should be preferred to any other direction, say $(s, 2s)$.

Therefore, Chapter 3 of this thesis introduces a novel comparison of extremal dependence, which considers all possible (ray-like) directions. More precisely, a random vector $\boldsymbol{X}$ is called *less tail dependent* than $\boldsymbol{Y}$, in short $\boldsymbol{X} \leq_{tdo} \boldsymbol{Y}$, if

$$\Lambda(\,\cdot\,;\boldsymbol{X}) \leq \Lambda(\,\cdot\,;\boldsymbol{Y}) \ . \tag{1.3}$$

Although the tail dependence function contains the entire information about the extremal behaviour of the random vector, it is rather difficult to scan the d-variate tail dependence function for specific properties in dimensions much higher than $d = 2$. To this end, we discuss measures of tail dependence that are consistent with the ordering (1.3) but condense the whole tail dependence into a single numerical quantity. Contrary to λ, this class of generalized measures provides a more comprehensive picture of the overall tail behaviour of the random vector and comprises a variety of natural quantities, such as the average or maximal tail dependence. In fact, even many formerly introduced coefficients, such as the conditional version of Spearman's ρ proposed by Schmid and Schmidt (2007), constitute measures of tail dependence.

Compared to $\leq_{tdo}$, the localized stochastic ordering of copulas $\leq_{loc}$ (i.e. a pointwise ordering on some neighbourhood of $\mathbf{0}$) is in some sense the strongest possible extremal ordering around $\mathbf{0}$. Still, using techniques from concordance ordering, we establish an equivalence between $\leq_{tdo}$ and $\leq_{loc}$ for some important classes of copulas like Archimedean copulas. For these classes, we find that $\boldsymbol{X} <_{tdo} \boldsymbol{Y}$ not only orders the limit relation

$$\lim_{s \searrow 0} \frac{C_{\boldsymbol{X}}(s\boldsymbol{w})}{s} < \lim_{s \searrow 0} \frac{C_{\boldsymbol{Y}}(s\boldsymbol{w})}{s}$$

but also states that $\boldsymbol{X}$ is in fact less likely than $\boldsymbol{Y}$ to attain very low values in both components simultaneously.

Chapter 4: A Markov product for tail dependence functions

The generalized Markov product for copulas

$$(C_1 *_C C_2)(u_1, u_2, u_3) := \int_0^{u_3} C\left(\partial_1 C_1(t, u_1), \partial_1 C_2(t, u_2)\right) \, \mathrm{d}t$$

finds extensive application as both a theoretical tool and a construction method for multivariate copulas. In Chapter 4, we make use of the similarities between copulas and tail dependence functions to introduce a (generalized) Markov product for tail dependence functions via

$$(\Lambda_1 *_C \Lambda_2)(w_1, w_2, w_3) := \int_0^{w_3} C\left(\partial_1 \Lambda_1\left(t, w_1\right), \partial_1 \Lambda_2\left(t, w_2\right)\right) \, \mathrm{d}t \; .$$

Considering our focus on the tail behaviour of copula families in Chapter 3, we then pose the following question: Is it possible to derive the tail dependence function of $(C_1 *_C C_2)$ simply from the generalized Markov product of the respective tail dependence functions, that is, does

$$(\Lambda\left(\,\cdot\,; C_1\right) *_C \Lambda\left(\,\cdot\,; C_2\right))(\boldsymbol{w}) = \Lambda\left(\boldsymbol{w}\,; C_1 *_C C_2\right) \tag{1.4}$$

hold? The answer is yes, we can indeed establish (1.4) if we impose certain conditions on the partial derivatives of the copulas C_1 and C_2.

After establishing the connection between the two Markov products, we investigate the algebraic and analytical aspects of the Markov product for tail dependence functions in more detail. Although the Markov products for copulas and for tail dependence functions share many properties, the concavity of the tail dependence functions induces several 'reduction' properties. Firstly, the Markov product reduces the tail dependence, i.e.

$$(\Lambda_1 *_C \Lambda_2)(w_1, w_2) \leq \min\left\{\Lambda_1(w_2, w_1), \Lambda_2(w_1, w_2)\right\} \; , \tag{1.5}$$

whenever C is negative quadrant dependent. Secondly, for the 'original' Markov product

$$(\Lambda_1 * \Lambda_2)(w_1, w_2) := \int_0^{\infty} \partial_1 \Lambda_1\left(t, w_1\right) \cdot \partial_1 \Lambda_2\left(t, w_2\right) \, \mathrm{d}t \; ,$$

the concavity enables us to characterize all idempotents (i.e. all tail dependence functions Λ with $\Lambda * \Lambda = \Lambda$) and limits of n-fold iterations $\Lambda^{*n} = \Lambda * \cdots * \Lambda$ as either $\Lambda^0(\boldsymbol{w}) = 0$ or $\Lambda^+(\boldsymbol{w}) = \min \boldsymbol{w}$.

We conclude Chapter 4 with an investigation of the Markov product for tail dependence functions from an operator-theoretic point of view. Similar to the connection between the Markov product for copulas and the Markov operators endowed with the composition, we establish an isomorphism between tail dependence functions and substochastic operators, such that the Markov product corresponds to the composition of substochastic operators. Most importantly, this rephrases the reduction property in Equation (1.5) in terms of a monotonicity property with respect to the well-known majorization order introduced by Hardy, Littlewood and Pólya (1952) and Ryff (1965).

Chapter 5: Stochastic monotonicity and the Markov product for copulas

The Markov product $C_1 * C_2$ of copulas generates a wide variety of behaviours by combining the underlying copulas C_1 and C_2 in a highly nonlinear manner. But this flexibility comes at a price: In general, the global behaviour of $C_1 * C_2$ is virtually unpredictable from either C_1 or C_2 alone. In Chapter 5, we investigate copulas where one factor, say C_2, does provide some indication about the behaviour of $C_1 * C_2$. More precisely, we study copulas C_2 that are maximal with regard to the action of the Markov product, i.e. fulfil

$$(C_1 * C_2)(\boldsymbol{u}) \leq C_2(\boldsymbol{u}) \tag{1.6}$$

for all copulas C_1. The fact that $C_1 * C_2$ is pointwise smaller than C_2 hints at the fact that Equation (1.6) captures some kind of positive dependence concept.

Indeed, using techniques from majorization theory, we show that C_2 fulfils the reduction property (1.6) if and only if its partial derivative $u \mapsto \partial_1 C_2(u, v)$ is decreasing for all $v \in [0, 1]$. Copulas with such a decreasing partial derivative are known as *stochastically increasing.* Examples include Gaussian, extreme-value and certain Archimedean copulas as well as more complex copulas constructed from these as building blocks. Pursuing the monotonicity of the partial derivatives further, we prove that if two copulas C_1 and C_2 are stochastically increasing, then so is their Markov product $C_1 * C_2$. Moreover, sequences of stochastically increasing copulas possess improved convergence properties, where the pointwise convergence implies the much stronger pointwise convergence of the partial derivatives.

Similar to our findings for tail dependence functions, the reduction property (1.6) enables us to characterize idempotents (i.e copulas for which $C * C = C$ holds) and limits of n-fold iterates $C^{*n} = C * \cdots * C$ of stochastically increasing copulas as ordinal sums of the independence copula. This greater variety of possible limit behaviours compared to the case of tail dependence functions in Chapter 4 is owed to the fact that copulas are, in general, not positive homogeneous.

Chapter 6: Rearranging copulas and dependence measures

Although measures of complete dependence have recently attracted considerable attention, both theoretically and practically, measures of monotone dependence are still ubiquitous in applications.[2] Consequently, it is quite natural to wonder whether it is possible to infer a general functional relationship $X_2 = f(X_1)$ using only measures of monotone dependence. At first glance, this seems impossible to achieve. After all, any measure of monotone dependence μ attains its maximal value $\mu(X_1, X_2) = 1$ whenever $X_2 = g(X_1)$ holds for some monotone function g, whereas any more general functional relationship can result in arbitrarily small values of μ. Furthermore, while μ is necessarily symmetric, complete dependence constitutes a fundamentally asymmetric concept.

In Chapter 6, we present a method to circumvent these 'deficiencies' of monotone dependence measures by rearranging the underlying dependence structure of X_1 and X_2. This so-called (SI)-rearrangement $C_{\boldsymbol{X}}^{\uparrow}$ depends only on the original copula $C_{\boldsymbol{X}}$ and is constructed

[2]Here, μ is called a measure of monotone dependence if $\mu(X_1, X_2) = 1$ if and only if $X_2 = g(X_1)$ holds a.s. for a monotone function g. Similarly, μ is a measure of complete dependence if $\mu(X_1, X_2) = 1$ if and only if $X_2 = f(X_1)$ holds a.s. for a measurable function f.

by decreasingly rearranging the conditional probabilities

$$\partial_1 C_{\boldsymbol{X}}(F_1(x_1), F_2(x_2)) = \mathbb{P}(X_2 \leq x_2 \mid X_1 = x_1)$$

with respect to x_1. Most importantly, we show that although $C_{\boldsymbol{X}}$ and $C^{\uparrow}_{\boldsymbol{X}}$ share the same degree of complete dependence, $\mu(C_{\boldsymbol{X}})$ and $\mu(C^{\uparrow}_{\boldsymbol{X}})$ can vary drastically for some measure μ of monotone dependence. This distinction is essential to construct a measure of complete dependence from μ: The (SI)-rearrangement transforms arbitrary functional dependence into a stochastically increasing relationship, which can in turn be accurately quantified using a measure of monotone dependence μ via

$$R_\mu(\boldsymbol{X}) := R_\mu(C_{\boldsymbol{X}}) := \mu(C^{\uparrow}_{\boldsymbol{X}}) \ . \tag{1.7}$$

We call measures R_μ of the above form (1.7) *rearranged dependence measures*, where possible choices for μ include Spearman's ρ, Kendall's τ or the Schweizer-Wolff measures σ_p with $1 \leq p < \infty$. In stark contrast to the properties of these underlying measures, R_μ constitutes a genuine measure of complete dependence that precisely detects arbitrary functional relationships. Rearranged dependence measures also conform to information theoretical concepts such as the data processing inequality and the self-equitability condition introduced in Kinney and Atwal (2014). For the special case of measures of concordance κ, we additionally obtain $\kappa(\boldsymbol{X}) \leq R_\kappa(\boldsymbol{X})$, i.e. the functional dependence between X_1 and X_2 is at least as strong as the monotone relationship. To practically apply the rearranged dependence measures, we provide a simple estimator $\widehat{R}_\mu$ for R_μ based on convergence results for the empirical checkerboard copulas developed in Junker et al. (2021).

2 Mathematical preliminaries

While the concept of independence is widely employed throughout all areas of probability theory and statistics, there is no universal notion of the opposite. Rather, it depends heavily on the setting, e.g., dependence between components of a random vector or serial dependence in a time series, and the preferred type of association, e.g., a linear, monotone or merely measurable relationship.

This chapter lays the groundwork to describe the dependence between components of a random vector by providing the necessary tools and notations. We start by introducing copulas, our primary tool to capture dependence, in Section 2.1 and discuss their connection to Markov operators and conditional expectations in Section 2.2. Afterwards, we outline different local and global dependence concepts in Sections 2.3 and 2.4 and present several copula families in Section 2.5. We conclude this chapter with a brief introduction to decreasing rearrangements and majorization theory in Section 2.6. While the sections concerning copula theory and stochastic dependence mainly follow Nelsen (2006), Joe (2015) and Durante and Sempi (2016), the review of majorization theory is based on Chong and Rice (1971) and Bennett and Sharpley (1988). Whenever we require further results, we will provide some additional references. As a notational convention, we write $\mathbb{R}_+ := [0, \infty)$ and use bold symbols to denote vectors, e.g. $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

2.1 Copulas

Copulas are central to modern dependence modelling as a means to separate the influence of the marginal distributions from the *dependence* governing the relationship between a collection of random variables. We will first present copulas from an axiomatic point of view utilizing d-increasing functions, *one possible* multivariate generalization of monotonicity. The d-increasing property is known in the literature under various terms, such as quasi-monotonicity and, in case of $d = 2$, supermodularity. Intuitively, a function F is called d-increasing if it associates a positive volume to each rectangle. In the following, $\operatorname{dom} F$ denotes the domain of F and will generally either be $[0, 1]^d$ or $\mathbb{R}_+^d = [0, \infty)^d$.

Definition 2.1.1. *Let F be a d-variate real-valued function with $d \geq 1$.*

1. *The F-volume of a rectangle $R := [\boldsymbol{a}, \boldsymbol{b}) \subseteq \operatorname{dom} F$ is defined as*

$$V_F(R) := \Delta_{a_d}^{b_d} \cdots \Delta_{a_1}^{b_1} F \ ,$$

where $\Delta_{a_k}^{b_k}$ denotes the k-th component difference operator

$$\Delta_{a_k}^{b_k} F(\boldsymbol{t}_k) := F(t_1, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_d) - F(t_1, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_d)$$

evaluated in $\boldsymbol{t}_k := (t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_d)$.

2. *F is called d-increasing if the F-volume of every rectangle $R := [\boldsymbol{a}, \boldsymbol{b}) \subseteq \operatorname{dom} F$ is positive, i.e. $V_F(R) \geq 0$. We denote the usual measure-theoretic extension of V_F onto the Borel σ-algebra by μ_F.*

A special class of d-increasing functions are copulas.

Definition 2.1.2. *A function $C : [0,1]^d \to [0,1]$ is called a d-copula if it fulfils the following properties*

1. *C is grounded, i.e. $C(\boldsymbol{u}) = 0$ if $u_k = 0$ holds for at least one coordinate k.*

2. *C has uniform margins, i.e. $C(\boldsymbol{u}) = u_k$ if all coordinates of $\boldsymbol{u}$ except for possibly u_k equal one.*

3. *C is d-increasing.*

We denote the set of all d-copulas by $\mathcal{C}_d$.

Considering that later chapters rely heavily upon 2-copulas, let us shortly restate the previous definition for dimension $d = 2$.

Example 2.1.3. *A function $C : [0,1]^2 \to [0,1]$ is called a 2-copula if it fulfils*

1. *$C(u,0) = 0$ and $C(0,v) = 0$ for all $u, v \in [0,1]$.*

2. *$C(u,1) = u$ and $C(1,v) = v$ for all $u, v \in [0,1]$.*

3. *For every rectangle $R = [u_1, u_2) \times [v_1, v_2)$, it holds that*

$$V_C(R) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0 \ .$$

At first glance, copulas appear to be of purely analytical interest, but Sklar's Theorem stresses the key role they play in the field of dependence modelling and stochastics in general. In the following, $\operatorname{im} F$ denotes the image of F.

Theorem 2.1.4 (Sklar's Theorem)**.** *Let $F_{\boldsymbol{X}}$ be a d-dimensional cumulative distribution function of an $\mathbb{R}^d$-valued random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with univariate margins $F_1, \ldots, F_d$. Then there exists a d-copula C such that for all $\boldsymbol{x} \in \mathbb{R}^d$ it holds*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \ , \tag{2.1}$$

where C is uniquely determined on $\operatorname{im} F_1 \times \cdots \times \operatorname{im} F_d$. Conversely, the expression on the right-hand side of (2.1) constitutes a d-variate cumulative distribution function for any combination of a d-copula C and univariate margins $F_1, \ldots, F_d$.

Theorem 2.1.4 immediately identifies d-copulas as the class of d-variate distribution functions with univariate uniform margins by combining univariate uniform distributions F_i for $1 \leq i \leq d$ with a d-copula C and observing that $F_i(x_i) = x_i$ holds.

Remark 2.1.5. *The normalization of a continuous univariate marginal distribution on $\mathbb{R}$ into a univariate uniform distribution on $[0,1]$ will later on play a crucial role in the comparison of different dependence concepts. Nonetheless, it is decidedly arbitrary as pointed out by Mikosch (2006) and various other problem-specific normalizations have been applied in the literature. For instance, Hoeffding (1940) uses uniformly distributed margins on $[-\frac{1}{2}, \frac{1}{2}]$ and Resnick (1987) presents the uses of Fréchet-marginals in multivariate extreme-value theory. For an extensive discussion, we refer to Embrechts (2009).*

Sklar's Theorem combines two distinct viewpoints relevant to dependence modelling: For practical applications, the second part entails that every combination as given in Equation (2.1) defines a proper cumulative distribution function. It thus provides a simple technique to model the behaviour of real-world phenomena by combining a specific copula (read: *dependence structure*) with arbitrary marginal behaviours. Of course, in a real-world setting, the statistician needs to strike a balance between the 'best possible' dependence structure and keeping the model as concise as possible to facilitate calculations.[1]

This thesis is mainly concerned with the implications of the first part of Sklar's Theorem. If all marginal distributions of the random vector $\boldsymbol{X}$ are continuous, i.e. fulfil $\operatorname{im} F_i = [0,1]$, the induced copula C is *unique* and we write $C_{\boldsymbol{X}} := C$. In this case, $C_{\boldsymbol{X}}$ completely determines any (scale-free) dependence between the components of the random vector. One special case are random vectors with independent components, whose corresponding copula is called the product copula:

Example 2.1.6. *Suppose $\boldsymbol{X}$ is a random vector with continuous univariate margins and d-copula $C_{\boldsymbol{X}}$. Then the components of $\boldsymbol{X}$ are independent if and only if $C_{\boldsymbol{X}} = \Pi$ holds, where Π denotes the independence d-copula (or product d-copula) defined as*

$$\Pi(\boldsymbol{u}) := \prod_{k=1}^{d} u_k \ .$$

Below, we collect some key properties of d-copulas and the set of all d-copulas, which we will rely upon throughout this thesis without explicitly citing the corresponding propositions.

Proposition 2.1.7. *The set $\mathcal{C}_d$ of all d-copulas is convex and compact with respect to uniform convergence. Moreover, pointwise and uniform convergence coincide, i.e. a sequence $(C_n)_{n\in\mathbb{N}}$ in $\mathcal{C}_d$ converges pointwise towards C if and only if it converges uniformly towards C.*

Proposition 2.1.8. *Let C be a d-copula, then the following holds:*

1. *C is bounded from below and from above by the lower and upper Fréchet-Hoeffding bound C^- and C^+, respectively, that is,*

$$C^-(\boldsymbol{u}) \le C(\boldsymbol{u}) \le C^+(\boldsymbol{u})$$

[1] Perhaps the most recent folklore example where this balance was struck inappropriately stems from the 2008 financial crisis, where the Gaussian copula was used to model the dependence between financial assets (see Salmon (2012) and Durante and Sempi (2016) and the references therein for an in-depth review). We will briefly discuss in Section 2.4 why the Gaussian copula is not an appropriate choice concerning the study of tail phenomena often relevant in finance.

holds for all $\boldsymbol{u} \in [0,1]^d$ *with*

$$C^-(\boldsymbol{u}) := \max\left\{\sum_{k=1}^{d} u_k - (d-1), 0\right\} \text{ and } C^+(\boldsymbol{u}) := \min\{u_1, \ldots, u_d\} \ .$$

While C^+ *is a d-copula for all* $d \geq 2$*,* C^- *is only a d-copula in dimension* $d = 2$ *but constitutes a pointwise lower bound even for* $d \geq 3$*.*

2. *C is increasing in each component.*

3. *C is Lipschitz continuous with Lipschitz constant* 1 *with respect to the* ℓ_1*-norm on* $\mathbb{R}^d$*, that is,*

$$|C(\boldsymbol{u}) - C(\boldsymbol{v})| \leq \sum_{k=1}^{d} |u_k - v_k| \ .$$

4. *The partial derivatives* $\partial_k C(\boldsymbol{u})$*,* $k = 1, \ldots, d$*, of C are Borel measurable and bounded. More precisely, they fulfil* $0 \leq \partial_k C(\boldsymbol{u}) \leq 1$*.*

5. *For* $d = 2$*, the functions* $t \mapsto \partial_1 C(u,t)$ *and* $t \mapsto \partial_2 C(t,v)$ *are increasing.*

Note that the partial derivative of a copula C is only defined almost everywhere, and the above theorem has to be read accordingly. For example, Assertion 4 in Proposition 2.1.8 states that $0 \leq \partial_k C(\boldsymbol{u}) \leq 1$ holds for all $u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_d$ in $[0,1]$ and almost all u_k in $[0,1]$ (with respect to the Lebesgue measure). For the remainder of this thesis, we will often suppress this fact in our notation.

2.2 Markov operators and the Markov product

The mapping $t \mapsto \partial_1 C(u,t)$ from Assertion 5 of Proposition 2.1.8 is not solely of analytical interest but contains a fundamental stochastic interpretation.

Proposition 2.2.1. *Let* $\boldsymbol{U} = (U_1, U_2)$ *be a random vector with* 2*-copula* $C_{\boldsymbol{U}}$ *and univariate uniformly distributed margins* U_1 *and* U_2 *on* $[0,1]$*, in short* $\boldsymbol{U} \sim C_{\boldsymbol{U}}$*. Then*

$$\mathbb{P}(U_2 \leq u_2 \mid U_1 = u_1) = \mathbb{E}(\mathbb{1}_{[0,u_2]}(U_2) \mid U_1 = u_1) = \partial_1 C_{\boldsymbol{U}}(u_1, u_2)$$

holds for all $u_2 \in [0,1]$ *and almost all* $u_1 \in [0,1]$*, where* $\mathbb{P}(U_2 \leq u_2 \mid U_1 = u_1)$ *and* $\mathbb{E}(\mathbb{1}_{[0,u_2]}(U_2) \mid U_1 = u_1)$ *denote the conditional probability and conditional expectation given* $U_1 = u_1$*, respectively.*

Consequently, one might wonder how the copula $C_{\boldsymbol{U}}$ can be related to $\mathbb{E}(f(U_2) \mid U_1 = u_1)$ for an arbitrary integrable function f using the usual measure theoretic progression. This generalization leads to the application of Markov operators, a class of linear operators on

$$L^1([0,1]) := L^1([0,1], \mathcal{B}([0,1]), \lambda)$$

closely related to 2-copulas, where $\mathcal{B}([0,1])$ denotes the Borel σ-algebra on $[0,1]$ and λ the Lebesgue measure.

Definition 2.2.2. *A linear operator $T : L^1([0,1]) \to L^1([0,1])$ is called a Markov operator if the following three properties are fulfilled:*

1. *T is positive, that is, $Tf \geq 0$ holds whenever $f \geq 0$.*
2. *T has the fixed point $\mathbb{1}_{[0,1]}$.*
3. *T preserves the integral, i.e. for all $f \in L^1([0,1])$, it holds that*

$$\int_0^1 Tf(t)\ \mathrm{d}t = \int_0^1 f(t)\ \mathrm{d}t\ .$$

Olsen, Darsow and Nguyen (1996) were the first to establish the direct link between 2-copulas and Markov operators, using ideas going back to Ryff (1963).

Theorem 2.2.3. *Let C be a 2-copula and T be a Markov operator. Then*

$$C_T(u,v) := \int_0^u T\mathbb{1}_{[0,v]}(t)\ \mathrm{d}t$$

defines a 2-copula and

$$T_C f(u) := \partial_u \int_0^1 \partial_2 C(u,t) f(t)\ \mathrm{d}t \tag{2.2}$$

for $f \in L^1([0,1])$ defines a Markov operator. This correspondence is one-to-one with $T_{C_T} = T$ and $C_{T_C} = C$ for all 2-copulas C and all Markov operators T.

Remark 2.2.4. *Originally, the correspondence between copulas and Markov operators was given for Markov operators T_C on $L^\infty([0,1])$, where T_C preserves the integral for any $f \in L^\infty([0,1])$. While any such L^∞-Markov operator can be uniquely extended to a Markov operator on $L^1([0,1])$, it remains to show that Equation (2.2) remains valid for integrable but not essentially bounded f. This follows from the representation result for linear operators stated in Theorem 2.3.9 of Dunford and Pettis (1940).*

The operator-theoretic formulation of copulas captures the behaviour of the conditional expectation of $f(U_2)$ given $U_1 = u$ alluded to at the beginning of this section.

Proposition 2.2.5. *Let C be a 2-copula and $\boldsymbol{U} = (U_1, U_2) \sim C$. Then*

$$T_C f(u) = \mathbb{E}(f(U_2) \mid U_1 = u)$$

holds for almost all $u \in [0,1]$ and for all $f \in L^1([0,1])$.

Proof. A proof of this result can be found in Trutschnig (2011). □

Markov operators not only describe the behaviour of the conditional expectation of U_2 given $U_1 = u$, but they also introduce a new view on copula theory. Noting that the composition $\circ$ of Markov operators again results in a Markov operator, one can define the so-called Markov product $*$ (or star-product) for 2-copulas as $C_1 * C_2 := C_{T_{C_1} \circ T_{C_2}}$.

Proposition 2.2.6. *The Markov product of the* 2*-copulas* C_1 *and* C_2 *is given by*

$$(C_1 * C_2)(u_1, u_2) = \int_0^1 \partial_2 C_1(u_1, t) \cdot \partial_1 C_2(t, u_2) \, \mathrm{d}t$$

and is again a 2*-copula. Furthermore, it fulfils* $T_{C_1 * C_2} = T_{C_1} \circ T_{C_2}$*.*

This simple product structure on $\mathcal{C}_2$ facilitates the study of various problems, some of which we will now discuss in more detail. The first is a connection between the algebraic concept of idempotency of a copula C, i.e. the property $C * C = C$, and the characterization of conditional expectations on $L^1([0,1])$.

Proposition 2.2.7. *For a* 2*-copula* C *and its corresponding Markov operator* T_C*, the following assertions are equivalent:*

1. C *is idempotent, i.e.* $C * C = C$*.*
2. T_C *is idempotent, i.e.* $T_C \circ T_C = T_C$*.*
3. T_C *is a conditional expectation restricted to* $L^1([0,1], \mathcal{B}([0,1]), \lambda)$*, i.e.*

$$T_C f = \mathbb{E}(f \mid \mathcal{G})$$

holds for all $f \in L^1([0,1])$*, where* $\mathcal{G} := \{A \in \mathcal{B}([0,1]) \mid T_C \mathbb{1}_A = \mathbb{1}_A\}$*.*

Proof. A proof can be found for the restriction to $L^\infty([0,1])$ in Durante and Sempi (2016) and for the restriction to $L^1([0,1])$ in Albanese and Sempi (2016). □

The next two results emphasize the role of invertible elements. We say C is left-invertible if there exists a 2-copula D such that $D * C = C^+$ holds, and right-invertible if the transposed copula $C^\top$ with $C^\top(u,v) := C(v,u)$ is left-invertible. In fact, the left-inverse fulfils $D = C^\top$, a property reminiscent of the product of matrices. But unlike matrices, a 2-copula can be left-invertible but not right-invertible and vice versa.

Proposition 2.2.8. *Every* 2*-copula* C *can be decomposed (although not uniquely) into a left-invertible copula* L *and a right-invertible copula* R *such that* $C = R * L$ *holds.*

The next result highlights the stochastic significance of the previous algebraic decomposition (see Darsow, Nguyen and Olsen (1992)).

Theorem 2.2.9. *Let* $\boldsymbol{X} = (X_1, X_2)$ *be a random vector with continuous univariate margins and* 2*-copula* $C_{\boldsymbol{X}}$*. Then the following are equivalent:*

1. X_2 *is completely dependent on* X_1*, i.e.* $X_2 = f(X_1)$ *holds almost surely for some measurable function* f*.*
2. $C_{\boldsymbol{X}}$ *is left-invertible, i.e.* $C_{\boldsymbol{X}}^\top * C_{\boldsymbol{X}} = C^+$*.*
3. $\partial_1 C_{\boldsymbol{X}}(u_1, u_2) \in \{0, 1\}$ *for all* $u_2 \in [0,1]$ *and almost all* $u_1 \in [0,1]$*.*
4. $T_{C_{\boldsymbol{X}}} f = f \circ \sigma$ *holds for all* $f \in L^1([0,1])$*, where* $\sigma : [0,1] \to [0,1]$ *is a* λ*-preserving transformation, i.e. fulfils* $\lambda(\sigma^{-1}(A)) = \lambda(A)$ *for all Borel measurable sets* A*.*

A copula fulfilling $\partial_1 C(u_1, u_2) \in \{0,1\}$ for all $u_2 \in [0,1]$ and almost all $u_1 \in [0,1]$ is called completely dependent.

2.3 Measures of dependence

Not surprisingly, Theorem 2.2.9 plays a central role in the construction of measures of complete dependence since the equivalence of the assertions therein allows for a precise detection of complete dependence. Before introducing specific measures of dependence, let us consider some axioms a 'good' or 'useful' measure of complete dependence should fulfil. Such sets of axioms have been proposed in the literature either explicitly by Rényi (1959) and by Schweizer and Wolff (1981) or implicitly by Dabrowska (1981). We essentially combine aspects from Dabrowska (1981) and Schweizer and Wolff (1981) to arrive at the following axioms:

Definition 2.3.1. *A measure of complete dependence is a function* $\mu : \boldsymbol{X} \mapsto \mu(\boldsymbol{X}) \in [0, \infty]$ *defined on the set of bivariate random vectors* $\boldsymbol{X} = (X_1, X_2)$ *with continuous univariate marginal distributions that satisfies the following axioms:*

1. $\mu(X_1, X_2)$ *exists and takes values in* $[0, 1]$.
2. $\mu(X_1, X_2) = 0$ *if and only if* X_1 *and* X_2 *are independent.*
3. $\mu(X_1, X_2) = 1$ *if and only if* $X_2 = f(X_1)$ *holds almost surely for some measurable function* f.
4. $\mu(f(X_1), X_2) = \mu(X_1, X_2)$ *for every measurable bijection* f.
5. $\mu(X_1, g(X_2)) = \mu(X_1, X_2)$ *for every strictly monotone function* g.
6. $\mu(X_1, X_2)$ *is a strictly increasing function of the absolute value of the coefficient of correlation for jointly normal distributed random vectors* (X_1, X_2)*, where the coefficient of correlation is defined as*

$$\mathrm{Corr}\,(X_1, X_2) := \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}} \ .$$

In short, $\mu(X_1, X_2) = f(|\mathrm{Corr}\,(X_1, X_2)|)$ *holds for some strictly increasing function* f.

Our set of axioms differs from the original ones in Rényi (1959) and Schweizer and Wolff (1981) in a few key aspects. Firstly, we view dependence to be a *directed* concept. As an example, consider the water levels of a river system as presented in Bücher, Irresberger and Weiss (2017). The river system consists of a main river and one tributary, such that, along the main river, one water depth gauge (A) is placed before and one gauge (B) is placed after both streams join. Then, raised water levels at (A) lead to raised levels at (B), while raised levels at (B) alone could stem from a flood of the tributary. Consequently, we do not impose symmetry in Definition 2.3.1, i.e. we do not require $\mu(X_1, X_2) = \mu(X_2, X_1)$. Furthermore, any asymmetric measure of complete dependence μ can easily be symmetrized either by setting

$$\overline{\mu}(X_1, X_2) := \frac{\mu(X_1, X_2) + \mu(X_2, X_1)}{2} \ , \tag{2.3}$$

which yields a measure of mutual complete dependence, i.e. $\overline{\mu}(X_1, X_2) = 1$ if and only if $X_2 = f(X_1)$ and $X_1 = g(X_2)$, or by setting

$$\mu_\vee(X_1, X_2) := \max\left\{\mu(X_1, X_2), \mu(X_2, X_1)\right\} \ ,$$

where $\mu_\vee(X_1, X_2)$ equals 1 if and only if either $X_2 = f(X_1)$ or $X_1 = g(X_2)$. In contrast, a symmetric measure cannot be easily turned into an asymmetric measure. Secondly, we require the 'if and only if' formulation in Property 3 of Definition 2.3.1 concerning the detection of complete dependence. This has been mentioned in Rényi (1959) but was ultimately reduced to '$\mu(X_1, X_2) = 1$ if $X_2 = f(X_1)$ holds for some measurable function f' in favour of convenience and a broader applicability.[2]

To the best of our knowledge, the first measure of (mutual) complete dependence fulfilling all but the fourth of the above axioms was introduced by Siburg and Stoimenov (2010). Exploiting the fact that copulas are Lipschitz continuous, they transfer the inner product structure of the Sobolev space onto $\mathcal{C}_2$ (or, more precisely, its linear hull $\operatorname{span}(\mathcal{C}_2)$).

Definition 2.3.2. *Let C_1 and C_2 be 2-copulas. Then*

$$\begin{aligned}\langle C_1\,,\, C_2\rangle_S &:= \int\limits_{[0,1]^2} \nabla C_1(\boldsymbol{u}) \cdot \nabla C_2(\boldsymbol{u})\ \mathrm{d}\lambda(\boldsymbol{u}) \\ &= \int\limits_0^1\int\limits_0^1 \partial_1 C_1(u_1,u_2)\cdot\partial_1 C_2(u_1,u_2) + \partial_2 C_1(u_1,u_2)\cdot\partial_2 C_2(u_1,u_2)\ \mathrm{d}u_1\ \mathrm{d}u_2\end{aligned}$$

defines the so-called Sobolev inner product on $\operatorname{span}(\mathcal{C}_2)$ *with the corresponding Sobolev norm* $\|\, C\,\|_S := \sqrt{\langle C\,,\, C\rangle_S}$.

$\|\cdot\|_S$ establishes a notion of distance on $\mathcal{C}_2$ and can be used to investigate complete dependence.

Theorem 2.3.3. *The Sobolev-norm $\|\cdot\|_S$ induces a measure ω of mutual complete dependence via*

$$\begin{aligned}\omega^2(C) &:= 3\,\|\, C - \Pi\,\|_S^2 = 3\,\|\, C\,\|_S^2 - 2 \\ &= 3\int\limits_0^1 (C^\top * C + C * C^\top)(t,t)\ \mathrm{d}t - 2\ .\end{aligned} \tag{2.4}$$

Proof. See Siburg and Stoimenov (2008a) and Siburg and Stoimenov (2010) for a proof. □

Conceptually, the integrand $C^\top * C + C * C^\top$ appearing in Equation (2.4) stems from Part 2 of Theorem 2.2.9 and quantifies the degree of invertibility of C and $C^\top$. Thus, exploiting the connection between invertibility and complete dependence, ω constitutes a measure of mutual complete dependence.

While Siburg and Stoimenov (2010) follow a geometric approach to quantify mutual complete dependence using ω, Trutschnig (2011) considers an analytic approach utilizing Markov operators and Markov kernels. It is well-known that the uniform convergence on $\mathcal{C}_2$ corresponds to the weak operator topology on the set of Markov operators (see Olsen et al. (1996)). Unfortunately, the L^∞-norm exhibits a major drawback in regard to complete dependence.

[2] Rényi (1959) writes on p.13 'It seems at the first sight natural to postulate that $\mu(X_1, X_2) = 1$ only if there is a strict dependence of the mentioned type between X_1 and X_2, but this condition is rather restrictive, and it is better to leave it out.'

Example 2.3.4. *The independence copula* Π *can be approximated arbitrarily close in the* L^∞*-norm using completely dependent copulas. More precisely, there exists a sequence* $(C_n)_{n\in\mathbb{N}}$ *of completely dependent copulas such that* $(C_n)_{n\in\mathbb{N}}$ *converges uniformly towards* Π*, i.e.*

$$\|C_n - \Pi\|_\infty \to 0$$

holds as n *tends towards infinity.*

This 'paradoxical' behaviour led Trutschnig (2011) to investigate the strong operator topology of Markov operators and metrizations thereof on $\mathcal{C}_2$, resulting in a family of metrics that are essentially L^p-distances of the partial derivatives.

Definition 2.3.5. *Let* C_1 *and* C_2 *be* 2*-copulas. Then*

$$D_p(C_1, C_2) := \Big(\int\limits_{[0,1]^2} |\partial_1 C_1(\boldsymbol{u}) - \partial_1 C_2(\boldsymbol{u})|^p \ \mathrm{d}\lambda(\boldsymbol{u}) \Big)^{\frac{1}{p}}$$

defines a metric on $\mathcal{C}_2$ *for any* $1 \le p < \infty$*. For* $p = \infty$*, the corresponding metric is*

$$D_\infty(C_1, C_2) := \sup_{u_2 \in [0,1]} \int\limits_{[0,1]} |\partial_1 C_1(u_1, u_2) - \partial_1 C_2(u_1, u_2)| \ \mathrm{d}\lambda(u_1) \ .$$

All D_p-metrics are strictly finer than the uniform convergence and guarantee the closure of the set of completely dependent copulas, hence circumventing the seemingly paradoxical behaviour of Example 2.3.4.

Proposition 2.3.6. *Convergence with respect to the* D_p*-metrics implies convergence with respect to the* d_∞*-metric on* $\mathcal{C}_2$*, that is,*

$$D_p(C_n, C) \to 0 \implies C_n \to C \ \textit{uniformly}$$

as n *tends towards infinity. Furthermore, the set of all completely dependent copulas is closed with respect to* D_p *for* $1 \le p \le \infty$*.*

Proof. See Trutschnig (2011) for a proof. □

Similar to the Sobolev norm in Theorem 2.3.3, the distance between C and Π with respect to the D_p-metric again induces a measure of complete dependence.

Theorem 2.3.7. *The* D_p*-metric for* $1 \le p < \infty$ *induces a measure* ζ_p *of complete dependence via*

$$\zeta_p(C) := \frac{D_p(C, \Pi)}{D_p(C^+, \Pi)} = \left(\frac{(p+1)(p+2)}{2} \right)^{1/p} D_p(C, \Pi) \ ,$$

where C *is a* 2*-copula and* Π *is the independence* 2*-copula.*

Proof. The cases $p = 1$ and $p = 2$ are treated in Trutschnig (2011) and Dette et al. (2013), respectively. An outline for a general $p \ge 1$ can be found in the unpublished article Li (2015). □

Despite the seemingly very different approaches, ω^2 and ζ_2 are closely linked via

$$\omega^2(C) = 3D_2^2(C, \Pi) + 3D_2^2(C^\top, \Pi) = \frac{\zeta_2^2(C) + \zeta_2^2(C^\top)}{2} ,$$

which is the aforementioned symmetrization stated in Equation (2.3). Independently, the directed measure ζ_2^2 was investigated by Dette et al. (2013) to create an asymmetric version of ω^2, called a measure of regression dependence

$$r(C) := 6 \int_{[0,1]^2} |\partial_1 C(u_1, u_2) - u_2|^2 \, \mathrm{d}\lambda(\boldsymbol{u}) = \zeta_2^2(C) . \tag{2.5}$$

Most notably, Dette et al. (2013) introduced an underlying *order of regression dependence* for random vectors based on univariate variability orders for the conditional distribution functions. This order[3] provides the means to discuss the degree of complete (or regression) dependence, allowing the comparison of intermediate values of r in $(0, 1)$.

2.3.1 Measures of concordance

Though complete dependence has gained much attention in recent years, one of the dependence concepts most commonly used is that of concordance. Intuitively, a random vector $\boldsymbol{Y}$ is more concordant than $\boldsymbol{X}$ if the components of $\boldsymbol{Y}$ have a higher probability than the components of $\boldsymbol{X}$ to simultaneously attain small or large values (see Joe (1990)).

Definition 2.3.8. *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two bivariate random vectors. We say $\boldsymbol{X}$ is smaller than $\boldsymbol{Y}$ with respect to the concordance ordering, in short $\boldsymbol{X} \leq_c \boldsymbol{Y}$, if*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \leq F_{\boldsymbol{Y}}(\boldsymbol{x}) \text{ and } \overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) \leq \overline{F}_{\boldsymbol{Y}}(\boldsymbol{x})$$

holds for all $\boldsymbol{x} \in \mathbb{R}^2$, where $\overline{F}$ denotes the survival function of F.

The concordance ordering for bivariate random vectors with continuous and identical margins $F_{X_i} = F_{Y_i}$, $i = 1, 2$, coincides with the pointwise ordering of copulas, i.e.

$$\boldsymbol{X} \leq_c \boldsymbol{Y} \Leftrightarrow C_{\boldsymbol{X}} \leq C_{\boldsymbol{Y}} .$$

This equivalence does not hold for d-variate random vectors with a dimension d strictly larger than 2.

Similar to the set of axioms postulated by Rényi (1959), an axiomatic approach to measures of concordance was established by Scarsini (1984):

Definition 2.3.9. *A mapping $\kappa : \mathcal{C}_2 \to [-1, 1]$ is called a measure of concordance if*

1. $\kappa(C^-) = -1 \leq \kappa(C) \leq 1 = \kappa(C^+)$ *holds for all* 2*-copulas* C.

2. $\kappa(\Pi) = 0$.

[3] While technically only a preorder, it is called an order in the context of stochastic orderings (see Shaked and Shanthikumar (2007)).

3. $\kappa(C_1) \leq \kappa(C_2)$ *whenever* $C_1 \leq C_2$ *holds for the* 2*-copulas* C_1 *and* C_2.

4. $\kappa(C^\top) = \kappa(C)$ *holds for all* 2*-copulas* C.

5. $\kappa(C^- * C) = \kappa(C * C^-) = -\kappa(C)$, *where* $*$ *denotes the Markov product.*

6. $\kappa(C_n) \to \kappa(C)$ *whenever* C_n *converges pointwise towards* C.

For a random vector $\boldsymbol{X}$ *with continuous univariate margins, we set* $\kappa(\boldsymbol{X}) := \kappa(C_{\boldsymbol{X}})$.

There is a wide range of concordance measures, among which the best-known measures are Spearman's ρ and Kendall's τ.

Theorem 2.3.10. *Spearman's* ρ *and Kendall's* τ, *defined as*

$$\rho(C) := 12 \int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}\lambda(\boldsymbol{u}) - 3 \text{ and } \tau(C) := 4 \int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}) - 1 \ ,$$

respectively, are measures of concordance.

Let us briefly emphasize the importance of the underlying concordance ordering when using Spearman's ρ and Kendall's τ.

Example 2.3.11. *To show that the concordance ordering cannot be dispensed with, we construct two copulas* C_1 *and* C_2 *which fulfil neither* $C_1 \leq C_2$ *nor* $C_2 \leq C_1$ *and result in reversely ordered measures of concordance, that is,*

$$\tau(C_2) < \tau(C_1) \text{ while } \rho(C_1) < \rho(C_2) \ .$$

We consider the diagonal 2*-copula (see also Definition 3.4.2)*

$$C_1(u, v) := \min\left\{u, v, \frac{u^2 + v^2}{2}\right\}$$

as well as the tent copula C_θ *with* $\theta \in [0, 1]$ *(see Example 6.1.5), whose support consists of the two line segments from* $(0, 0)$ *to* $(\theta, 1)$ *and from* $(\theta, 1)$ *to* $(1, 0)$. *Then it holds for* $C_2 := C_\theta$ *with* $\theta = 0.65$ *(see Examples 5.6 and 5.14 in Nelsen (2006)) that*

$$\tau(C_2) = 2\theta - 1 = 0.3 < \frac{1}{3} = \tau(C_1)$$

as well as

$$\rho(C_1) = 5 - \frac{3\pi}{2} \approx 0.288 < 0.3 = 2\theta - 1 = \rho(C_2) \ .$$

Thus, dismissing the underlying concordance order $\leq_c$ *and simply ordering via some measure of concordance* can *result in contradictory notions of concordance.*

2.4 Tail dependence

Fundamentally different from the previous dependence concepts governing the global behaviour of two random variables, the tail dependence quantifies extremal dependence between multiple random variables. Introduced by Sibuya (1960) and further discussed by Joe (1997), the tail dependence coefficient of a random vector $\boldsymbol{X}$ is defined as

$$\lambda(\boldsymbol{X}) := \lim_{s \searrow 0} \mathbb{P}(X_2 \leq F_2^{-1}(s), \ldots, X_d \leq F_d^{-1}(s) \mid X_1 \leq F_1^{-1}(s)) \ , \tag{2.6}$$

whenever the limit exists. It captures the extremal behaviour of the whole random vector in case one component attains extremely low values.

Despite its wide-spread use, the tail dependence coefficient as presented in Equation (2.6) has some severe drawbacks, which we shall consider more carefully now. Firstly, the influence of the univariate marginal distributions F_i of X_i on $\lambda(\boldsymbol{X})$ is, at first glance, unclear. A short calculation remedies this problem and yields

$$\lambda(\boldsymbol{X}) = \lim_{s \searrow 0} \frac{C_{\boldsymbol{X}}(s, \ldots, s)}{s} = \lim_{s \searrow 0} \frac{C_{\boldsymbol{X}}(s\mathbf{1})}{s} \ , \tag{2.7}$$

where $C_{\boldsymbol{X}}$ denotes the unique copula of $\boldsymbol{X}$ and $\mathbf{1} = (1, \ldots, 1)$. The tail dependence coefficient $\lambda(\boldsymbol{X})$, or $\lambda(C_{\boldsymbol{X}})$, is therefore a margin-free measure of extremal dependence and, in particular, independent of the preferred units. However, Equation (2.7) also shows that $\lambda(\boldsymbol{X})$ is solely affected by the behaviour of the copula along the diagonal, thus ignoring any additional information contained in a neighbourhood around zero. This suggests the following natural generalization of λ.

Definition 2.4.1. *For a d-copula C, the (lower) tail dependence function $\Lambda : \mathbb{R}_+^d \to \mathbb{R}_+$ is defined as*

$$\Lambda(\boldsymbol{w}) := \Lambda(\boldsymbol{w}\,; C) := \lim_{s \searrow 0} \frac{C(s\boldsymbol{w})}{s} \ ,$$

provided that the limit exists for all $\boldsymbol{w} \in \mathbb{R}_+^d$. We denote the set of all d-variate tail dependence functions by $\mathcal{T}_d$.

The next theorem reveals how the tail dependence function influences the behaviour of C around zero.

Theorem 2.4.2. *Suppose C is a d-copula and $L : \mathbb{R}_+^d \to \mathbb{R}_+$ is positive homogeneous of order 1, i.e. $L(s\boldsymbol{w}) = sL(\boldsymbol{w})$ holds for all $s > 0$ and $\boldsymbol{w} \in \mathbb{R}_+^d$. Then the following are equivalent:*

1. *The tail dependence function $\Lambda(\boldsymbol{w}\,; C)$ exists for all $\boldsymbol{w} \in \mathbb{R}_+^d$ and equals $L(\boldsymbol{w})$.*
2. *L is the leading part of a uniform lower tail expansion of C, i.e.*

$$C(\boldsymbol{u}) = L(\boldsymbol{u}) + R(\boldsymbol{u})\, \|\boldsymbol{u}\|_1 = L(\boldsymbol{u}) + R(\boldsymbol{u})\,(u_1 + \ldots + u_d) \ ,$$

where $R : [0,1]^d \to \mathbb{R}$ is a bounded function fulfilling $R(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\|_1 \to 0$.

In addition, the tail dependence function Λ is concave.

Proof. A proof can be found in Jaworski (2006) or Jaworski (2010).[4] □

Theorem 2.4.2 shows that the tail dependence function $\Lambda(\,\cdot\,; C)$ is the directional derivative of C in zero such that the remainder term $C(\boldsymbol{u}) - L(\boldsymbol{u})$ has a certain asymptotic behaviour. Furthermore, C is (Fréchet-) differentiable in 0 if and only if $\Lambda(\boldsymbol{w}; C) = 0$ holds for all $\boldsymbol{w} \in \mathbb{R}_+^d$. Essentially, any continuity in the directional derivative of C within a neighbourhood of zero requires a tail independent copula C, an observation also found in Example 2.6.11 of Durante and Sempi (2016) and underlined by the next example.

Example 2.4.3. *Suppose $\boldsymbol{X} = (X_1, X_2)$ is a jointly normal distributed random vector with correlation coefficient* $\operatorname{Corr}(X_1, X_2) \in [-1, 1]$*. Then $\boldsymbol{X}$ is tail independent, i.e. fulfils $\lambda(\boldsymbol{X}) = 0$, for* $\operatorname{Corr}(X_1, X_2) < 1$*, and tail dependent with $\lambda(\boldsymbol{X}) = 1$ for* $\operatorname{Corr}(X_1, X_2) = 1$*. This makes jointly normal distributed random vectors an often unsuitable and even dangerous model in the context of financial assets since extreme simultaneous losses are severely underestimated.*

As $\Lambda(\,\cdot\,; C)$ is a local approximation of C around zero, the usual d_∞-metric is unable to discriminate between different tail behaviours of copulas. Let us illustrate this point using the patchwork technique described in Durante, Fernández Sánchez and Sempi (2013). For any given d-variate tail dependence function Λ, there exists a family of copulas with tail dependence function Λ, which is dense in the set of all d-copulas w.r.t. the uniform topology. This also implies that the class of all tail dependent d-copulas is dense in $\mathcal{C}_d$. On the other hand, the class of d-copulas which do not allow for a tail dependence function is also dense in $\mathcal{C}_d$ w.r.t. the uniform topology.

Before investigating how the tail dependence function can be used to quantify the extremal behaviour of the underlying random vector, we briefly present some theoretical properties of Λ. Many of these properties are directly linked to the properties of the corresponding copula C and the fact that Λ is the directional derivative of C in $\mathbf{0}$.

Proposition 2.4.4. *A function $\Lambda : \mathbb{R}_+^d \to \mathbb{R}_+$ is the tail dependence function of a d-copula C if and only if*

a. Λ is bounded from below by 0 and from above by $\Lambda^+ := \Lambda(\,\cdot\,; C^+)$, and

b. Λ is d-increasing, and

c. Λ is positive homogeneous of order 1.

Furthermore, for any tail dependence function Λ, we have

1. Λ is Lipschitz continuous. More precisely, it holds for all $\boldsymbol{v}$ and $\boldsymbol{w} \in \mathbb{R}_+^d$

$$|\Lambda(\boldsymbol{v}) - \Lambda(\boldsymbol{w})| \le \sum_{k=1}^{d} |v_k - w_k| \;.$$

2. $w_1 \mapsto \partial_1 \Lambda(w_1, w_2, \ldots, w_d)$ is decreasing for almost all $w_1 \in \mathbb{R}_+$ and all $w_2, \ldots, w_d \in \mathbb{R}_+$.

[4] While, to the best of our knowledge, not discussed in the literature, Theorem 2.4.2 can be extended to hold for quasi-d-copulas. The first part relies on the theory of B-differentiability discussed in Scholtes (2012), while the concavity follows from Theorem 2.10 in König (2003) whenever the quasi-d-copula is supermodular in a neighbourhood of zero.

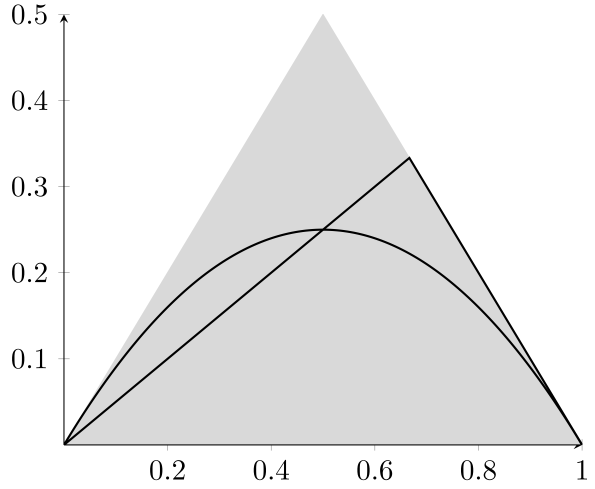


Figure 2.1: Example of two valid tail dependence functions (depicted in black) lying in the area between 0 and $\min\{t, 1-t\}$ (depicted in grey).

Finally, for $d = 2$, $w_2 \mapsto \partial_1 \Lambda(w_1, w_2)$ is increasing for almost all $w_1 \in \mathbb{R}_+$ and all $w_2 \in \mathbb{R}_+$.

Proof. For a proof, see again Jaworski (2006) or Jaworski (2010). □

As is the case for d-copulas, the partial derivatives of d-variate tail dependence functions are only defined almost everywhere. We will often suppress this fact in our notation.

In dimension $d = 2$, a characterization of tail dependence functions simpler than the one provided in Proposition 2.4.4 is known.

Example 2.4.5. *Due to the positive homogeneity of the tail dependence function Λ, we can reduce the domain of Λ from $\mathbb{R}_+^d$ to the simplex $\mathcal{S}^{d-1} := \left\{ \boldsymbol{w} \in \mathbb{R}_+^d \mid \|\boldsymbol{w}\|_1 = 1 \right\}$ with no loss of information. Most notably, for dimension $d = 2$, this results in a univariate function $\widetilde{\Lambda} : [0,1] \to [0, 1/2]$ such that*

$$\begin{aligned} \Lambda(\boldsymbol{w}) &= \|\boldsymbol{w}\|_1 \Lambda\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_1}\right) \\ &= (w_1 + w_2)\Lambda\left(\frac{w_1}{w_1 + w_2}, 1 - \frac{w_1}{w_1 + w_2}\right) \\ &= (w_1 + w_2)\widetilde{\Lambda}\left(\frac{w_1}{w_1 + w_2}\right) \end{aligned} \tag{2.8}$$

holds. From Einmahl, Krajina and Segers (2008) and Gudendorf and Segers (2010), it then follows that Λ is a bivariate tail dependence function if and only if $\widetilde{\Lambda} : [0,1] \to [0, 1/2]$ is a concave function fulfilling $0 \le \widetilde{\Lambda}(t) \le \min\{t, 1-t\}$. In short, any concave function lying in the grey area depicted in Figure 2.1 induces a valid bivariate tail dependence function via Equation (2.8)*.*

As the reduced function $\widetilde{\Lambda}$ contains all necessary information of Λ, we will oftentimes use $\widetilde{\Lambda}$ instead of the actual tail dependence function Λ. The next remark briefly establishes the corresponding notation in the bivariate case for later reference.

Remark 2.4.6. *Due to the positive homogeneity of Λ, in the bivariate case, we will often only consider Λ on the unit simplex $\mathcal{S}^1 := \left\{ \boldsymbol{w} \in \mathbb{R}^2_+ \mid \boldsymbol{w} = (t, 1-t) \text{ with } t \geq 0 \right\}$ and identify $\mathcal{S}^1$ with $[0,1]$ such that*

$$\widetilde{\Lambda}(t) := \Lambda(t, 1-t) \ .$$

2.5 Constructing families of copulas

While we have extensively discussed the uses of copulas in dependence modelling, we are still lacking examples of copula families capturing specific types of dependence. Thus, we now present some common copula families, which we will use both as illustrative examples and for theoretical applications.

2.5.1 Extreme-value copulas

Although Proposition 2.4.4 provides a theoretical existence result for copulas with a given tail dependence function, we have yet to construct such copulas. To do so, we introduce the (lower) extreme-value copulas.

Proposition 2.5.1. *Suppose Λ is a bivariate tail dependence function. Then*

$$C^{EV}(u_1, u_2\,; \Lambda) := \exp\left(\log(u_1) + \log(u_2) + \Lambda(-\log(u_1), -\log(u_2))\right) \tag{2.9}$$

is a 2-copula called extreme-value copula. The survival copula of C^{EV}, defined as

$$C^{LEV}(u_1, u_2\,; \Lambda) := \widehat{C^{EV}}(u_1, u_2\,; \Lambda) := u_1 + u_2 - 1 + C^{EV}(1-u_1, 1-u_2) \ ,$$

is called lower extreme-value copula.

The next proposition states that the lower extreme-value copula indeed exhibits the prescribed tail behaviour.

Proposition 2.5.2. *Suppose Λ is a bivariate tail dependence function and C^{LEV} its corresponding lower extreme-value copula. Then it holds for all $\boldsymbol{w} \in \mathbb{R}^d_+$ that*

$$\Lambda\left(\boldsymbol{w}\,; C^{LEV}\right) = \Lambda(\boldsymbol{w}) \ .$$

Remark 2.5.3. *While the construction of a d-copula with a given tail dependence function is quite straightforward in dimension $d = 2$, the general construction for $d > 2$ is more involved and can be found in the proof of Proposition 6 of Jaworski (2006).*

2.5.2 Archimedean copulas

The next class of copulas we consider are the so-called Archimedean copulas, whose multivariate behaviour is again determined by a single (although this time, univariate) generator function.

Definition 2.5.4. *We call a function $\phi : [0,1] \to \overline{\mathbb{R}}_+ := [0, \infty]$ an Archimedean generator if it is continuous, strictly decreasing and fulfils $\phi(1) = 0$. Furthermore, we call ϕ strict if*

$$\lim_{s \searrow 0} \phi(s) = \infty \ .$$

Definition 2.5.5. *A d-copula C is called Archimedean if there exists an Archimedean generator ϕ with*

$$C(\boldsymbol{u}) = \phi^{[-1]}\left(\sum_{k=1}^{d} \phi(u_k)\right) \tag{2.10}$$

for $\boldsymbol{u} \in [0,1]^d$, where $\phi^{[-1]}(x) := \inf\{t \in [0,1] \mid \phi(t) \leq x\}$ denotes the generalized inverse of ϕ.

Remark 2.5.6. *There exist conflicting definitions of Archimedean copulas in the literature, either using Equation* (2.10) *or the converse*

$$C(\boldsymbol{u}) = \phi\left(\sum_{k=1}^{d} \phi^{[-1]}(u_k)\right)$$

with an appropriately modified definition of ϕ. We have adopted the version stated in Equation (2.10) *to simplify some of our later results and proofs.*

A necessary and sufficient condition on ϕ to be the generator of an Archimedean d-copula was developed by McNeil and Nešlehová (2009) and utilizes the concept of d-monotonicity.

Definition 2.5.7. *A function $f : (0,\infty) \to \mathbb{R}$ is called d-monotone if for all $0 \leq k \leq d-2$*

$$(-1)^k f^{(k)}(x) \geq 0$$

holds for all $x \in (0,\infty)$ and $(-1)^{d-2} f^{(d-2)}$ is decreasing and convex, where $f^{(k)}$ denotes the k-th derivative of f

Theorem 2.5.8. *Let ϕ be an Archimedean generator. Then C from Equation* (2.10) *is an Archimedean d-copula with generator ϕ if and only if $\phi^{[-1]}$ is d-monotone.*

Example 2.5.9. *In dimension $d = 2$, Theorem 2.5.8 states that C is a 2-copula if and only if $\phi^{[-1]}$ is convex. Unfortunately, in higher dimensions, this characterization becomes increasingly complex. For example, in dimension $d = 3$, C is a 3-copula with Archimedean generator ϕ if and only if $\phi^{[-1]}$ is twice-differentiable, decreasing and $\phi^{[-1]\prime}$ is decreasing and convex.*

Archimedean copulas have become a popular tool both for theoretical as well as practical considerations. Their simple form allows for the explicit calculation of many dependence measures, thus enabling a precise understanding of their induced dependencies. Additionally, there exist fast and reliable methods to sample from Archimedean copulas or to fit them to a given data set. This simplicity is also their biggest drawback for practical applications since they only allow for a rather limited range of possible dependencies, all of which are necessarily exchangeable. For higher dimensional models, this problem is alleviated by using Archimedean copulas mainly as building blocks, for example in the setting of pair-copula constructions (see Joe, Li and Nikoloulopoulos (2010) and Czado (2019)) or nested Archimedean copulas (see McNeil (2008)). Empirical studies have suggested that many pairs of stocks can be modelled quite successfully by Archimedean copulas (see, e.g., Bücher, Dette and Volgushev (2012) and Trede and Savu (2013)). This immediately leads to the question of how to calculate the tail dependence of Archimedean copulas in order to quantify the probability of large simultaneous losses.

To obtain a closed-form expression for the tail dependence of an Archimedean copula, the Archimedean generator ϕ is often assumed to be regularly varying. This condition poses in practice only a slight restriction since it is fulfilled by virtually all relevant Archimedean generators as demonstrated by Charpentier and Segers (2009). For an extensive treatment of regularly varying functions and related topics, we refer to Bingham, Goldie and Teugels (1987).

Definition 2.5.10. *A positive measurable function f on $\mathbb{R}_+$ is called regularly varying at ∞ with index $\alpha \in \mathbb{R}$ if*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha} \tag{2.11}$$

holds for all $t \in \mathbb{R}_+$. f is said to be slowly varying at ∞ if α equals 0, and rapidly varying at ∞ (i.e. regularly varying with parameter $\alpha = \infty$) if

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = \begin{cases} 0 & t < 1 \\ 1 & t = 1 \\ \infty & t > 1 \end{cases}$$

When no ambiguity is possible, we will call all slowly, regularly and rapidly varying functions regularly varying functions with $\alpha \in \overline{\mathbb{R}}$. f is regularly varying at zero with parameter $\alpha \in \overline{\mathbb{R}}$ if $x \mapsto f\left(\frac{1}{x}\right)$ is regularly varying at infinity with parameter $-\alpha$.

Example 2.5.11. *Slowly varying functions include (iterated) logarithms or functions possessing a limit at infinity. Furthermore, any product, sum and power of slowly varying functions is again slowly varying.*

Example 2.5.12. *Any polynomial x^{α} is regularly varying with parameter α. Furthermore, for any regularly varying function f, Equation (2.11) yields that $\ell(x) := x^{-\alpha} f(x)$ is slowly varying. Thus, f is regularly varying with parameter $\alpha \in \mathbb{R}$ if and only if $f(x) = x^{\alpha}\ell(x)$ holds for some slowly varying function ℓ. Examples of functions which are not regularly varying include oscillating functions such as $\sin(x)$.*

With these examples in mind to illustrate the behaviour of regularly varying functions, we now state the closed-form expression for the tail behaviour of Archimedean copulas.

Lemma 2.5.13. *Let C be an Archimedean d-copula with Archimedean generator ϕ. If ϕ is regularly varying at 0 with parameter $-\alpha$, then*

$$\Lambda\left(\boldsymbol{w}\,; C\right) = \begin{cases} 0 & \text{if } \alpha = 0 \\ \left(\sum\limits_{k=1}^{d} w_k^{-\alpha}\right)^{-\frac{1}{\alpha}} & \text{if } \alpha \in (0,\infty) \\ \min\limits_{k=1,\dots,d} w_k & \text{if } \alpha = \infty \end{cases}$$

holds for all $\boldsymbol{w} \in \mathbb{R}_+^d$.

2.5.3 Checkerboard copulas

After introducing extreme-value and Archimedean copulas, both of which rely on lower dimensional functions to generate more complex behaviour, we will now consider a construction that exploits the underlying structure of the set of 2-copulas. To do so, we investigate the relation between copulas and doubly stochastic matrices. Viewing copulas as a continuous analogue of the latter, we provide a construction method for copulas from doubly stochastic matrices. But first, we investigate the converse relation, that is, how to discretize a copula into a doubly stochastic matrix.

Example 2.5.14. *For any dimension $n \in \mathbb{N}$ and any 2-copula C, the induced volume (see Definition 2.1.1)*

$$(A_n)_{k\ell} := n \cdot V_C\left(\left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{\ell-1}{n}, \frac{\ell}{n}\right]\right)$$

defines a doubly stochastic matrix $A_n \in \mathbb{R}^{n \times n}$, i.e. a matrix consisting of nonnegative entries for which all row and column sums are equal to one.

Conversely, to associate a doubly stochastic matrix with a 2-copula, we need to interpolate the discrete values continuously on $[0,1]^2$. This can be done using so-called partitions of unity (see Li, Mikusiński, Sherwood and Taylor (1997) for an extensive introduction). Here, we apply the partition of unity $\phi_k(s) := \mathbb{1}_{\left[\frac{k-1}{n}, \frac{k}{n}\right)}(s)$, $k = 1, \dots, n$, resulting in an absolutely continuous copula. In turn, partitions $\{\phi_k\}_{k=1,\dots,n}$ of higher regularity yield smoother copulas, an example being the Bernstein polynomials leading to the Bernstein copulas with second order continuous partial derivatives (see Durante and Sempi (2016)).

Definition 2.5.15. *Suppose $A = (a_{k\ell})_{k,\ell=1,\dots,n} \in \mathbb{R}^{n \times n}$ is a doubly stochastic matrix. Then*

$$C_n^{\#}(A)(u_1, u_2) := n \sum_{k,\ell=1}^{n} a_{k\ell} \int_0^{u_1} \mathbb{1}_{\left[\frac{k-1}{n}, \frac{k}{n}\right)}(s)\ \mathrm{d}s \int_0^{u_2} \mathbb{1}_{\left[\frac{\ell-1}{n}, \frac{\ell}{n}\right)}(t)\ \mathrm{d}t$$

is a 2-copula called the checkerboard copula of A.

Combining Example 2.5.14 and Definition 2.5.15, we can smooth any possibly singular behaviour of the original copula C and generate an absolutely continuous checkerboard copula that is, in some sense, close to C.

Definition 2.5.16. *For a 2-copula C, the induced checkerboard copula is defined as*

$$C_n^{\#}(C) := C_n^{\#}(A_n)$$

for the doubly stochastic matrix A_n with

$$(A_n)_{k\ell} := n \cdot V_C\left(\left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{\ell-1}{n}, \frac{\ell}{n}\right]\right) .$$

Example 2.5.17. *Consider the upper Fréchet-Hoeffding bound C^+ and its induced measure μ_{C^+} (see Definition 2.1.1). While the support of μ_{C^+} is evenly distributed on the diagonal from $(0,0)$ to $(1,1)$, the associated checkerboard copulas are supported on the squares $\left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{k-1}{n}, \frac{k}{n}\right]$ for $k = 1, \dots, n$. Figure 2.2 illustrates the aforementioned smoothing property and hints at the approximation properties for increasing resolutions n.*

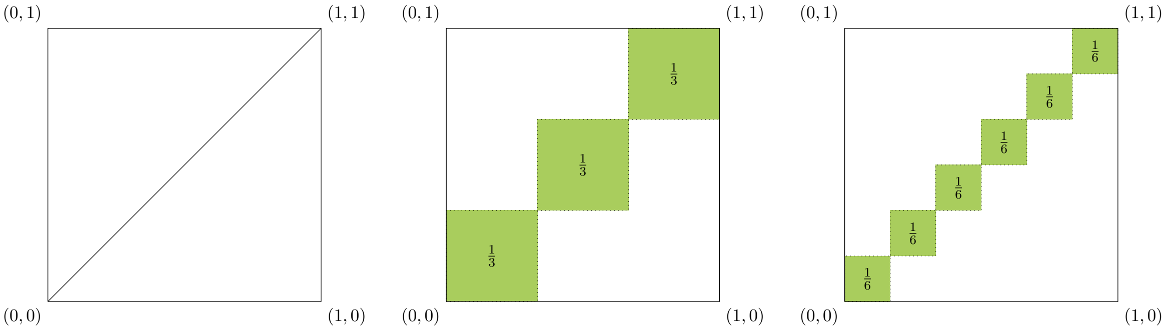


Figure 2.2: Plots depicting the support of the measures of C^+, $C_3^\#(C^+)$ and $C_6^\#(C^+)$, respectively. The numbers denote the respective mass evenly distributed in each square.

In fact, every copula can be approximated by its induced checkerboard copulas with respect to various topologies, a result especially useful in our discussion of complete dependence.

Proposition 2.5.18. *Suppose C is a 2-copula. Then*

$$D_p(C_n^\#(C), C) \to 0$$

as n tends to infinity, where D_p for $1 \le p < \infty$ denotes the metric introduced in Definition 2.3.5.

Thus, the absolutely continuous checkerboard copulas are dense in the set of all 2-copulas $\mathcal{C}_2$ with respect to various modes of convergence, including uniform convergence and D_p-convergence.

2.6 Rearrangement and majorization of functions

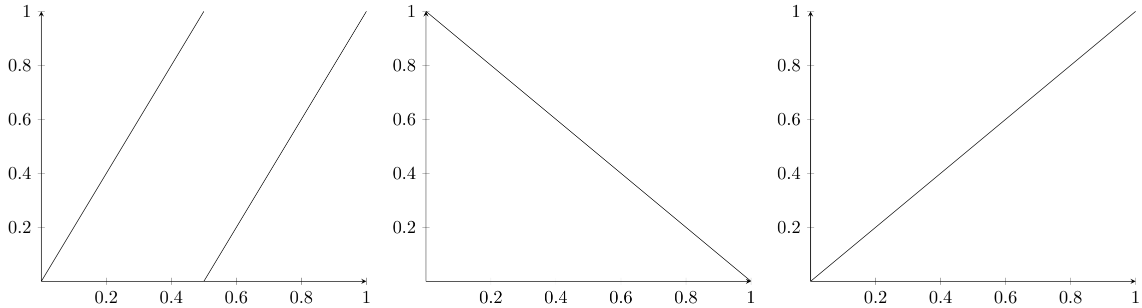
In Theorem 2.2.9, we have seen that a copula C is completely dependent if and only if

$$\partial_1 C(u_1, u_2) = T_C \mathbb{1}_{[0,u_2]}(u_1) = \partial_1 C^+(\sigma(u_1), u_2)$$

holds for some λ-preserving function $\sigma : [0,1] \to [0,1]$. Therefore, for fixed $u_2 \in [0,1]$, the exact behaviour of $\partial_1 C(u_1, u_2)$ at certain points $u_1 \in [0,1]$ is less important than its 'similarity' to $\partial_1 C^+(u_1, u_2)$. One approach to render this 'similarity' mathematically precise is the concept of the (decreasing) rearrangement of functions. Put plainly, two Borel measurable functions f and g (for simplicity both defined on $[0,1]$) are called rearrangements of each other if their distribution functions

$$\lambda(\{f > t\}) = \lambda(\{g > t\})$$

are equal for all $t \in \mathbb{R}$. The decreasing rearrangement f^* of f is then simply the unique right-continuous and decreasing function that is a rearrangement of f. Most importantly, while the decreasing rearrangement reduces the variation of the original function, it preserves the L^p-norms. This property will allow us later on to rearrange the partial derivatives of copulas while leaving measures of complete dependence based on L^p-norms, such as r and ζ_p, invariant. We now briefly collect all necessary notation and concepts, while we refer to Chong and Rice (1971) and Bennett and Sharpley (1988) for a thorough overview.



(a) f (b) f^* and μ_f (c) f_*

Figure 2.3: Plot of $f(x) = 2x \bmod 1$, its decreasing rearrangement f^* and its increasing rearrangement f_*.

Definition 2.6.1. *Let f be a real-valued Borel measurable function on $[0,1]$ and let λ denote the Lebesgue-measure. The decreasing rearrangement f^* of f is defined as*

$$f^*(x) := \inf\{t \mid \mu_f(t) \leq x\}$$

for $x \in [0,1]$, where $\mu_f(t) := \lambda(\{x \in [0,1] \mid f(x) > t\})$ is the distribution function of f. Analogously, the increasing rearrangement f_* of f is given by

$$f_*(x) := f^*(1-x) \ .$$

Example 2.6.2. *To illustrate the concept of a rearrangement, consider the λ-preserving transformation $f(x) = 2x \bmod 1$ on $[0,1]$. A straightforward calculation yields*

$$\begin{aligned} \mu_f(t) &= \lambda(\{x \in [0,1] \mid 2x \bmod 1 > t\}) \\ &= \lambda\left(\left(\frac{t}{2}, \frac{1}{2}\right) \cup \left(\frac{t+1}{2}, 1\right)\right) = 1 - t \end{aligned}$$

and $f^*(x) = 1 - x$. Figure 2.3 depicts the plots of f, f^* and f_*.

The literature contains various definitions of decreasing rearrangements, ranging from symmetric to asymmetric and from univariate to multivariate ones. We follow the asymmetric and univariate approach presented in Ryff (1970), Chong and Rice (1971) and Day (1972) and, due to a lack of one comprehensive introduction, provide separate references to each of the results and their respective proofs. Take special note of the definition of μ_f in Definition 2.6.1 since in some references, the f therein is replaced by $|f|$, in which case the corresponding 'rearrangement' f^* is a genuine rearrangement for nonnegative functions f only.

Proposition 2.6.3. *Let f and g be real-valued Borel measurable functions on $[0,1]$. Then the following assertions hold:*

1. *f^* is decreasing and right-continuous on $[0,1]$.*
2. *$f \leq g$ implies $f^* \leq g^*$.*

3. *The decreasing rearrangement is L^p-invariant, i.e.* $\|f\|_p = \|f^*\|_p$ *for* $1 \le p \le \infty$.

4. *There exists a λ-preserving transformation* $\sigma : ([0,1], \mathcal{B}([0,1])) \to ([0,1], \mathcal{B}([0,1]))$ *such that* $f = f^* \circ \sigma$ *holds almost everywhere.*

Proof. Property 1 is stated in Theorem 4.2, Properties 2 and 3 can be found in Proposition 4.3, and Property 4 is stated in Theorem 6.2 of Chong and Rice (1971). □

Remark 2.6.4. *Using the adjoint operator T'_σ of $T_\sigma f := f \circ \sigma$ for some λ-preserving transformation σ, we can 'invert' the relation given in Part 4 of Proposition 2.6.3 such that $f^* = T'_\sigma f$ holds. Note, however, that this does not imply the existence of a λ-preserving transformation $\tilde{\sigma}$ fulfilling $f^* = f \circ \tilde{\sigma}$.*

One origin of majorization theory is the investigation of income inequalities, but today it has applications in many branches of mathematics, physics and economics. It was discussed by Hardy et al. (1952) for vectors and generalized to functions by Ryff (1965). We refer to Bennett and Sharpley (1988) and Marshall, Olkin and Arnold (2011) for a comprehensive treatment of majorization.

Definition 2.6.5. *Let f and $g \in L^1([0,1])$. Then f is majorized by g, denoted by $f \preceq g$, if*

$$\int_0^t f^*(s)\, \mathrm{d}s \le \int_0^t g^*(s)\, \mathrm{d}s$$

holds for all $t \in [0,1]$, as well as

$$\int_0^1 f^*(s)\, \mathrm{d}s = \int_0^1 g^*(s)\, \mathrm{d}s \; .$$

The above definition of the majorization order $\preceq$ relies on the following idea: Whenever f and g are decreasing functions with the same mass, f is, intuitively, less concentrated than g if

$$\int_0^t f(s)\, \mathrm{d}s \le \int_0^t g(s)\, \mathrm{d}s$$

holds for all $t \in [0,1]$. By first rearranging arbitrary functions f and g into decreasing functions, we can apply this ordering for general measurable functions. The next proposition provides a more explicit link between the definition of $\preceq$ and the notion of the degree of concentration for functions.

Proposition 2.6.6. *For f and $g \in L^1([0,1])$, the following statements are equivalent:*

1. *f is majorized by g, i.e. $f \preceq g$.*

2. *For every convex function $\phi : \mathbb{R} \to \mathbb{R}$, we have*

$$\int_0^1 \phi(f(s))\, \mathrm{d}s \le \int_0^1 \phi(g(s))\, \mathrm{d}s \; .$$

3. *There exists a Markov operator* T *on* $L^1([0,1])$ *such that* $f = Tg$.

Proof. The equivalence of 1 and 2 is contained in Theorem 2.5 of Chong (1974), while that of 1 and 3 is shown in Theorem 4.9 of Day (1973). □

Proposition 2.6.7. *For* f *and* $g \in L^1([0,1])$, *the following inequalities hold:*

1. *If* $f^*g^* \in L^1([0,1])$, *then*

$$\int_0^1 |f^*(s)g_*(s)| \;\mathrm{d}s \leq \int_0^1 |f(s)g(s)| \;\mathrm{d}s \leq \int_0^1 |f^*(s)g^*(s)| \;\mathrm{d}s \ .$$

2. $f^* - g^* \preceq f - g \preceq f^* - g_*$.

Proof. The proofs of 1, called the Hardy-Littlewood inequality, and 2 can be found in (6.2) and (6.1) of Day (1972) (see also Theorem 13.4 in Chong and Rice (1971)). □

3 Comparing the extremal behaviour of copulas

Facing the choice between different portfolios, how do you determine which portfolio has the highest likelihood to result in a 'worst-case' outcome? That is, which portfolio is most likely to incur extremely large losses in all stocks simultaneously? The tail dependence function of a d-variate random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ aims to capture such a worst-case behaviour of $\boldsymbol{X}$. For a random vector $\boldsymbol{X}$ with continuous marginal distributions X_i, $i = 1, \dots, d$, it is given by

$$\Lambda\left(\boldsymbol{w}\,;\boldsymbol{X}\right) := w_1 \lim_{s\searrow 0} \mathbb{P}(X_2 \le F_2^{-1}(sw_2), \dots, X_d \le F_d^{-1}(sw_d) \mid X_1 \le F_1^{-1}(sw_1))$$

for $\boldsymbol{w} = (w_1, \dots, w_d) \in \mathbb{R}_+^d$ and describes the comovement in the extreme tail of the components $X_1, \dots, X_d$.

One common approach to compare the extremal behaviour of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is to order them using the tail dependence coefficient

$$\lambda(\boldsymbol{X}) := \lim_{s\searrow 0} \frac{C_{\boldsymbol{X}}(s \cdot \mathbf{1})}{s} = \Lambda\left(\mathbf{1}\,;\boldsymbol{X}\right) \ ,$$

and to call $\boldsymbol{Y}$ more tail dependent than $\boldsymbol{X}$ if $\lambda(\boldsymbol{X}) \le \lambda(\boldsymbol{Y})$. Reasonable as this may seem, Figure 3.1 illustrates the drawbacks of this approach for dimension $d = 2$: Even though $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ possess the same tail dependence coefficient, their overall tail behaviour varies drastically and is, simply put, not comparable.

Thus, we propose a more comprehensive stochastic order which allows for a meaningful comparison of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Given two d-variate random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we say $\boldsymbol{X}$ is less tail dependent than $\boldsymbol{Y}$, in short $\boldsymbol{X} \le_{tdo} \boldsymbol{Y}$, if

$$\Lambda\left(\boldsymbol{w}\,;\boldsymbol{X}\right) \le \Lambda\left(\boldsymbol{w}\,;\boldsymbol{Y}\right)$$

holds for all $\boldsymbol{w} \in \mathbb{R}_+^d$. The tail dependence order now provides a consistent notion of the degree of tail dependence but can be difficult to (visually) scan for certain aspects for dimensions d much larger than 2. To alleviate this problem, we introduce measures of tail dependence that condense complex phenomena into a single numerical value but remain consistent with $\le_{tdo}$ (think, e.g., of the average or maximal extremal dependence). The most prominent example of these measures of tail dependence is the aforementioned tail dependence coefficient λ, as $\boldsymbol{X} \le_{tdo} \boldsymbol{Y}$ implies $\lambda(\boldsymbol{X}) \le \lambda(\boldsymbol{Y})$. Although λ is already widely applied in the literature (see, for example, Christoffersen, Errunza, Jacobs and Langlois (2012) and Chabi-Yo, Ruenzi and Weigert (2018)), it only provides a consistent notion of the degree of tail dependence in the presence of an underlying order such as $\le_{tdo}$. And, most importantly, λ only incorporates the extremal behaviour along the diagonal $\boldsymbol{w} = (1, \dots, 1)$ and discards any additional information

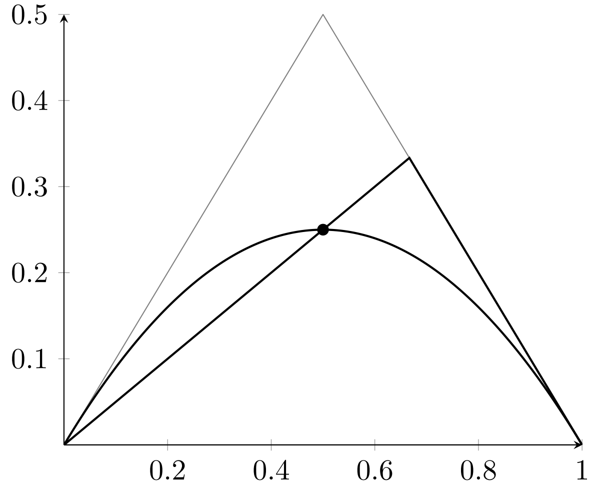


Figure 3.1: Plot depicting the tail dependence functions $\Lambda_1(s) = s(1-s)$ and $\Lambda_2(s) = \min\{s/3, 1-s\}$ and their respective tail dependence coefficient $\lambda_1 = \lambda_2 = 1/4$ (see Proposition 2.5.1 for the derivation of *a* corresponding copula). The upper bound Λ^+ is depicted in grey.

concerning the extremal behaviour of the components of the random vector $\boldsymbol{X}$ away from the diagonal. This focus on the diagonal is contrary to empirical evidence gathered in recent years, which often asserts distinctly asymmetric dependencies, for instance, between financial assets (see Okimoto (2008)).

Therefore, we introduce a broader class of consistent measures of tail dependence that can distinguish between various extremal characteristics undetected by the tail dependence coefficient alone. Among these measures of tail dependence are the average and maximal tail dependence given by

$$\mu_p(\boldsymbol{X}) := \|\widetilde{\Lambda}\,(\,\cdot\,;\boldsymbol{X})\,\|_p$$

for $p = 1$ and $p = \infty$, respectively. Furthermore, many well-known measures of extremal dependence from the literature are consistent with the tail dependence order $\leq_{tdo}$. This includes, for example, the conditional Spearman's ρ introduced in Schmid and Schmidt (2007) or the extremal coefficient given in Frahm (2006).

Having introduced a concept to compare the extremal behaviour of two random vectors, we then turn to the question raised at the beginning of this chapter: If a random vector $\boldsymbol{X}$ is less tail dependent than $\boldsymbol{Y}$, that is, if $\boldsymbol{X} \leq_{tdo} \boldsymbol{Y}$, is a worst-case scenario, where all components attain small values simultaneously, less likely to occur for $\boldsymbol{X}$ than for $\boldsymbol{Y}$? More precisely, does

$$\Lambda\,(\boldsymbol{w}\,;\boldsymbol{X}) < \Lambda\,(\boldsymbol{w}\,;\boldsymbol{Y}) \implies C_{\boldsymbol{X}}(\boldsymbol{u}) \leq C_{\boldsymbol{Y}}(\boldsymbol{u}) \tag{3.1}$$

hold for all $\boldsymbol{u}$ in some neighbourhood $B_\epsilon(\boldsymbol{0}) \cap [0,1]^d$? Considering the approximation

$$C_{\boldsymbol{X}}(\boldsymbol{u}) = \Lambda\,(\boldsymbol{u}\,;\boldsymbol{X}) + R(\boldsymbol{u})\,(u_1 + \ldots + u_d)$$

for $\boldsymbol{u} \in B_\epsilon(\boldsymbol{0}) \cap [0,1]^d$ with $R(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\|_1 \to 0$ established by Jaworski (2006) (see Section 2.4 for a more detailed overview), the tail dependence function should in some sense dominate the behaviour of the copula at least for small values of $\|\boldsymbol{u}\|_1$. Trivially, the pointwise ordering $C_{\boldsymbol{X}}(\boldsymbol{u}) \leq C_{\boldsymbol{Y}}(\boldsymbol{u})$ in some neighbourhood of zero implies $\boldsymbol{X} \leq_{tdo} \boldsymbol{Y}$, whereas the converse result in general does not hold. Nevertheless, we establish (3.1) on any cone bounded away from the axes. Furthermore, we present various copula families, such as the extreme-value and Archimedean copulas, for which (3.1) can be recovered.

Note that the 'tail dependence ordering', the corresponding order-preserving measures based on the L^p-norms and some of their properties have been investigated for dimension $d = 2$ in the diploma thesis by Harder (2013) under the supervision of Karl Friedrich Siburg.

This chapter is structured as follows: Section 3.1 introduces the ordering of tail dependence and its key properties. Our new measures of tail dependence are presented in Section 3.2, together with examples comparing them to the tail dependence coefficient. Section 3.3 then investigates the connection between the localized stochastic order and the tail dependence order for general copulas, whereas Section 3.4 focuses on certain parametric classes.

3.1 Tail dependence order

For the rest of this chapter, we assume that all random vectors possess continuous univariate marginal distributions and a tail dependence function.[1] Since our aim is to investigate the tail behaviour of some random vector $\boldsymbol{X}$, our proposed tail dependence order is based on its tail dependence function $\Lambda(\,\cdot\,;\boldsymbol{X})$.

Definition 3.1.1. *We say that a random vector $\boldsymbol{X}$ is less tail dependent than $\boldsymbol{Y}$, written $\boldsymbol{X} \leq_{tdo} \boldsymbol{Y}$, if and only if*

$$\Lambda(\boldsymbol{w};\boldsymbol{X}) \leq \Lambda(\boldsymbol{w};\boldsymbol{Y})$$

holds for all $\boldsymbol{w} \in \mathbb{R}^d_+$. Similarly, $\boldsymbol{X}$ is strictly less tail dependent than $\boldsymbol{Y}$, written $\boldsymbol{X} <_{tdo} \boldsymbol{Y}$, if and only if

$$\Lambda(\boldsymbol{w};\boldsymbol{X}) < \Lambda(\boldsymbol{w};\boldsymbol{Y})$$

holds for all $\boldsymbol{w} \in (0,\infty)^d$.

Whenever the two underlying random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are of no particular importance, we simply write $\Lambda_i(\boldsymbol{w})$, $i = 1,2$, instead of $\Lambda(\boldsymbol{w};\boldsymbol{X})$ and $\Lambda(\boldsymbol{w};\boldsymbol{Y})$. We point out that $\leq_{tdo}$ is only a preorder but following common practice, we will nevertheless call $\leq_{tdo}$ the *tail dependence order*.

Proposition 3.1.2. *The tail dependence order $\leq_{tdo}$ is a preorder, i.e. it is reflexive and transitive, but neither antisymmetric nor total. Furthermore,*

1. *the greatest and maximal elements of this preorder are copulas with a tail dependence function equal to $\Lambda(\boldsymbol{w};C^+)$.*

2. *the least and minimal elements are copulas with a tail dependence function equal to zero.*

[1] See Feidt, Genest and Nešlehová (2010) for a discussion of tail dependence in the presence of discontinuous univariate marginal distributions.

Proof. The first assertion follows immediately from the properties of the pointwise ordering on the space of continuous functions. Furthermore, $\leq_{tdo}$ is not antisymmetric due to $\Lambda(\boldsymbol{w};\Pi) = \Lambda(\boldsymbol{w};C^-)$ and $\Pi \neq C^-$. To see that $\leq_{tdo}$ is not total, consider the tail dependence functions $0.5 \cdot \min\{2w_1, w_2, \ldots, w_d\}$ and $0.5 \cdot \min\{w_1, w_2, \ldots, 2w_d\}$ constructed via the gluing technique introduced in Siburg and Stoimenov (2008b). To obtain the least and greatest elements, we apply Proposition 2.4.4, stating that

$$0 \leq \Lambda(\boldsymbol{w};C) \leq \Lambda\left(\boldsymbol{w};C^+\right)$$

holds for all d-copulas C. As these bounds are sharp, every copula with $0 = \Lambda(\boldsymbol{w};C)$ or $\Lambda(\boldsymbol{w};C^+) = \Lambda(\boldsymbol{w};C)$ is a least or greatest element, respectively. If C is a maximal element, then for every d-copula D, it holds that

$$C \leq_{tdo} D \implies D \leq_{tdo} C\ .$$

Thus, for $D = C^+$, we have $\Lambda(\boldsymbol{w};C) = \Lambda(\boldsymbol{w};C^+)$. Analogously, every minimal element must be tail independent. □

The tail dependence function $\Lambda(\boldsymbol{w};\boldsymbol{X})$ of a random vector $\boldsymbol{X}$ is, by definition, a function on $\mathbb{R}^d_+$. As discussed in Example 2.4.5, Λ is positive homogeneous of order 1, allowing us to restrict Λ to the compact set

$$\mathcal{S}^{d-1} := \left\{ \boldsymbol{w} \in \mathbb{R}^d_+ \;\middle|\; \sum_{\ell=1}^{d} w_\ell = 1 \right\}$$

with no loss of information. We will denote this restricted function by $\widetilde{\Lambda}$ and, abusing notation, still call $\widetilde{\Lambda}$ the tail dependence function.

Example 3.1.3. *For a bivariate random vector $\boldsymbol{X}$, the tail dependence coefficient is given by*

$$\lambda(\boldsymbol{X}) = 2\widetilde{\Lambda}(1/2;\boldsymbol{X}). \tag{3.2}$$

For an easier analysis, the following example presents different types of such restricted bivariate tail dependence functions. Note that some are symmetric with respect to the point 1/2, which is the case, for instance, if the two random variables X_1 and X_2 are exchangeable (see Definition 3.2.3 below). However, other tail dependence functions may exhibit a marked asymmetric pattern. Moreover, some tail dependence functions are ordered while others are not.

Example 3.1.4.

1. *Let $\boldsymbol{X}$ follow a bivariate jointly normal distribution with correlation $\rho \in [-1,1]$. Then, by Example 2.4.3, $\boldsymbol{X}$ is tail independent, that is, it fulfils $\Lambda(\boldsymbol{w};\boldsymbol{X}) = 0$ for all $\boldsymbol{w} \in \mathbb{R}^2_+$, whenever $\rho < 1$ holds, and tail dependent with $\Lambda(\boldsymbol{w};\boldsymbol{X}) = \min\{w_i\}$ in case of $\rho = 1$.*
2. *Example 2.4.5 yields a straightforward construction method for arbitrary tail dependence functions. Two asymmetric tail dependence functions are depicted in Figure 3.2.*

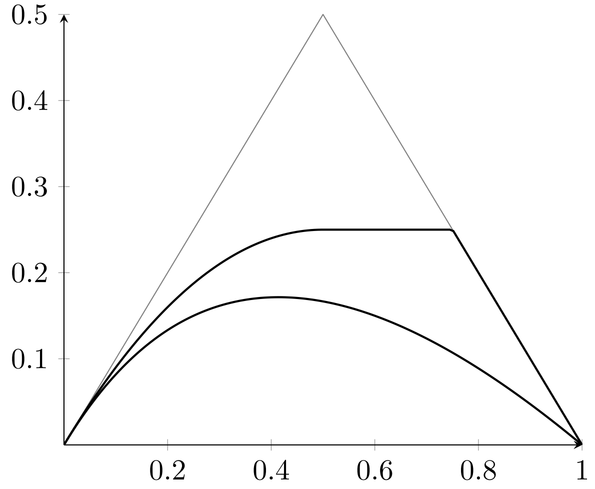


Figure 3.2: Following Example 2.4.5 and Remark 2.4.6, this plot depicts two asymmetric tail dependence functions (in black) and the upper bound $\min\{t, 1-t\}$ (in grey).

Remark 3.1.5. *An order similar to $\leq_{tdo}$ was introduced in Li (2013), who orders the copula values along rays. That is, C_1 is smaller than C_2 in the so-called 'tail orthant order', in short $C_1 \leq_{too} C_2$, if for all $\boldsymbol{w} \in \mathbb{R}_+^d$ there exists a $t_{\boldsymbol{w}} > 0$ such that*

$$C_1(s\boldsymbol{w}) \leq C_2(s\boldsymbol{w})$$

holds for all $s \leq t_{\boldsymbol{w}}$. While $\leq_{too}$ implies $\leq_{tdo}$, the converse does not hold. To see this, consider the Archimedean Joe 2-copula (see Family (6) in Charpentier and Segers (2009)) with parameter $\theta = 2$, which we will simply denote by C. Then gluing C and C^+ with respect to the first and second component (see Siburg and Stoimenov (2008b)) results in the copulas C_1 and C_2, respectively. On the one hand, as C is tail independent, so are C_1 and C_2 and $\Lambda(\boldsymbol{w}\,; C_1) = \Lambda(\boldsymbol{w}\,; C_2)$ holds for all $\boldsymbol{w} \in \mathbb{R}_+^d$. On the other hand, C_1 and C_2 are strictly ordered conversely along the directions $\boldsymbol{w}_1 = (\frac{1}{2}, 1)$ and $\boldsymbol{w}_2 = (1, \frac{1}{2})$, and therefore neither $C_1 \leq_{too} C_2$ nor $C_2 \leq_{too} C_1$ can hold.

Returning to the tail dependence order, the following simple result shows that the tail dependence order implies the ordering of the tail dependence coefficient but not vice versa.

Proposition 3.1.6. *If $\boldsymbol{X} \leq_{tdo} \boldsymbol{Y}$ then $\lambda(\boldsymbol{X}) \leq \lambda(\boldsymbol{Y})$ but the converse is generally not true.*

Proof. The first assertion is trivial in view of $\lambda(\boldsymbol{X}) = \Lambda(\boldsymbol{1}\,; \boldsymbol{X})$. The second assertion follows from the counterexamples in Figure 3.1. $\square$

3.2 Measures of tail dependence

While from a theoretical point of view, the tail dependence function encodes all necessary information about the extremal behaviour of the underlying random vector, a practitioner may want to pinpoint more specific aspects of the tail dependence into a single quantity. A possible application could be the quantification of the overall extremal behaviour or of certain diversification effects. This leads to the construction of *measures of tail dependence*, which are mappings μ that associate to each random vector $\boldsymbol{X}$ a number $\mu(\boldsymbol{X})$ reflecting certain aspects of the underlying tail behaviour. In order to be able to compare different values of such a measure, we are looking for measures that are monotone with respect to the tail dependence order.

Definition 3.2.1. *A measure of tail dependence is a function $\mu : \boldsymbol{X} \mapsto \mu(\boldsymbol{X}) \in [0, \infty]$ defined on the set of all d-variate random vectors $\boldsymbol{X}$ that satisfies the monotonicity condition*

$$\boldsymbol{X} \leq_{tdo} \boldsymbol{Y} \Longrightarrow \mu(\boldsymbol{X}) \leq \mu(\boldsymbol{Y}) \ .$$

Recall that the tail dependence function can be reduced to the unit simplex $\mathcal{S}^{d-1}$.

Theorem 3.2.2. *Suppose Λ is a d-variate tail dependence function. Then all of the following quantities are measures of tail dependence:*

1. *The maximal tail dependence*
$$\max_{\boldsymbol{s} \in \mathcal{S}^{d-1}} \Lambda(\boldsymbol{s}) \ .$$
2. *The average tail dependence*
$$\int_{\mathcal{S}^{d-1}} \Lambda(\boldsymbol{s}) \ \mathrm{d}\boldsymbol{s} \ .$$
3. *Any L^p-norm with $1 \leq p < \infty$*
$$\left(\int_{\mathcal{S}^{d-1}} \Lambda(\boldsymbol{s})^p \ \mathrm{d}\boldsymbol{s} \right)^{1/p} \ .$$

Proof. All of the above quantities are well defined since $\Lambda(\boldsymbol{s})$ is continuous and hence integrable on $\mathcal{S}^{d-1}$ according to Proposition 2.4.4. The monotonicity condition is obviously satisfied by all quantities. □

It is important to note the fundamental role played by the underlying tail dependence order here. Of course, every measure μ can be used to define a preorder $\leq_\mu$ on the d-variate random vectors by setting

$$\boldsymbol{X} \leq_\mu \boldsymbol{Y} :\Longleftrightarrow \mu(\boldsymbol{X}) \leq \mu(\boldsymbol{Y}).$$

This preorder $\leq_\mu$ is always reflexive, transitive and total but, in general, neither symmetric nor antisymmetric. Without the underlying tail dependence order, however, this construction may lead to grave inconsistencies. For instance, take the two measures of tail dependence from Theorem 3.2.2 for dimension $d = 2$ given by

$$\mu(\Lambda) := \max_{s \in [0,1]} \Lambda(s, 1-s) \ \text{ and } \nu(\Lambda) := \int_0^1 \Lambda(s, 1-s) \ \mathrm{d}s \ ,$$

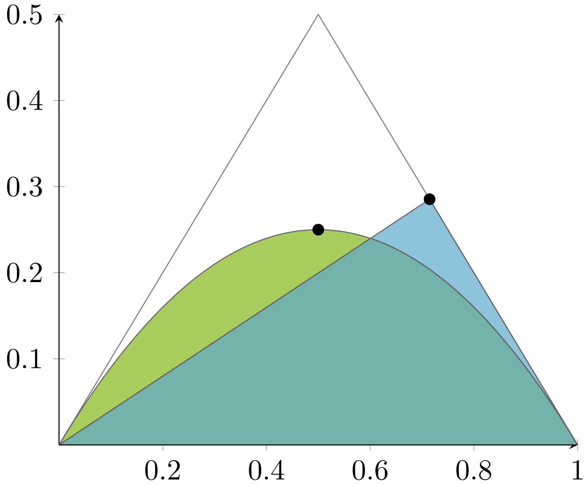


Figure 3.3: Plot depicting $\widetilde{\Lambda}_1(s) = \min\{0.4s, 1-s\}$ and $\widetilde{\Lambda}_2(s) = s(1-s)$. While the green and blue shaded areas mark the respective average tail dependence ν, the dot marks the maximal value μ.

and the tail dependence functions $\widetilde{\Lambda}_1(s) = \min\{0.4s, 1-s\}$ and $\widetilde{\Lambda}_2(s) = s(1-s)$ illustrated in Figure 3.3. Then Λ_1 and Λ_2 are not ordered with respect to $\leq_{tdo}$, however, they fulfil $\mu(\Lambda_1) > \mu(\Lambda_2)$ as well as $\nu(\Lambda_1) < \nu(\Lambda_2)$, so that we would end up with the inconsistent relations

$$\boldsymbol{X} >_\mu \boldsymbol{Y} \text{ and } \boldsymbol{X} <_\nu \boldsymbol{Y} ,$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random vectors with tail dependence function Λ_1 and Λ_2, respectively.

In practical applications, the choice of a measure of tail dependence will depend on the particular aim of the investigation— in some cases the maximal tail dependence may be appropriate while in others, the average is more relevant. In any case, 'evaluation measures' $\Lambda(\boldsymbol{s_0})$ for $\boldsymbol{s_0} \in \mathcal{S}^{d-1}$, in particular the classical tail dependence coefficient λ, are rather arbitrary choices. Only in very special cases does the tail dependence coefficient provide a reasonable measure of tail dependence, as we will subsequently explain.

Definition 3.2.3. *A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is called exchangeable if*

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \mathbb{P}(X_1 \leq x_{\pi(1)}, \ldots, X_d \leq x_{\pi(d)})$$

holds for all $\boldsymbol{x} \in \mathbb{R}^d$ and every permutation π of $\{1, \ldots, d\}$. This is equivalent to saying that the joint distribution functions of all permutated random vectors $\boldsymbol{X}^\pi$ coincide.

Proposition 3.2.4. *If $\boldsymbol{X}$ is exchangeable with tail dependence function Λ, then*

$$\lambda(\boldsymbol{X}) = d \cdot \Lambda(1/d, \ldots, 1/d) = d \cdot \max_{\boldsymbol{s} \in \mathcal{S}^{d-1}} \Lambda(\boldsymbol{s}) = d \cdot \|\widetilde{\Lambda}\|_\infty .$$

If $\widetilde{\Lambda}$ is strictly concave, then λ is the unique maximum.

Proof. $\Lambda(\,\cdot\,;\boldsymbol{X})$ is a continuous function on a compact set and as such, it attains its maximum value in some point $\boldsymbol{w} \in \mathcal{S}^{d-1}$. Due to the exchangeability of $\boldsymbol{X}$, every permutation of $\boldsymbol{w}$ maximizes $\Lambda(\,\cdot\,;\boldsymbol{X})$ as well. Moreover, the set $\mathbb{O}$ of local optima is convex as $\Lambda(\,\cdot\,;\boldsymbol{X})$ is concave. Therefore, we only need to verify that $\frac{1}{d}\mathbf{1}$ lies in $\mathbb{O}$. A direct calculation yields

$$\frac{1}{d!}\sum_{\pi\in\mathcal{P}_d}\boldsymbol{w}^{\pi} = \frac{1}{d!}(d-1)!\sum_{\ell=1}^{d} w_\ell \mathbf{1} = \frac{1}{d}\mathbf{1}\ ,$$

where $\mathcal{P}_d$ denotes the set of all possible permutations of $\{1,\dots,d\}$. The last claim follows immediately from the strict concavity of $\widetilde{\Lambda}$. □

We now give examples of measures of tail dependence that are somewhat more intricate than the measures described in Theorem 3.2.2.

Example 3.2.5. *Suppose Λ_1 and Λ_2 are bivariate tail dependence functions. Then*

$$\Lambda_1 \le \Lambda_2 \implies C^{LEV}(\,\cdot\,,\Lambda_1) \le C^{LEV}(\,\cdot\,,\Lambda_2)$$

follows immediately from Proposition 2.5.1. Thus, for any concordance measure κ (see Section 2.3) it holds

$$\Lambda_1 \le \Lambda_2 \implies \kappa(C^{LEV}(\,\cdot\,,\Lambda_1)) \le \kappa(C^{LEV}(\,\cdot\,,\Lambda_2))$$

and $\Lambda \mapsto \kappa(C^{LEV}(\,\cdot\,,\Lambda))$ is a measure of tail dependence.

Example 3.2.6. *Spearman's ρ for lower extreme-value 2-copulas yields the measure of tail dependence*

$$\mu(\Lambda) := \rho\left(C^{LEV}(\,\cdot\,,\Lambda)\right) = 12\int_0^1 \frac{1}{\left(2-\Lambda(t,1-t)\right)^2}\ \mathrm{d}t - 3\ .$$

The final class of measures of tail dependence we present is motivated directly by the majorization theory outlined in Section 2.6.

Proposition 3.2.7. *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-variate random vectors. Then $\boldsymbol{X} \le_{tdo} \boldsymbol{Y}$ implies*

$$\int_0^\infty \phi(\partial_1\Lambda(t,w_2,\dots,w_d\,;\boldsymbol{X}))\ \mathrm{d}t \le \int_0^\infty \phi(\partial_1\Lambda(t,w_2,\dots,w_d\,;\boldsymbol{Y}))\ \mathrm{d}t$$

for all $(w_2,\dots,w_d) \in \mathbb{R}_+^{d-1}$ and all increasing convex functions $\phi:\mathbb{R}_+\to\mathbb{R}_+$ such that $\phi(0)=0$ and $t\mapsto\phi(\partial_1\Lambda(t,w_2,\dots,w_d\,;C^+))$ is integrable.

Proof. The result follows from the equivalence of $\Lambda(\,\cdot\,;\boldsymbol{X}) \le \Lambda(\,\cdot\,;\boldsymbol{Y})$ and

$$\int_0^{w_1} \partial_1\Lambda(t,w_2,\dots,w_d\,;\boldsymbol{X})\ \mathrm{d}t \le \int_0^{w_1} \partial_1\Lambda(t,w_2,\dots,w_d\,;\boldsymbol{Y})\ \mathrm{d}t$$

for all $w_1,\dots,w_d \ge 0$. An application of Theorem 2.1 in Chong (1974) then gives the result. □

Example 3.2.8. *Given an increasing convex function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ as in Proposition 3.2.7 and a d-variate random vector $\boldsymbol{X}$,*

$$\mu_\phi(\boldsymbol{X}) := \int\limits_{\mathbb{R}_+^d} \phi(\partial_1 \Lambda\,(\boldsymbol{w}\,;\boldsymbol{X}))\ \mathrm{d}\boldsymbol{w}$$

is a measure of tail dependence. Similar to the convex order of random variables, measures of this type quantify the degree of concentration of a tail dependence function.

Finally, we point out that functional combinations $f(\mu, \nu)$ of measures of tail dependence μ and ν, where f is real-valued and increasing in each argument, are likewise measures of tail dependence. A simple example is any linear combination $a\mu + b\nu$ with $a, b \geq 0$. Such combinations can be useful in applications where one would like to consider different aspects of tail dependence simultaneously.

Example 3.2.9. *Frahm (2006) introduced the so-called 'extremal dependence coefficient'*

$$\varepsilon_L := \frac{\lambda}{2-\lambda} = \frac{\Lambda\,(1/2)}{1-\Lambda\,(1/2)} = f(\Lambda\,(1/2))$$

with $f(\mu) := \frac{\mu}{1-\mu}$, which investigates the asymptotic dependence between the componentwise minima and maxima of a pair (X_1, X_2) of random variables. Since the function f is increasing, the quantity ε_L is a measure of tail dependence.

Example 3.2.10. *Schmid and Schmidt (2007) introduced a local conditional version of Spearman's ρ by setting*

$$\rho(\varepsilon\,; C) := \frac{\int\limits_{[0,\varepsilon]^d} C(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u} - \left(\frac{\varepsilon^2}{2}\right)^d}{\frac{\varepsilon^{d+1}}{d+1} - \left(\frac{\varepsilon^2}{2}\right)^d}\,,$$

where an application of the dominated convergence theorem yields

$$\rho_L(C) := \lim_{\varepsilon \searrow 0} \rho(\varepsilon\,; C) = (d+1)\,\|\Lambda\,(\,\cdot\,; C)\|_{L^1([0,1]^d)}\ .$$

Therefore, $\rho_L(C)$ is a measure of tail dependence and quantifies the scaled average tail dependence. We will now briefly explore a consequence of the assumption $\rho_L(C_1) < \rho_L(C_2)$. It implies the existence of an $\varepsilon^* > 0$ such that

$$\int\limits_{[0,\varepsilon]^d} C_1(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u} \leq \int\limits_{[0,\varepsilon]^d} C_2(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u}$$

holds for all $\varepsilon < \varepsilon^*$. Thus, a strict ordering of this measure leads to an ordering of the averages of the copula on $[0,\varepsilon]^d$. In the next section, we will investigate similar results under the assumption of $C_1 <_{tdo} C_2$.

3.3 The tail behaviour of general copulas

While the copula C uniquely determines the limiting behaviour $\Lambda(\boldsymbol{w}\,; C)$, every tail dependence function Λ captures a wide variety of tail behaviours from different copulas. For example, both the product copula $\Pi(\boldsymbol{u}) = u_1 u_2$ as well as the lower Fréchet-Hoeffding bound $C^-(\boldsymbol{u}) = \max\{u_1 + u_2 - 1, 0\}$ for $\boldsymbol{u} \in [0,1]^2$ are tail independent, i.e. they fulfil $\Lambda(\boldsymbol{w}) = 0$ for all $\boldsymbol{w} \in \mathbb{R}^2_+$, even though their stochastic behaviour is markedly different. Thus, when considering the approximation of C established in Theorem 2.4.2,

$$C(\boldsymbol{u}) = \Lambda(\boldsymbol{u}\,; C) + R(\boldsymbol{u})\,(u_1 + \ldots + u_d)$$

with $R(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\|_1 \to 0$, the function R may encode a wide variety of local behaviours. We are therefore interested in the extent to which the tail dependence function alone dominates the behaviour of the underlying copula near zero, such that

$$\Lambda(\boldsymbol{w}\,; C_1) < \Lambda(\boldsymbol{w}\,; C_2) \implies C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u}) \tag{3.3}$$

holds for all $\boldsymbol{u}$ in some neighbourhood $B_\varepsilon(\boldsymbol{0}) \cap [0,1]^d$. Here, we define the neighbourhood $B_\varepsilon(\boldsymbol{0})$ by $\left\{x \in \mathbb{R}^d \mid \|x\|_1 \le \epsilon\right\}$, although the choice of the norm is irrelevant for our purposes. Note that we do not consider $\Lambda(\boldsymbol{w}\,; C_1) \le \Lambda(\boldsymbol{w}\,; C_2) \implies C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u})$, due to the counterexample stated in Remark 3.1.5.

The right-hand side of (3.3) defines an adapted version of the usual stochastic dominance ordering, which has briefly been considered in Hua (2012) under the term 'ultimate usual stochastic order.' It is reminiscent of the concept of the germ of a function.

Definition 3.3.1. *Let C_1 and C_2 be d-copulas. We say*

1. *C_1 is smaller than C_2 in the lower orthant order, in short $C_1 \le C_2$, if $C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u})$ holds for all $\boldsymbol{u} \in [0,1]^d$.*
2. *C_1 is smaller than C_2 in the local lower orthant order, in short $C_1 \le_{loc} C_2$, if there exists a neighbourhood $B_\varepsilon(\boldsymbol{0})$ of zero such that $C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u})$ holds for all $\boldsymbol{u} \in B_\varepsilon(\boldsymbol{0}) \cap [0,1]^d$.*

Note that $\le_{loc}$ is distinctly weaker than $\le$, which is easily seen from patchwork or gluing techniques (see Durante and Sempi (2016)). We start by discussing a straightforward connection between $\le_{tdo}$ and $\le_{loc}$.

Proposition 3.3.2. *For any two d-copulas C_1 and C_2, $C_1 \le_{loc} C_2$ implies $C_1 \le_{tdo} C_2$.*

Proof. The result follows immediately from the definitions. □

Unfortunately, the following example shows that (3.3) cannot hold in general. This example was communicated to us by Piotr Jaworski.

Example 3.3.3 (Example due to Piotr Jaworski)**.** *Keeping in mind the ray-like definition of the tail dependence function, we use the path (t, t^α) that 'bends' around any given cone. Consider the Marshall-Olkin copula $M_\alpha(\boldsymbol{u}) = \min\left\{u_1^{1-\alpha} u_2, u_1\right\}$ with parameter $\alpha \in (0,1)$ and the Archimedean Clayton copula C_ϑ with $\vartheta > 0$ given by $C_\vartheta(\boldsymbol{u}) = \left(u_1^{-\vartheta} + u_2^{-\vartheta} - 1\right)^{-\frac{1}{\vartheta}}$. While M_α is strictly smaller than C_ϑ in the tail dependence order with*

$$\Lambda(\boldsymbol{w}\,; M_\alpha) = 0 < \Lambda(\boldsymbol{w}\,; C_\vartheta) \text{ for all } \boldsymbol{w} \in (0,\infty)^2 \,,$$

setting $\boldsymbol{u} := (t, t^\alpha)$ *yields*

$$C_\vartheta(t,t^\alpha) = \frac{t^{1+\alpha}}{\left(t^\vartheta + t^{\alpha\vartheta} - t^{(1+\alpha)\vartheta}\right)^{\frac{1}{\vartheta}}} = \frac{t}{\left(t^{(1-\alpha)\vartheta} + 1 - t^\vartheta\right)^{\frac{1}{\vartheta}}} < t = M_\alpha(t,t^\alpha) \ .$$

Thus, even the strict tail dependence ordering does not imply an ordering of the values of the underlying copulas.

While the general result does not hold, the next result shows that the difficulties arise only in the vicinity of the axes.

Theorem 3.3.4. *Suppose the d-copulas* C_1 *and* C_2 *fulfil* $C_1 <_{tdo} C_2$. *Then for any cone*[2] $S \subset (0,\infty)^d$ *such that* $S \cup \{\mathbf{0}\}$ *is closed in* $[0,\infty)^d$, *there exists an* $\varepsilon > 0$ *such that*

$$C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u}) \text{ for all } \boldsymbol{u} \in S \cap B_\varepsilon(\mathbf{0}) \ .$$

Proof. Our proof follows a standard technique in convex analysis (see, e.g., Scholtes (2012)). For ease of notation, let S_0 denote $S \cup \{0\}$ and define $f : S_0 \to [-1,1]$ as $f(\boldsymbol{u}) := C_2(\boldsymbol{u}) - C_1(\boldsymbol{u})$. Due to $C_1 <_{tdo} C_2$, the directional derivative of f in $\mathbf{0}$ exists and is strictly positive, i.e. $f'(\mathbf{0}\,;\,\boldsymbol{s}) > 0$ for all $\boldsymbol{s} \in S$, and f is Lipschitz continuous with constant 2. We will show via contradiction that $\mathbf{0}$ is a local minimum of f on S_0, which in turn implies

$$C_2(\boldsymbol{u}) - C_1(\boldsymbol{u}) = f(\boldsymbol{u}) \ge 0$$

in a neighbourhood of $\mathbf{0}$. Thus, assume there exists a sequence $S \ni \boldsymbol{w}_n \to \mathbf{0}$ and $f(\boldsymbol{w}_n) \le f(\mathbf{0})$. We will decompose $\boldsymbol{w}_n$ into $\boldsymbol{s}_n := \frac{\boldsymbol{w}_n}{\|\boldsymbol{w}_n\|_1}$ and $r_n := \|\boldsymbol{w}_n\|_1$. Then $\boldsymbol{s}_n$ has a convergent subsequence, again denoted by $\boldsymbol{s}_n$, with $\boldsymbol{s}_n \to \boldsymbol{s}^* \in S$ and it holds that

$$\begin{aligned} 0 &\ge \frac{f(\boldsymbol{w}_n) - f(\mathbf{0})}{r_n} = \frac{f(r_n\boldsymbol{s}_n) - f(\mathbf{0})}{r_n} \\ &= \frac{f(r_n\boldsymbol{s}_n) - f(r_n\boldsymbol{s}^*)}{r_n} + \frac{f(r_n\boldsymbol{s}^*) - f(\mathbf{0})}{r_n} \to f'(\mathbf{0}\ ;\ \boldsymbol{s}^*) > 0 \ , \end{aligned}$$

which is a contradiction. The convergence of the first term follows from the Lipschitz continuity of f,

$$0 \le \frac{|f(r_n\boldsymbol{s}_n) - f(r_n\boldsymbol{s}^*)|}{r_n} \le \frac{2\,\|r_n\boldsymbol{s}_n - r_n\boldsymbol{s}^*\|_1}{r_n} = 2\,\|\boldsymbol{s}_n - \boldsymbol{s}^*\|_1 \to 0 \ . \qquad \square$$

3.4 The tail behaviour of certain copula families

We will now focus on specific copula families which provide better control on the behaviour near the axes to recover (3.3). The main idea behind our approach is the reduction of the behaviour of C to a parameter linked to the tail dependence function. The first and most immediate class are the (lower) extreme-value 2-copulas, which are uniquely determined in the bivariate case by their tail dependence function Λ (see Section 2.5).

[2] A set S is called cone if $\boldsymbol{w} \in S$ implies $\lambda\boldsymbol{w} \in S$ for any $\lambda > 0$.

Proposition 3.4.1. *The tail dependence order is equivalent to the lower orthant order for lower extreme-value* 2*-copulas. That is,*

$$\Lambda(\boldsymbol{w}\,; C_1) \leq \Lambda(\boldsymbol{w}\,; C_2) \Leftrightarrow \widehat{C}_1 \leq \widehat{C}_2 \Leftrightarrow C_1 \leq C_2 \Leftrightarrow C_1 \leq_{loc} C_2$$

holds for all lower extreme-value 2*-copulas* C_1 *and* C_2*, where* $\widehat{C}_1$ *denotes the survival copula of* C_1*.*

Proof. The proof follows immediately from Equation (2.9), Proposition 2.5.1 and the fact that for all 2-copulas C_1 and C_2, $C_1 \leq C_2$ holds if and only if $\widehat{C}_1 \leq \widehat{C}_2$ holds. □

Another easily parametrized class of 2-copulas for which the strict tail dependence ordering implies the local orthant ordering are the diagonal copulas.

Definition 3.4.2. *A* 2*-copula* C *is called a diagonal copula if*

$$C(\boldsymbol{u}) = \min\left\{u_1, u_2, \frac{\delta(u_1) + \delta(u_2)}{2}\right\}$$

holds for some function $\delta : [0,1] \to [0,1]$ *fulfilling*

1. $\delta(t) \leq t$ *for all* $t \in [0,1]$*,*
2. $\delta(1) = 1$*,*
3. δ *is increasing, and,*
4. $|\delta(t) - \delta(s)| \leq 2\,|t - s|$ *for all* $s, t \in [0,1]$*.*

δ *is called the diagonal section of* C*.*

Proposition 3.4.3. *For diagonal* 2*-copulas, the strict extremal order implies the local orthant order.*

Proof. Let C_1 and C_2 be diagonal 2-copulas with diagonal sections δ_1 and δ_2, respectively. Due to C_1 and C_2 admitting a tail dependence function, we have

$$\lim_{s \searrow 0} \frac{\delta_2(s) - \delta_1(s)}{s} = \delta_2'(0) - \delta_1'(0) = \Lambda(\mathbf{1}\,; C_2) - \Lambda(\mathbf{1}\,; C_1) > 0\ .$$

Thus, $\delta_2 - \delta_1$ must be nonnegative on $[0, \varepsilon)$ for some $\varepsilon > 0$, which implies

$$C_1(\boldsymbol{u}) = \min\left\{u_1, u_2, \frac{\delta_1(u_1) + \delta_1(u_2)}{2}\right\} \leq \min\left\{u_1, u_2, \frac{\delta_2(u_1) + \delta_2(u_2)}{2}\right\} = C_2(\boldsymbol{u})$$

for $\boldsymbol{u} = (u_1, u_2) \in [0, \varepsilon)^2$. □

Lastly, we investigate the more elaborate case of Archimedean d-copulas. We may restrict our subsequent analysis to strict Archimedean generators ϕ, that is, generators fulfilling

$$\lim_{s \searrow 0} \phi(s) = \infty\ ,$$

as the Archimedean copula is otherwise necessarily equal to zero in a neighbourhood around zero (see Charpentier and Segers (2009)). The following lemma gives a reformulation of this fact in our language of stochastic orders.

Lemma 3.4.4. *Let C be an Archimedean d-copula with nonstrict generator ϕ and let $\widetilde{C}$ be an arbitrary d-copula. Then $C \leq_{tdo} \widetilde{C}$ is equivalent to $C \leq_{loc} \widetilde{C}$.*

Proof. Let C be an Archimedean d-copula with nonstrict generator ϕ. Following Section 3.2 in Charpentier and Segers (2009), there exists an $\varepsilon > 0$ such that

$$C(\boldsymbol{u}) = 0 \quad \text{for all } \boldsymbol{u} \in B_\varepsilon(\mathbf{0}) \cap [0,1]^d$$

and necessarily $C \leq_{loc} \widetilde{C}$ for all d-copulas $\widetilde{C}$. Moreover, $\Lambda(\boldsymbol{w}\,; C)$ is identically zero for all $\boldsymbol{w} \in \mathbb{R}_+^d$ and therefore $\Lambda(\boldsymbol{w}\,; C) = 0 \leq \Lambda\left(\boldsymbol{w}\,; \widetilde{C}\right)$ for all copulas $\widetilde{C}$. Thus $C \leq_{tdo} \widetilde{C}$ is equivalent to $C \leq_{loc} \widetilde{C}$. □

Note that for strict Archimedean generators, the generalized inverse reduces to the usual inverse. Let us now state the main theorem before proving the necessary technical results.

Theorem 3.4.5. *Let C_1 and C_2 be Archimedean d-copulas with regularly varying generators ϕ_1 and ϕ_2, respectively. Then*

$$C_1 <_{tdo} C_2 \implies C_1 \leq_{loc} C_2 \ .$$

We will briefly outline the proof, whose detailed version is given at the end of this chapter. It follows a localized version of Chapter 4.4 in Nelsen (2006).

Outline of the proof of Theorem 3.4.5. The proof consists of several steps:

1. $C_1 \leq_{loc} C_2$ whenever $(\phi_1 \circ \phi_2^{-1})$ is subadditive around ∞ (see Proposition 3.4.7).

2. $(\phi_1 \circ \phi_2^{-1})$ is subadditive around ∞ if $\frac{\phi_1(x)}{\phi_2(x)}$ is increasing around 0 (see Proposition 3.4.8).

3. $\lambda(C_1) < \lambda(C_2)$ implies that $\frac{\phi_1(x)}{\phi_2(x)}$ is increasing around 0. □

In Nelsen (2006), the lower orthant ordering of two Archimedean copulas C_1 and C_2 is characterized in terms of the global subadditivity of $(\phi_1 \circ \phi_2^{-1})$. We can relax this condition and only require subadditivity for large values to ensure the local orthant ordering $\leq_{loc}$. Such a localized version has previously also been considered in Hua (2012), where the generator is assumed to be the Laplace transform of a positive random variable.

Definition 3.4.6. *We say a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is subadditive around ∞ if there exists an $M \geq 0$ such that*

$$f(x+y) \leq f(x) + f(y)$$

holds for all $x, y \in [M, \infty)$.

Proposition 3.4.7. *Let C_1 and C_2 be Archimedean d-copulas generated by ϕ_1 and ϕ_2, respectively. Then $C_1 \leq_{loc} C_2$ whenever $(\phi_1 \circ \phi_2^{-1})$ is subadditive around ∞.*

Proof. Consider $f : \mathbb{R}_+ \to \mathbb{R}_+$ defined by $f(x) := (\phi_1 \circ \phi_2^{-1})(x)$. As f is subadditive around ∞, we have

$$f\left(\sum_{\ell=1}^{d} x_\ell\right) \leq \sum_{\ell=1}^{d} f(x_\ell) \qquad \text{for all } x_1, \dots, x_d \geq M \ .$$

Applying the strictly decreasing function ϕ_1^{-1} to the above inequality yields

$$\phi_2^{-1}\left(\sum_{\ell=1}^{d} x_\ell\right) = \left(\phi_1^{-1} \circ f\right)\left(\sum_{\ell=1}^{d} x_\ell\right) \geq \phi_1^{-1}\left(\sum_{\ell=1}^{d} f(x_\ell)\right) \qquad \text{for all } x_1, \dots, x_d \geq M \ .$$

Since ϕ_2 is strictly decreasing, $x_\ell \geq M$ implies $\phi_2^{-1}(x_\ell) \leq \phi_2^{-1}(M)$. Therefore, for any $\boldsymbol{u} \in B_\varepsilon(\mathbf{0}) \cap [0,1]^d$ with $\varepsilon := \phi_2^{-1}(M) > 0$ and $x_\ell := \phi_2(u_\ell) \geq \phi_2(\varepsilon) = M$, we have

$$\begin{aligned}\phi_2^{-1}\left(\sum_{\ell=1}^{d} \phi_2(u_\ell)\right) &= \phi_2^{-1}\left(\sum_{\ell=1}^{d} x_\ell\right) \geq \phi_1^{-1}\left(\sum_{\ell=1}^{d} f(x_\ell)\right) \\ &= \phi_1^{-1}\left(\sum_{\ell=1}^{d} \left(\phi_1 \circ \phi_2^{-1}\right)\left(\phi_2(u_\ell)\right)\right) = \phi_1^{-1}\left(\sum_{\ell=1}^{d} \phi_1(u_\ell)\right) \ .\end{aligned}$$

This yields $C_1(\boldsymbol{u}) \leq C_2(\boldsymbol{u})$ for all $\boldsymbol{u} \in B_\varepsilon(\mathbf{0}) \cap [0,1]^d$ and thus $C_1 \leq_{loc} C_2$. □

Unfortunately, the subadditivity of $\phi_1 \circ \phi_2^{-1}$ around ∞ is rather difficult to validate and we therefore present a sufficient criterion similar to Corollary 4.5 in Nelsen (2006).

Proposition 3.4.8. *Let C_1 and C_2 be Archimedean d-copulas generated by ϕ_1 and ϕ_2, respectively. Then $(\phi_1 \circ \phi_2^{-1})$ is subadditive around ∞ whenever $\frac{\phi_1}{\phi_2}$ is increasing on $(0, \varepsilon)$ for some $\varepsilon > 0$.*

Proof. Define $g : (0,\infty) \to (0,\infty)$ as $g(x) := \frac{f(x)}{x}$ and $f(x) := (\phi_1 \circ \phi_2^{-1})(x)$ as before. We now verify that g is decreasing on (M, ∞) with $M := \phi_2(\varepsilon)$. As ϕ_2 is strictly decreasing for all $x \geq 0$, we have for $M \leq x \leq y$

$$0 < \phi_2^{-1}(y) \leq \phi_2^{-1}(x) \leq \phi_2^{-1}(M) = \varepsilon \ .$$

Combining this with the fact that $g \circ \phi_2 = \frac{\phi_1}{\phi_2}$ is increasing on $(0, \varepsilon)$, we have

$$g(x) = (g \circ \phi_2)\left(\phi_2^{-1}(x)\right) \geq (g \circ \phi_2)\left(\phi_2^{-1}(y)\right) = g(y) \ ,$$

so g is decreasing for $x, y \geq M$. Thus, for all $x, y \in [M, \infty)$

$$x\left(g(x+y) - g(x)\right) + y\left(g(x+y) - g(y)\right) \leq 0$$

or equivalently

$$f(x+y) = (x+y)g(x+y) \leq xg(x) + yg(y) = f(x) + f(y) \ .$$

Thus, f is subadditive around ∞. □

Up to this point, our approach was entirely independent of the tail dependence function of the Archimedean copulas. To ensure the existence of a tail dependence function in the following, we will assume that the Archimedean generators are regularly varying (see Section 2.5 for a detailed account). We will show that ordered tail dependence functions result in ordered parameters of regular variation, allowing us to apply Proposition 3.4.8 for Archimedean copulas with regularly varying generators.

Lemma 3.4.9. *Let C_1 and C_2 be Archimedean d-copulas with strict Archimedean generators ϕ_1 and ϕ_2. Furthermore, suppose ϕ_1 and ϕ_2 are regularly varying at 0 with parameters $-\alpha_1$ and $-\alpha_2 \in [-\infty, 0]$, respectively. Then the following are equivalent:*

1. $C_1 <_{tdo} C_2$.
2. $\lambda(C_1) < \lambda(C_2)$.
3. $\alpha_1 < \alpha_2$.

Proof. The first implication 1 to 2 is immediate. For the second implication 2 to 3, consider the possible combinations of $\lambda \in [0, 1]$:

1. If $\lambda(C_1) = 0$ and $\lambda(C_2) \in (0, 1)$, then
$$0 = \lambda(C_1) < \lambda(C_2) = d^{-\frac{1}{\alpha_2}}$$
and thus $\alpha_2 > 0 = \alpha_1$ holds.
2. If both $\lambda(C_1)$ and $\lambda(C_2)$ take values in $(0, 1)$, then $\alpha_1 < \alpha_2$ due to
$$d^{-\frac{1}{\alpha_1}} = \lambda(C_1) < \lambda(C_2) = d^{-\frac{1}{\alpha_2}}$$
3. If $\lambda(C_1) \in (0, 1)$ and $\lambda(C_2) = 1$, then
$$d^{-\frac{1}{\alpha_1}} = \lambda(C_1) < \lambda(C_2) = 1$$
and thus $\alpha_1 < \infty = \alpha_2$.
4. Lastly, if $\lambda(C_1) = 0$ and $\lambda(C_2) = 1$, then $a_1 = 0 < \infty = a_2$.

The last implication 3 to 1 follows from Example 3.8 of Li (2013). □

Proof of Theorem 3.4.5. To show $C_1 \leq_{loc} C_2$, we will invoke Proposition 3.4.8, i.e. show that $\psi(x) := \frac{\phi_1(x)}{\phi_2(x)}$ is increasing in a neighbourhood of zero. An application of Proposition 3.4.7 then yields the desired result. Due to Lemma 3.4.9, we have $\alpha_1 < \alpha_2$, where ϕ_i is regularly varying with coefficient $-\alpha_i$ in 0. Furthermore, as ϕ_1 and ϕ_2 are positive, convex, regularly varying functions, Lemma A.1. in Charpentier and Segers (2009) yields

$$\lim_{s \searrow 0} \frac{s\phi_i'(s)}{\phi_i(s)} = -\alpha_i$$

for $i = 1$ and 2. Here, ϕ_i' denotes an increasing representative of the derivative of ϕ_i, which is only defined almost everywhere. ψ is almost everywhere differentiable as the ratio of continuous and almost everywhere differentiable functions, where

$$\frac{\phi_1'(s)\phi_2(s) - \phi_1(s)\phi_2'(s)}{\phi_2(s)^2}$$

is a representative of the derivative of ψ, which we will denote by ψ'. This implies

$$\begin{aligned}\lim_{s \searrow 0} \frac{s\psi'(s)}{\psi(s)} &= \lim_{s \searrow 0} \left(s \frac{\phi_1'(s)\phi_2(s) - \phi_1(s)\phi_2'(s)}{\phi_2(s)^2} \right) \frac{\phi_2(s)}{\phi_1(s)} \\ &= \lim_{s \searrow 0} \frac{s\phi_1'(s)}{\phi_1(s)} - \frac{s\phi_2'(s)}{\phi_2(s)} = -\alpha_1 - (-\alpha_2) = \alpha_2 - \alpha_1 > 0 \ .\end{aligned}$$

Due to $\psi \geq 0$, ψ' must be positive on $(0, \varepsilon)$ for some $\varepsilon > 0$. Finally, given $x_1, x_2 \in (0, \varepsilon)$ with $x_1 \leq x_2$, we have that ϕ_1 and ϕ_2 are absolutely continuous on $[x_1, x_2]$. This yields that ψ as the ratio of absolutely continuous functions on $[x_1, x_2]$ is absolutely continuous on $[x_1, x_2]$ and therefore increasing on $(0, \varepsilon)$ due to

$$\psi(x_2) - \psi(x_1) = \int_{x_1}^{x_2} \psi'(s) \, \mathrm{d}s \geq 0 \ . \qquad \square$$

4 A Markov product for tail dependence functions

The Markov product is an important tool in the analysis of bivariate copulas. Originally introduced by Darsow et al. (1992) as

$$(C_1 * C_2)(\boldsymbol{u}) = \int_0^1 \partial_2 C_1(u_1, t) \cdot \partial_1 C_2(t, u_2) \, \mathrm{d}t \ , \tag{4.1}$$

the Markov product of the 2-copulas C_1 and C_2 has since then been generalized in several ways. By using a weighting of the two factors C_1 and C_2 other than $\Pi(u, v) = u \cdot v$ and by allowing for more than two 2-copulas C_1 and C_2, we arrive at the generalized Markov product (see Section 5.5 in Durante and Sempi (2016))

$$\phi_C \left(C_1, \ldots, C_d\right)(\boldsymbol{u}) := \int_0^1 C\left(\partial_1 C_1\left(t, u_1\right), \ldots, \partial_1 C_d\left(t, u_d\right)\right) \, \mathrm{d}t \ . \tag{4.2}$$

For dimensions d higher than 2, this generalized Markov product has become a popular method to construct d-copulas, as it combines d bivariate building blocks into one d-copula, while for dimensions $d = 2$, it is an important tool to investigate, for instance, complete dependence (see Section 2.2).

Considering the analytical similarities between copulas and tail dependence functions, one may wonder whether it is possible to derive the tail dependence function of the d-copula ϕ_C simply from the newly-defined product

$$\phi_C \left(\Lambda_1, \ldots, \Lambda_d\right)(\boldsymbol{w}) := \int_0^\infty C\left(\partial_1 \Lambda_1\left(t, w_1\right), \ldots, \partial_1 \Lambda_d\left(t, w_d\right)\right) \, \mathrm{d}t \tag{4.3}$$

for their respective tail dependence functions.

In this chapter, we will show that the answer is affirmative— but one has to impose certain Sobolev-type regularity conditions on the underlying copulas $C_1, \ldots, C_d$. Then, the generalized Markov products on $\mathcal{C}_2^d$ and $\mathcal{T}_2^d$ do indeed commute with the tail dependence operation, such that

$$\phi_C \left(\Lambda_{C_1}, \ldots, \Lambda_{C_d}\right) = \Lambda_{\phi_C(C_1, \ldots, C_d)} \ , \tag{4.4}$$

where we used the shorthand $\Lambda_C(\boldsymbol{w}) := \Lambda\left(\boldsymbol{w}\,; C\right)$. Similar results have been established in the context of vine-copulas by Joe et al. (2010) and more recently by Jaworski (2015).

Analytically, the Markov products on $\mathcal{C}_2^d$ and $\mathcal{T}_2^d$ share many properties, but the additional concavity of tail dependence functions turns out to have far reaching consequences regarding the behaviour of the (generalized) Markov product. Aside from stronger convergence results, the concavity induces a fundamental reduction property:

$$\phi_C\left(\Lambda_1,\ldots,\Lambda_d\right)(\boldsymbol{w}) \leq \min_{\substack{k,\ell=1,\ldots,d\\ \ell\neq k}} \Lambda_k(w_\ell,w_k)\ , \tag{4.5}$$

whenever C is negative quadrant dependent, i.e. whenever $C(\boldsymbol{u}) \leq \Pi(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$.

The concavity of tail dependence functions is also the key ingredient in our analysis of iterates and idempotents of the Markov product. Since $\phi_C\left(\Lambda_1,\ldots,\Lambda_d\right)$ creates a d-variate tail dependence function from d bivariate ones, an iterative application requires the limitation to $d=2$. Hence, we consider a Markov product for tail dependence functions analogous to the original Markov product (4.1) and define

$$(\Lambda_1 * \Lambda_2)(\boldsymbol{w}) := \int_0^\infty \partial_2\Lambda_1(w_1,t)\cdot\partial_1\Lambda_2(t,w_2)\ \mathrm{d}t\ .$$

Note that $*$ indeed combines two bivariate tail dependence functions into a new bivariate tail dependence function. In particular, for a bivariate tail dependence function Λ, we can now define the n-fold iteration

$$\Lambda^{*n} := \underbrace{\Lambda * \ldots * \Lambda}_{n \text{ times}}\ .$$

We then characterize the limits $\lim_{n\to\infty}\Lambda^{*n}$, and thereby the behaviour of idempotents (i.e. tail dependence functions fulfilling $\Lambda*\Lambda=\Lambda$), as either $\Lambda^0(\boldsymbol{w})=0$ or $\Lambda^+(\boldsymbol{w})=\min\boldsymbol{w}$ using the concavity of Λ. This characterization for the Markov product on $\mathcal{T}_2$ is in stark contrast to the behaviour of $*$ on $\mathcal{C}_2$, where, for example, the iterated application of C^- yields a circular pattern, that is, we have

$$\left(C^-\right)^{*(2k)} = C^+ \text{ and } \left(C^-\right)^{*(2k+1)} = C^-\ ,$$

resulting in a 2-periodic orbit.

Up until now, we have considered the Markov product as a purely analytical operation. But the Markov product for copulas (4.1) can also be constructed from an isomorphism between copulas and Markov operators as seen in Section 2.2. In the last part of this chapter, we adopt this point of view and establish a similar relationship between tail dependence functions and substochastic operators known from majorization theory. Essentially, a substochastic operator T is nonexpansive with respect to the L^p-norms $\|\cdot\|_p$ for $1\leq p\leq\infty$ and maps positive functions f onto positive functions Tf, thereby generalizing the usual Markov operators.

This chapter is based on Siburg and Strothmann (2021a) and is structured as follows: Section 4.1 introduces the Markov product for tail dependence functions and establishes a link to the original Markov product for copulas. Section 4.2 discusses the reduction property (4.5) unique to tail dependence functions, while Section 4.3 employs the reduction property to derive the behaviour of iterates and idempotents. Lastly, Section 4.4 establishes a connection between $(\mathcal{T}_2,*)$ and the substochastic operators equipped with the composition.

4.1 A Markov product for tail dependence functions

We investigate a generalized version of the Markov product introduced by Darsow et al. (1992), which was discussed in Durante and Sempi (2016) and, in the context of vine-copulas, in Jaworski (2015).

Definition 4.1.1. *Let $C_1, \ldots, C_d$ be 2-copulas and let C be a d-copula. Then, the $(d+1)$-copula*

$$\phi_{u_0,C}(C_1, \ldots, C_d)(u_1, \ldots, u_d) := \int_0^{u_0} C\left(\partial_1 C_1(t, u_1), \ldots, \partial_1 C_d(t, u_d)\right)\, \mathrm{d}t$$

is called the C-lifting of the copulas $C_1, \ldots, C_d$. Furthermore, we define the d-copula

$$\begin{aligned}\phi_C(C_1, \ldots, C_d)(u_1, \ldots, u_d) &:= \int_0^{1} C\left(\partial_1 C_1(t, u_1), \ldots, \partial_1 C_d(t, u_d)\right)\, \mathrm{d}t \\ &= \phi_{1,C}(C_1, \ldots, C_d)(u_1, \ldots, u_d)\end{aligned}$$

to be the generalized Markov product of $C_1, \ldots, C_d$ induced by C.

Similar to this construction of higher dimensional copulas from bivariate copulas, we introduce an operation on bivariate tail dependence functions.

Definition 4.1.2. *Let $\Lambda_1, \ldots, \Lambda_d \in \mathcal{T}_2$ and $C \in \mathcal{C}_d$. We call*

$$\phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) := \int_0^{w_0} C\left(\partial_1 \Lambda_1(t, w_1), \ldots, \partial_1 \Lambda_d(t, w_d)\right)\, \mathrm{d}t$$

the C-lifting of the tail dependence functions $\Lambda_1, \ldots, \Lambda_d$. Similarly, we define the generalized Markov product of $\Lambda_1, \ldots, \Lambda_d$ induced by C as

$$\phi_C(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) := \int_0^{\infty} C\left(\partial_1 \Lambda_1(t, w_1), \ldots, \partial_1 \Lambda_d(t, w_d)\right)\, \mathrm{d}t\ .$$

We now verify that the C-lifting and the generalized Markov product do in fact generate new tail dependence functions.

Theorem 4.1.3. *Suppose C is a d-copula and $\Lambda_1, \ldots, \Lambda_d \in \mathcal{T}_2$. Then $\phi_{\cdot,C}(\Lambda_1, \ldots, \Lambda_d)$ and $\phi_C(\Lambda_1, \ldots, \Lambda_d)$ are $(d+1)$-variate and d-variate tail dependence functions, respectively.*

Proof. The tail dependence functions Λ_ℓ for $\ell = 1, \ldots, d$ are positive, increasing in each component, Lipschitz continuous and thus have partial derivatives almost everywhere. Moreover, the partial derivatives attain values in $[0,1]$. Therefore, we have

$$\begin{aligned}0 \le \phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) &\le \int_0^{\infty} C^+\left(\partial_1 \Lambda_1(t, w_1), \ldots, \partial_1 \Lambda_d(t, w_d)\right)\, \mathrm{d}t \\ &\le \min_{\ell=1,\ldots,d} \|\partial_1 \Lambda_\ell(t, w_\ell)\|_1 = \min_{\ell=1,\ldots,d} \left(\lim_{t\to\infty} \Lambda_\ell(t, w_\ell)\right) \\ &\le \min_{\ell=1,\ldots,d} w_\ell < \infty\ ,\end{aligned}$$

which establishes the existence of the integral. The last inequality is due to Λ_ℓ being increasing in each component and bounded from above by Λ^+. Thus, we can define

$$\phi(\boldsymbol{w}) := \int_0^{w_0} C\left(\partial_1\Lambda_1(t,w_1),\ldots,\partial_1\Lambda_d(t,w_d)\right)\,\mathrm{d}t\ .$$

It remains to verify Properties a. to c. from the characterization of $(d+1)$-variate tail dependence functions of Proposition 2.4.4. For the first property, note that due to all copulas being bounded from above by C^+ and due to all tail dependence functions having bounded partial derivatives between 0 and 1, it holds

$$\begin{aligned}0 &\le \int_0^{w_0} C\left(\partial_1\Lambda_1(t,w_1),\ldots,\partial_1\Lambda(t,w_d)\right)\,\mathrm{d}t\\ &\le \begin{cases}\int_0^{w_0} 1\,\mathrm{d}t = w_0\\ \int_0^{\infty} C^+\left(\partial_1\Lambda_1(t,w_1),\ldots,\partial_1\Lambda_d(t,w_d)\right)\,\mathrm{d}t \le \min\limits_{\ell=1,\ldots,d} w_\ell\end{cases} = \Lambda^+(w_0,\ldots,w_d)\ .\end{aligned}$$

The $(d+1)$-increasing property of ϕ needs to be verified on every rectangle $R \subset \mathbb{R}_+^{d+1}$. A direct calculation identical to that of Proposition 2.2 in Durante, Klement and Quesada-Molina (2008) with $v_\ell < w_\ell$ yields

$$V_\phi\left(\bigtimes_{\ell=0}^{d}[v_\ell,w_\ell]\right) = \int_{v_0}^{w_0} V_C\left(\bigtimes_{\ell=1}^{d}\left[\partial_1\Lambda_\ell(t,v_\ell),\partial_1\Lambda_\ell(t,w_\ell)\right)\right)\,\mathrm{d}t \ge 0\ ,$$

which holds due to $\partial_1\Lambda_\ell(t,v_\ell) \le \partial_1\Lambda_\ell(t,w_\ell)$. Lastly, the positive homogeneity can be established via a change of variables and via the positive homogeneity of order 0 of the partial derivatives of Λ (see, e.g., Joe et al. (2010)), yielding

$$\begin{aligned}\phi(s\boldsymbol{w}) &= \int_0^{sw_0} C\left(\partial_1\Lambda_1(t,sw_1),\ldots,\partial_1\Lambda_d(t,sw_d)\right)\,\mathrm{d}t\\ &= \int_0^{sw_0} C\left(\partial_1\Lambda_1(t/s,w_1),\ldots,\partial_1\Lambda_d(t/s,w_d)\right)\,\mathrm{d}t\\ &= s\int_0^{w_0} C\left(\partial_1\Lambda_1(z,w_1),\ldots,\partial_1\Lambda_d(z,w_d)\right)\,\mathrm{d}z = s\phi(\boldsymbol{w})\ .\end{aligned}$$

By Proposition 2.4.4, we can thus find a $(d+1)$-copula C with $\Lambda(\boldsymbol{w}\,;C) = \phi(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathbb{R}_+^{d+1}$. The proof that $\phi_C(\Lambda_1,\ldots,\Lambda_d)$ is a d-variate tail dependence function is a simplified version of this proof. □

Remark 4.1.4. *Note that the first part of the proof of Theorem 4.1.3 only requires that all Λ_ℓ are 2-increasing functions bounded from below by 0 and from above by Λ^+. Furthermore, ϕ is positive homogeneous of order one if $\partial_1\Lambda_\ell(t,w_\ell)$ is homogeneous of order zero for all $\ell = 1,\ldots,d$.*

The next proposition gives some basic algebraic properties for both $\phi_C(\Lambda_1, \ldots, \Lambda_d)$ and $\phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)$.

Proposition 4.1.5. *Suppose $\Lambda_1, \ldots, \Lambda_d$ are bivariate tail dependence functions and C is a d-copula. Then it holds:*

1. $\Lambda^+ = \Lambda(\,\cdot\,; C^+)$ *is the unit element in the sense that if* $\Lambda_\ell = \Lambda^+$, *then*

$$\phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) = \phi_{\min\{w_0,w_\ell\},C_{-\ell}}(\Lambda_1, \ldots, \Lambda_{\ell-1}, \Lambda_{\ell+1}, \ldots, \Lambda_d)(\boldsymbol{w_{-\ell}})$$

   with $\boldsymbol{w_{-\ell}} := (w_1, \ldots, w_{\ell-1}, w_{\ell+1}, \ldots, w_d)$ *and* $C_{-\ell} := C(u_1, \ldots, u_{\ell-1}, 1, u_{\ell+1}, \ldots, u_d)$.

2. $\Lambda^0 := \Lambda(\,\cdot\,; \Pi)$ *is the null element in the sense that if* $\Lambda_\ell = \Lambda^0$, *then*

$$\phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) = \Lambda\left(w_0, \ldots, w_d\,; \Pi^{d+1}\right) = 0 \ .$$

3. *If C is convex (concave) in the ℓ-th component, then $\phi_{w_0,C}(\,\cdot\,)$ is convex (concave) in the ℓ-th component.*[1]

4. *For every permutation π on $\{1, \ldots, d\}$ with inverse permutation τ, we have*

$$\phi_C(\Lambda_1, \ldots, \Lambda_d)(w_{\tau(1)}, \ldots, w_{\tau(d)}) = \phi_{C^\pi}\left(\Lambda_{\pi(1)}, \ldots, \Lambda_{\pi(d)}\right)(w_1, \ldots, w_d) \ ,$$

   where $C^\pi(u_1, \ldots, u_d) := C(u_{\pi(1)}, \ldots, u_{\pi(d)})$.

5. *If $C \leq D$ pointwise, then* $\phi_{\cdot,C}(\Lambda_1, \ldots, \Lambda_d) \leq \phi_{\cdot,D}(\Lambda_1, \ldots, \Lambda_d)$.

Remark 4.1.6. *The term 'unit-element' stems from the bivariate case, where the generalized Markov product $\phi_C : \mathcal{T}_2 \times \mathcal{T}_2 \to \mathcal{T}_2$ fulfils*

$$\phi_C\left(\Lambda^+, \Lambda\right)(w_1, w_2) = \Lambda(w_1, w_2) \ .$$

Proof. 1. Without loss of generality, we consider $\ell = 1$. As $\partial_1\Lambda(t, w_1\,; C^+) = \mathbb{1}_{[0,w_1]}(t)$, we have

$$\begin{aligned}
\phi_{w_0,C}(\Lambda_1, \ldots, \Lambda_d)(w_1, \ldots, w_d) &= \int_0^{\min\{w_0,w_1\}} C(1, \partial_1\Lambda_2(t, w_2), \ldots, \partial_1\Lambda_d(t, w_d))\, \mathrm{d}t \\
&= \int_0^{\min\{w_0,w_1\}} C_{-1}(\partial_1\Lambda_2(t, w_2), \ldots, \partial_1\Lambda_d(t, w_d))\, \mathrm{d}t \\
&= \phi_{\min\{w_0,w_1\},C_{-1}}(\Lambda_2, \ldots, \Lambda_d)(w_2, \ldots, w_d) \ .
\end{aligned}$$

2. The second result follows from $\Lambda(\boldsymbol{w}\,; \Pi) = 0$ for $\boldsymbol{w} \in \mathbb{R}^d_+$ and $C(0, \boldsymbol{v}) = 0$ with $\boldsymbol{v} \in [0,1]^{d-1}$.

3. The third result follows immediately from the componentwise convexity (concavity).

[1] Chapter 5 discusses such componentwise convex and concave copulas in more detail.

4. A direct calculation yields

$$\begin{aligned}\phi_C\left(\Lambda_1,\dots,\Lambda_d\right)\left(w_{\tau(1)},\dots,w_{\tau(d)}\right) &= \int_0^\infty C\left(\partial_1\Lambda_1(t,w_{\tau(1)}),\dots,\partial_1\Lambda_d(t,w_{\tau(d)})\right)\,\mathrm{d}t \\ &= \int_0^\infty C^\pi\left(\partial_1\Lambda_{\pi(1)}(t,w_1),\dots,\partial_1\Lambda_{\pi(d)}(t,w_d)\right)\,\mathrm{d}t \\ &= \phi_{C^\pi}\left(\Lambda_{\pi(1)},\dots,\Lambda_{\pi(d)}\right)(w_1,\dots,w_d)\ .\end{aligned}$$

5. The last assertion follows immediately from the pointwise inequality $C \leq D$. □

We now derive some convergence results for the Markov product of tail dependence functions. In the case of 2-copulas, Siburg and Stoimenov (2008a) and Trutschnig (2011) introduced different metrics allowing for the joint convergence of the (generalized) Markov product for 2-copulas, i.e.

$$\phi_C(C_{n,1},C_{n,2}) \to \phi_C(C_1,C_2)\ ,$$

whenever $C_{n,i}$ converges in some sense towards C_i, $i=1,2$. We will present very similar conditions in the case of tail dependence functions. Note, however, that due to the different domains of copulas and tail dependence functions and the concavity of the latter, convergence of the partial derivatives coincides with the L^1-convergence of the partial derivatives in the case of copulas and with pointwise convergence in the case of tail dependence functions.

Proposition 4.1.7. *Suppose* $\Lambda_1,\dots,\Lambda_d \in \mathcal{T}_2$ *and* C *is a* d*-copula.*

1. Let $C_n \in \mathcal{C}_d$ *with* $C_n \to C$ *pointwise. Then*

$$\phi_{w_0,C_n}\left(\Lambda_1,\dots,\Lambda_d\right) \to \phi_{w_0,C}\left(\Lambda_1,\dots,\Lambda_d\right)$$

pointwise for all $w_0 \in \mathbb{R}_+$.

2. Let $\Lambda_{n,i} \in \mathcal{T}_2$ *with* $\Lambda_{n,i} \to \Lambda_i$ *pointwise. Then*

$$\phi_{w_0,C}\left(\Lambda_{n,1},\dots,\Lambda_{n,d}\right) \to \phi_{w_0,C}\left(\Lambda_1,\dots,\Lambda_d\right)$$

pointwise.

3. Let $\Lambda_{n,i} \in \mathcal{T}_2$ *with* $\|\partial_1\Lambda_{n,i}(\,\cdot\,,w)-\partial_1\Lambda_i(\,\cdot\,,w)\|_{L^1(\mathbb{R}_+)} \to 0$ *for all* $w \in \mathbb{R}_+$*, then*

$$\phi_C\left(\Lambda_{n,1},\dots,\Lambda_{n,d}\right) \to \phi_C\left(\Lambda_1,\dots,\Lambda_d\right)$$

pointwise.

Proof. 1. A combination of $C_n(\partial_1\Lambda_1(t,w_1),\dots,\partial_1\Lambda_d(t,w_d)) \leq \partial_1\Lambda_1(t,w_1)$ and the dominated convergence theorem yields the desired result.

2. Due to the Lipschitz continuity of C, we have

$$\begin{aligned}&\left|\phi_{w_0,C}\left(\Lambda_{n,1},\ldots,\Lambda_{n,d}\right)(\boldsymbol{w})-\phi_{w_0,C}\left(\Lambda_1,\ldots,\Lambda_d\right)(\boldsymbol{w})\right|\\ \leq&\sum_{i=1}^{d}\int_0^{w_0}\left|\partial_1\Lambda_{n,i}(t,w_i)-\partial_1\Lambda_i(t,w_i)\right|\,\mathrm{d}t\ .\end{aligned}$$

Thus, it suffices to consider each integral separately. As tail dependence functions are concave, Lemma 1 in Tsuji (1952) yields that for all fixed $w_i\in\mathbb{R}_+$, $\Lambda_{n,i}(t,w_i)\to\Lambda_i(t,w_i)$ holds pointwise if and only if $\partial_1\Lambda_{n,i}(t,w_i)\to\partial_1\Lambda_i(t,w_i)$ holds for almost all $t\in[0,w_0]$. Thus, an application of the dominated convergence theorem in combination with $0\leq\partial_1\Lambda_{n,i}\leq 1$ yields the desired result.

3. Again, due to the Lipschitz continuity of C, we have

$$\left|\phi_C\left(\Lambda_{n,1},\ldots,\Lambda_{n,d}\right)(\boldsymbol{w})-\phi_C\left(\Lambda_1,\ldots,\Lambda_d\right)(\boldsymbol{w})\right|\leq\sum_{i=1}^{d}\int_0^{\infty}\left|\partial_1\Lambda_{n,i}(t,w_i)-\partial_1\Lambda_i(t,w_i)\right|\,\mathrm{d}t\ ,$$

which converges to zero. □

Analogously to the binary product $*$ on $\mathcal{C}_2\times\mathcal{C}_2$ induced by Π, we introduce $*$ on $\mathcal{T}_2\times\mathcal{T}_2$ via

$$(\Lambda_1*\Lambda_2)(w_1,w_2):=\phi_\Pi\left(\Lambda_1^\top,\Lambda_2\right)(w_1,w_2)=\int_0^\infty\partial_2\Lambda_1(w_1,t)\cdot\partial_1\Lambda_2(t,w_2)\,\mathrm{d}t\ .$$

Its properties closely resemble those of the Markov product on $\mathcal{C}_2\times\mathcal{C}_2$. In particular, Λ^+ and Λ^0 are the unit and null element of $*$, respectively, and $*$ is skew-symmetric, that is,

$$(\Lambda_1*\Lambda_2)^\top=\Lambda_2^\top*\Lambda_1^\top\ .$$

With these basic algebraic properties, we will develop two conditions under which the Markov product commutes with the tail dependence function, i.e.

$$\Lambda\left(\boldsymbol{w}\,;C_1*C_2\right)=\left(\Lambda\left(\,\cdot\,;C_1\right)*\Lambda\left(\,\cdot\,;C_2\right)\right)(\boldsymbol{w})\ .\tag{4.6}$$

The following example shows that we require some condition to ensure that the Markov product commutes with the tail dependence operation.

Example 4.1.8. *Consider the tail independent* 2*-copulas* C^- *and* Π*. While they fulfil*

$$\Lambda\left(\boldsymbol{w}\,;C^-\right)=0=\Lambda\left(\boldsymbol{w}\,;\Pi\right)$$

for all $\boldsymbol{w}\in\mathbb{R}_+^2$*, their behaviour under the Markov product for copulas is distinctly different:*

$$(\Pi*\Pi)(\boldsymbol{u})=\Pi(\boldsymbol{u})\leq C^+(\boldsymbol{u})=(C^-*C^-)(\boldsymbol{u})$$

for all $\boldsymbol{u} \in [0,1]^2$. *Now assume that the Markov product and tail dependence operation commute unconditionally, then*

$$\begin{aligned}(\Lambda(\,\cdot\,;\Pi) * \Lambda(\,\cdot\,;\Pi))(\boldsymbol{w}) &= \Lambda(\boldsymbol{w};\Pi)\\ &< \Lambda\left(\boldsymbol{w};C^+\right)\\ &= \left(\Lambda\left(\,\cdot\,;C^-\right) * \Lambda\left(\,\cdot\,;C^-\right)\right)(\boldsymbol{w})\\ &= (\Lambda(\,\cdot\,;\Pi) * \Lambda(\,\cdot\,;\Pi))(\boldsymbol{w})\end{aligned}$$

for all $\boldsymbol{w} \in (0,\infty)^2$, *which is a contradiction.*

Our first approach to a condition that ensures (4.6) utilizes the Lipschitz continuity of C and follows an idea from Jaworski (2015). Theorem 7 therein derives the tail behaviour of the C-lifting

$$\Lambda(w_0,\ldots,w_d;\phi_{\cdot,C}(C_1,\ldots,C_d)) = \phi_{w_0,C}(\Lambda(\,\cdot\,;C_1),\ldots,\Lambda(\,\cdot\,;C_d))(w_1,\ldots,w_d)$$

under a Sobolev-type condition imposed on $C_1,\ldots,C_d$.

Theorem 4.1.9. *Suppose that* C *is a* d*-copula and that* $C_1,\ldots,C_d$ *are* 2*-copulas, which fulfil the Sobolev-type condition*

$$\lim_{s\searrow 0}\int_0^\infty \left|\partial_1 C_i(st,sw)\mathbb{1}_{[0,\frac{1}{s}]}(t) - \partial_1\Lambda(t,w;C_i)\right|\,\mathrm{d}t = 0 \tag{4.7}$$

for all $w \in \mathbb{R}_+$ *and all* $i = 1,\ldots,d$. *Then,*

$$\phi_C(\Lambda(\,\cdot\,;C_1),\ldots,\Lambda(\,\cdot\,;C_d))(\boldsymbol{w}) = \Lambda(\boldsymbol{w};\phi_C(C_1,\ldots,C_d))$$

for all $\boldsymbol{w} \in \mathbb{R}_+^d$, *or, equivalently,*

$$\begin{array}{ccc}\mathcal{C}_2^d & \xrightarrow{\phi_C} & \mathcal{C}_d\\ {\scriptstyle\Lambda(\cdot\,;C_i)}\big\downarrow & \circlearrowleft & \big\downarrow{\scriptstyle\Lambda(\cdot\,;C)}\\ \mathcal{T}_2^d & \xrightarrow[\phi_C]{} & \mathcal{T}_d\end{array}$$

Proof. The Lipschitz continuity and groundedness of C yield

$$\begin{aligned}&\left|C(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d))\,\mathbb{1}_{[0,\frac{1}{s}]}(\tau) - C(\partial_1\Lambda(\tau,w_1;C_1),\ldots,\partial_1\Lambda(\tau,w_d;C_d))\right|\\ &\le \sum_{\ell=1}^d \left|\partial_1 C_\ell(s\tau,sw_\ell)\mathbb{1}_{[0,\frac{1}{s}]}(\tau) - \partial_1\Lambda(\tau,w_\ell;C_\ell)\right| .\end{aligned}$$

Thus,

$$\begin{aligned}&|\Lambda(\boldsymbol{w};\phi_C(C_1,\ldots,C_d)) - \phi_C(\Lambda(\,\cdot\,;C_1),\ldots,\Lambda(\,\cdot\,;C_d))(\boldsymbol{w})|\\ &\le \lim_{s\searrow 0}\sum_{\ell=1}^d\int_0^\infty \left|\partial_1 C_\ell(s\tau,sw_\ell)\mathbb{1}_{[0,\frac{1}{s}]}(\tau) - \partial_1\Lambda(\tau,w_\ell;C_\ell)\right|\,\mathrm{d}\tau = 0\,.\end{aligned}$$

□

Using the concept of strict tail dependence functions yields a more feasible sufficient condition for Theorem 4.1.9. We call a tail dependence function strict if it has margins in the sense of Nelsen (2006).

Definition 4.1.10. *Let Λ be a bivariate tail dependence function. We call Λ strict if*

$$\lim_{t\to\infty} \Lambda(w_1, t) = w_1 \quad \text{and} \quad \lim_{t\to\infty} \Lambda(t, w_2) = w_2$$

hold for all $(w_1, w_2) \in \mathbb{R}_+^2$.

Remark 4.1.11. *Assume that in addition to the almost everywhere pointwise convergence of the partial derivatives, the tail dependence functions of all 2-copulas C_i are strict. Then an application of Scheffé's Lemma (see Novinger (1972)) yields*

$$\lim_{s\searrow 0} \int_0^\infty \left| \partial_1 C_i(st, sw) \mathbb{1}_{[0,\frac{1}{s}]}(t) - \partial_1 \Lambda\left(t, w\,; C_i\right) \right| \,\mathrm{d}t = 0$$

for all $i = 1, \ldots, d$, which in turn implies

$$\phi_C\left(\Lambda\left(\,\cdot\,; C_1\right), \ldots, \Lambda\left(\,\cdot\,; C_d\right)\right)(\boldsymbol{w}) = \Lambda\left(\boldsymbol{w}\,; \phi_C\left(C_1, \ldots, C_d\right)\right)$$

due to Theorem 4.1.9.

Example 4.1.12. *Suppose C_ϕ is an Archimedean 2-copula with generator ϕ, which is regularly varying in 0 with parameter $-\alpha \in (-\infty, 0)$. Then its tail dependence function equals (see Lemma 2.5.13)*

$$\Lambda\left(\boldsymbol{w}\,; C_\phi\right) = \left(w_1^{-\alpha} + w_2^{-\alpha}\right)^{-1/\alpha} .$$

Thus, $\Lambda(\,\cdot\,; C_\phi)$ is strict. The convergence of the partial derivatives follows similarly to the proof of Theorem 3.1 in Charpentier and Segers (2009) when combined with Theorem 1 in Charpentier and Segers (2007). Therefore, Archimedean copulas fulfil the conditions given in Remark 4.1.11, whenever ϕ is differentiable in a neighbourhood of zero.

The next approach to a condition that ensures (4.6) does not utilize the Lipschitz continuity of the copula C and yields a different condition in terms of the convergence of the partial derivatives.

Theorem 4.1.13. *Suppose C is a d-copula, $C_1, \ldots, C_d$ are 2-copulas and that their generalized Markov product possesses a tail dependence function. If*

$$\lim_{s\searrow 0} \partial_1 C_i(st, sw) = \partial_1 \Lambda\left(t, w\,; C_i\right)$$

holds for all $w \in \mathbb{R}_+$, almost all $t \in \mathbb{R}_+$ and all $i = 1, \ldots, d$, then

$$\phi_C\left(\Lambda\left(\,\cdot\,; C_1\right), \ldots, \Lambda\left(\,\cdot\,; C_d\right)\right)(\boldsymbol{w}) \le \Lambda\left(\boldsymbol{w}\,; \phi_C\left(C_1, \ldots, C_d\right)\right)$$

for all $\boldsymbol{w} \in \mathbb{R}_+^d$. Additionally, if there exists an $\ell \in \{1, \ldots, d\}$ such that

$$\partial_1 C_\ell(s\tau, sw_\ell) \mathbb{1}_{[0,\frac{1}{s}]}(\tau) \le g_{w_\ell}(\tau)$$

for all $w_\ell \in [0, 1]$ and some family $(g_w)_{w\in[0,1]}$ of integrable functions, it holds

$$\phi_C\left(\Lambda\left(\,\cdot\,; C_1\right), \ldots, \Lambda\left(\,\cdot\,; C_d\right)\right)(\boldsymbol{w}) = \Lambda\left(\boldsymbol{w}\,; \phi_C\left(C_1, \ldots, C_d\right)\right) .$$

Proof. By the definition of the tail dependence function and an application of Fatou's lemma for positive measurable functions, it holds that

$$\begin{aligned}
\Lambda\left(\boldsymbol{w}\,;\phi_C\left(C_1,\ldots,C_d\right)\right) &= \lim_{s\searrow 0}\frac{1}{s}\int_0^1 C\left(\partial_1 C_1(t,sw_1),\ldots,\partial_1 C_d(t,sw_d)\right)\,\mathrm{d}t\\
&= \lim_{s\searrow 0}\frac{1}{s}\int_0^{\frac{1}{s}} C\left(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d)\right)s\,\mathrm{d}\tau\\
&= \lim_{s\searrow 0}\int_{\mathbb{R}_+} C\left(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d)\right)\mathbb{1}_{[0,\frac{1}{s}]}(\tau)\,\mathrm{d}\tau\\
&\geq \int_{\mathbb{R}_+}\lim_{s\searrow 0} C\left(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d)\right)\mathbb{1}_{[0,\frac{1}{s}]}(\tau)\,\mathrm{d}\tau\\
&= \int_{\mathbb{R}_+} C\left(\partial_1\Lambda\left(\tau,w_1\,;C_1\right),\ldots,\partial_1\Lambda\left(\tau,w_d\,;C_d\right)\right)\,\mathrm{d}\tau\\
&= \phi_C\left(\Lambda\left(\,\cdot\,;C_1\right),\ldots,\Lambda\left(\,\cdot\,;C_d\right)\right)(\boldsymbol{w})\ .
\end{aligned}$$

If component ℓ's partial derivative is dominated by an integrable function g_{w_ℓ}, we have that for $\tau \leq 1/s$

$$\begin{aligned}
C\left(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d)\right) &\leq C^+\left(\partial_1 C_1(s\tau,sw_1),\ldots,\partial_1 C_d(s\tau,sw_d)\right)\\
&\leq \partial_1 C_\ell(s\tau,sw_\ell) \leq g_{w_\ell}(\tau)\ .
\end{aligned}$$

The desired result follows from the dominated convergence theorem. □

Example 4.1.14. *If C has a tail dependence function and continuous second-order partial derivatives, then Joe et al. (2010) have shown that*

$$\lim_{s\searrow 0}\partial_1 C(st,sw) = \partial_1\Lambda\left(t,w\,;C\right)$$

holds for almost all t and all $w\in\mathbb{R}_+$, thus fulfilling the condition of Theorem 4.1.13.

Example 4.1.15. *Consider the lower extreme-value copula C^{LEV} induced by the tail dependence function $\Lambda\in\mathcal{T}_2$. It follows from Proposition 2.5.1 that C^{LEV} is the survival copula of the extreme value copula C^{EV} induced by Λ. Thus, employing the positive homogeneity of degree 0 of $\partial_1\Lambda(u,v)$, we have*

$$\begin{aligned}
\partial_1 C^{LEV}(su,sv) &= 1-\partial_1 C^{EV}(1-su,1-sv)\\
&= 1-\frac{C^{EV}(1-su,1-sv)}{1-su}\left(1-\partial_1\Lambda(-\log(1-su),-\log(1-sv))\right)\\
&= 1-\underbrace{\frac{C^{EV}(1-su,1-sv)}{1-su}}_{\to 1\ as\ s\searrow 0}\left(1-\partial_1\Lambda\left(1,\underbrace{\frac{\log(1-sv)}{\log(1-su)}}_{\to v/u\ as\ s\searrow 0}\right)\right)\ .
\end{aligned}$$

As $t \mapsto \partial_1 \Lambda(w,t)$ is increasing and bounded from below by 0 and from above by 1, we have

$$\lim_{s \searrow 0} \partial_1 C^{LEV}(su, sv) = \partial_1 \Lambda \left(1, \frac{v}{u}\right) = \partial_1 \Lambda (u, v)$$

for almost all u and all $v \in [0,1]$.

The lower bound behaviour stated in Theorem 4.1.13 is generally the best result possible, as can be seen from the following example.

Example 4.1.16. *Consider the lower Fréchet-Hoeffding bound C^-, which is symmetric and left invertible, i.e. $(C^-)^\top * C^- = C^+$. Then an application of Theorem 5.5.3 in Durante and Sempi (2016) yields*

$$\phi_C \left(C^-, C^-\right) = C^- * C^- = \left(C^-\right)^\top * C^- = C^+ \ .$$

Hence, for $\boldsymbol{w} = (w_1, w_2) \in \mathbb{R}_+^2$,

$$\begin{aligned} \phi_C \left(\Lambda \left(\,\cdot\, ; C^-\right), \Lambda \left(\,\cdot\, ; C^-\right)\right)(\boldsymbol{w}) = 0 &\leq \min\{w_1, w_2\} \\ &= \Lambda \left(\boldsymbol{w}\,; C^+\right) = \Lambda \left(\boldsymbol{w}\,; \phi_C \left(C^-, C^-\right)\right) \ , \end{aligned}$$

which is strict for every $\boldsymbol{w} \in (0,\infty)^2$.

Let us now study some examples investigating the behaviour of the Markov product on $\mathcal{T}_2$ for different 2-copulas C.

Example 4.1.17. *Let C be a d-copula and $\Lambda_1, \dots, \Lambda_d \in \mathcal{T}^+$, where*

$$\begin{aligned} \mathcal{T}^+ &:= \left\{ \Lambda \in \mathcal{T}_2 \;\middle|\; \partial_1 \Lambda(\boldsymbol{w}) = \alpha \mathbb{1}_{[0, \frac{\beta}{\alpha} w_2]}(w_1) \text{ for some } \alpha, \beta \in (0,1] \right\} \\ &= \left\{ \Lambda \in \mathcal{T}_2 \mid \Lambda(\boldsymbol{w}) = \min\{\alpha w_1, \beta w_2\} \text{ for some } \alpha, \beta \in (0,1] \right\} \ . \end{aligned}$$

Then

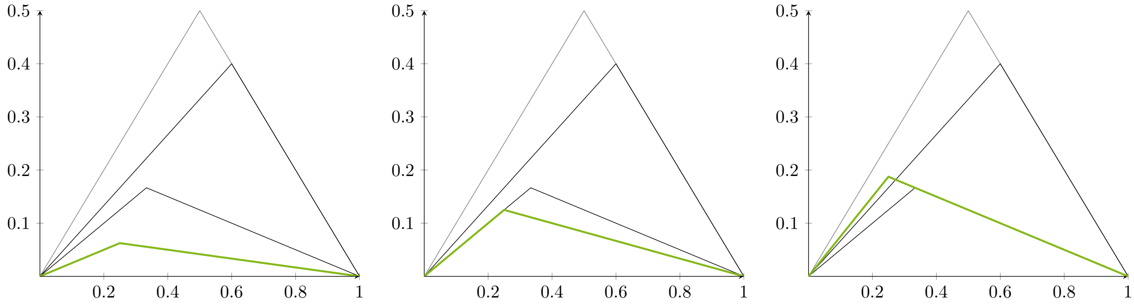
$$\phi_C (\Lambda_1, \dots, \Lambda_d)(\boldsymbol{w}) = C(\alpha_1, \dots, \alpha_d)\, \Lambda \left(\frac{\beta_1}{\alpha_1} w_1, \dots, \frac{\beta_d}{\alpha_d} w_d \,; C^+ \right) \ .$$

The influence of the choice of C on the product is depicted in Figure 4.1. For the two tail dependence functions $\Lambda_1(w_1,w_2) = \min\left\{\frac{2w_1}{3}, w_2\right\}$ and $\Lambda_2(w_1,w_2) = \min\left\{\frac{w_1}{2}, \frac{w_2}{4}\right\}$, the resulting (restricted) generalized Markov product $\phi_C(\Lambda_1, \Lambda_2)$ is shown by the green line for the choices $C = C^-$, $C = \Pi$ and $C = C^+$, respectively.

Example 4.1.18. *Taking the product of $\Lambda_1 \in \mathcal{T}^+$ and an arbitrary $\Lambda_2 \in \mathcal{T}_2$ yields*

$$\phi_C (\Lambda_1, \Lambda_2)(w_1, w_2) = \int_0^{\frac{\beta}{\alpha} w_1} C(\alpha, \partial_1 \Lambda_2(t, w_2))\, \mathrm{d}t \ .$$

The above expression can be explicitly calculated for some choices of C:



(a) $\phi_{C^-}(\Lambda_1, \Lambda_2)$. (b) $\phi_{\Pi}(\Lambda_1, \Lambda_2)$. (c) $\phi_{C^+}(\Lambda_1, \Lambda_2)$.

Figure 4.1: Plots of the product $\phi_C(\Lambda_1, \Lambda_2)(t, 1-t)$ (green line) for different choices of C as described in Example 4.1.17. The tail dependence functions $\Lambda_1(t, 1-t) = \min\left\{\frac{2t}{3}, 1-t\right\}$ and $\Lambda_2(t, 1-t) = \min\left\{\frac{t}{2}, \frac{1-t}{4}\right\}$ are depicted in black, the upper bound Λ^+ in grey.

1. *If $C = C^-$, we have*

$$\phi_{C^-}(\Lambda_1, \Lambda_2)(w_1, w_2) = \Lambda_2(p^* \wedge \frac{\beta}{\alpha} w_1, w_2) + (\alpha - 1)\Lambda\left(p^*, \frac{\beta}{\alpha} w_1\,; C^+\right)$$

since the monotonicity of $t \mapsto \partial_1 \Lambda_2(t, w_2)$ yields the existence of a $p^* = p^*(\alpha, w_2) \geq 0$ with

$$\partial_1 \Lambda_2(t, w_2) + \alpha - 1 \geq 0 \text{ for all } t \leq p^* \text{ and } \partial_1 \Lambda_2(t, w_2) + \alpha - 1 \leq 0 \text{ for all } t > p^*\,.$$

2. *If $C = \Pi$, then $\phi_\Pi(\Lambda_1, \Lambda_2)(w_1, w_2) = \Lambda_2(\beta w_1, \alpha w_2)$.*

3. *By a similar argument as in 1, for $C = C^+$, it holds*

$$\phi_{C^+}(\Lambda_1, \Lambda_2)(w_1, w_2) = \Lambda\left(\alpha p^*, \beta w_1\,; C^+\right) + \Lambda_2\left(\frac{\beta}{\alpha} w_1, w_2\right) - \Lambda_2\left(p^* \wedge \frac{\beta}{\alpha} w_1, w_2\right)\,,$$

where $p^* = p^*(1 - \alpha, w_2)$.

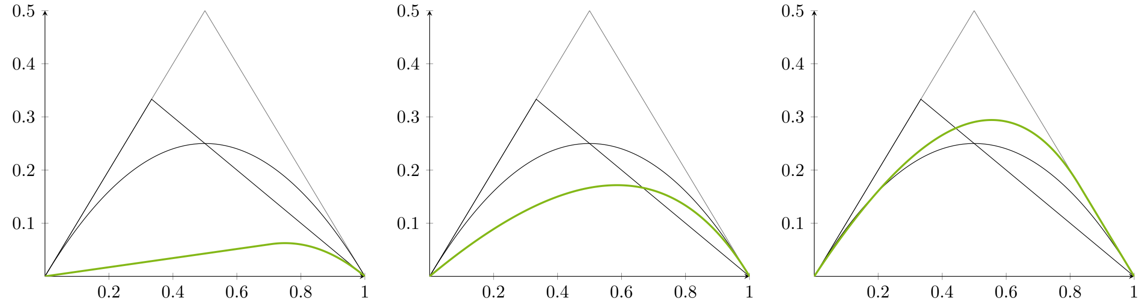
The influence of the choice of C on the product is depicted in Figure 4.2, where the resulting (restricted) generalized Markov product $\phi_C(\Lambda_1, \Lambda_2)$ is shown by the green line for the choices $C = C^-$, $C = \Pi$ and $C = C^+$, respectively.

4.2 Monotonicity of the Markov product

Figures 4.1 and 4.2 already suggest a monotonicity of the Markov product whenever C fulfils a negative dependence property. In this section, we will treat this property in more detail.

Theorem 4.2.1. *Let $\Lambda_1, \ldots, \Lambda_d \in \mathcal{T}_2$ and $C \in \mathcal{C}_d$ be negative quadrant dependent, i.e. $C(\boldsymbol{u}) \leq \Pi(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$. Then, for $k \in \{1, \ldots, d\}$ and $\boldsymbol{w} \in \mathbb{R}^d_+$,*

$$\phi_C(\Lambda_1, \ldots, \Lambda_d)(\boldsymbol{w}) \leq \phi_\Pi(\Lambda_1, \ldots, \Lambda_d)(\boldsymbol{w}) \leq \min_{\substack{\ell=1,\ldots,d \\ \ell \neq k}} \Lambda_k(w_\ell, w_k)\,.$$



(a) $\phi_{C^-}(\Lambda_1, \Lambda_2)$. (b) $\phi_\Pi(\Lambda_1, \Lambda_2)$. (c) $\phi_{C^+}(\Lambda_1, \Lambda_2)$.

Figure 4.2: Plots of the product $\phi_C(\Lambda_1, \Lambda_2)(t, 1-t)$ (green line) for different choices of C as described in Example 4.1.18. The tail dependence functions $\Lambda_1(t, 1-t) = \min\left\{t, \frac{1-t}{2}\right\}$ and $\Lambda_2(t, 1-t) = t(1-t)$ are depicted in black, the upper bound Λ^+ in grey.

This result decidedly contrasts with the behaviour of the Markov product for 2-copulas, where for example

$$C^- \leq C^+ = C^- * C^- .$$

Note that Theorem 4.2.1 is incorrect without the assumption that $C \leq \Pi$, as can be seen in Figure 4.2 (c). We will give two different proofs of Theorem 4.2.1. The first one is given below, while the second proof is deferred to Section 4.4 since it uses the theory of substochastic operators developed there.

Proof. Due to $\Lambda_1 \leq \Lambda^+$, we have

$$\int_0^t \partial_1 \Lambda_1(s, w_1) \, \mathrm{d}s \leq \int_0^t \partial_1 \Lambda^+(s, w_1) \, \mathrm{d}s$$

for all w_1 and $t \in [0, \infty)$. Hardy's Lemma (see Proposition 2.3.6 in Bennett and Sharpley (1988)) yields for any nonnegative decreasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ that

$$\int_0^\infty \partial_1 \Lambda_1(s, w_1) f(s) \, \mathrm{d}s \leq \int_0^\infty \partial_1 \Lambda^+(s, w_1) f(s) \, \mathrm{d}s = \int_0^{w_1} f(s) \, \mathrm{d}s .$$

Thus, for all tail dependence functions $\Lambda_2, \ldots, \Lambda_d$ and any $\boldsymbol{w} \in \mathbb{R}_+^d$, it holds

$$\begin{aligned} \phi_\Pi(\Lambda_1, \ldots, \Lambda_d)(\boldsymbol{w}) &= \int_0^\infty \partial_1 \Lambda_1(s, w_1) \partial_1 \Lambda_2(s, w_2) \cdots \partial_1 \Lambda_d(s, w_d) \, \mathrm{d}s \\ &\leq \int_0^{w_1} \partial_1 \Lambda_2(s, w_2) \cdots \partial_1 \Lambda_d(s, w_d) \, \mathrm{d}s \\ &= \phi_{w_1, \Pi}(\Lambda_2, \ldots, \Lambda_d)(w_2, \ldots, w_d) . \end{aligned}$$

An application of Property 4 from Proposition 4.1.5 yields the desired result. □

Corollary 4.2.2. *Let C be an idempotent 2-copula, i.e. a copula fulfilling $C * C = C$. Then for all $\boldsymbol{w} \in \mathbb{R}_+^2$,*

$$\Lambda(\boldsymbol{w}\,; C * C) \geq (\Lambda(\,\cdot\,; C) * \Lambda(\,\cdot\,; C))(\boldsymbol{w}) \ .$$

Proof. Theorem 4.2.1 in combination with $C * C = C$ immediately yields

$$\Lambda(\boldsymbol{w}\,; C * C) = \Lambda(\boldsymbol{w}\,; C) \geq (\Lambda(\,\cdot\,; C) * \Lambda(\,\cdot\,; C))(\boldsymbol{w}) \ . \qquad \square$$

The next lemma shows that the behaviour of the restricted bivariate tail dependence function $\widetilde{\Lambda}$ at zero and at one has a surprising influence on the global behaviour of Λ, connecting the derivative of $\widetilde{\Lambda}$ in zero to the behaviour of Λ near ∞.

Lemma 4.2.3. *Let Λ be a bivariate tail dependence function. Then the right-hand derivative of $\widetilde{\Lambda}$ in 0 fulfils $\widetilde{\Lambda}'(0) \in [0,1]$ and*

$$\lim_{y\to\infty} \Lambda(x,y) = \widetilde{\Lambda}'(0) \cdot x \tag{4.8}$$

holds for all $x \in \mathbb{R}_+$.

Proof. As tail dependence functions are concave, the right-hand derivative $\widetilde{\Lambda}'(0)$ exists and lies between 0 and 1. Then, for all $x \in (0,\infty)$, it holds that

$$\begin{aligned}\lim_{y\to\infty} \Lambda(x,y) &= \lim_{y\to\infty} (x+y)\Lambda\left(\frac{x}{x+y}, \frac{y}{x+y}\right) \\ &= \lim_{y\to\infty} x\frac{x+y}{x}\widetilde{\Lambda}\left(\frac{x}{x+y}\right) = \lim_{t\searrow 0} x\frac{\widetilde{\Lambda}(t)}{t} = x \cdot \widetilde{\Lambda}'(0) \ .\end{aligned}$$

The equality for $x = 0$ follows immediately. $\square$

In light of Lemma 4.2.3, we can now strengthen Theorem 4.2.1 for bivariate tail dependence functions at zero and at one. Due to the concavity of tail dependence functions and Remark 2.4.6, an application of Theorem 4.2.1 yields

$$(\widetilde{\Lambda_1 * \Lambda_2})'(0) = \lim_{s\searrow 0} \frac{(\widetilde{\Lambda_1 * \Lambda_2})(s)}{s} \leq \lim_{s\searrow 0} \frac{\min\left\{\widetilde{\Lambda_1}(s), \widetilde{\Lambda_2}(s)\right\}}{s} = \min\left\{\widetilde{\Lambda_1}'(0), \widetilde{\Lambda_2}'(0)\right\} \ ,$$

despite of Figure 4.3 suggesting a much stronger result.

Proposition 4.2.4. *For Λ_1 and $\Lambda_2 \in \mathcal{T}_2$, it holds*

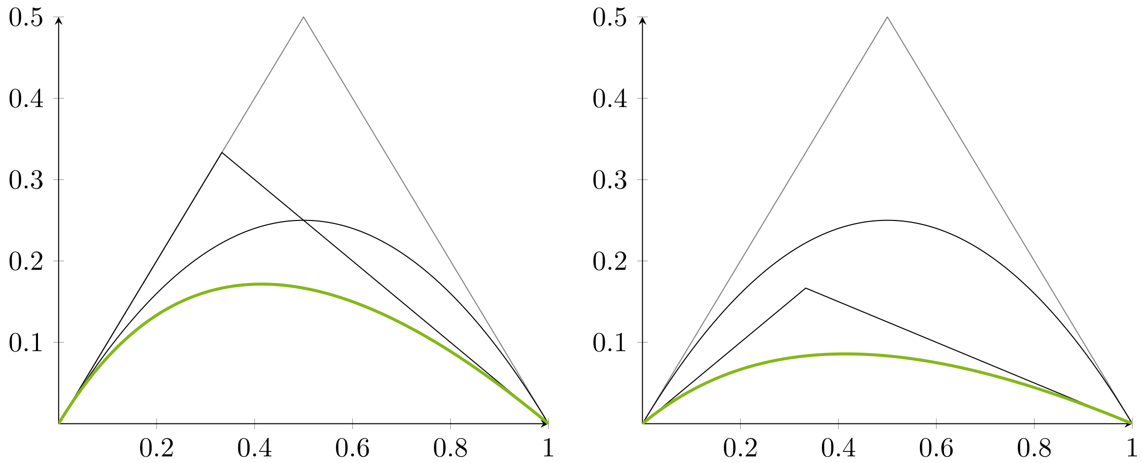
$$(\widetilde{\Lambda_1 * \Lambda_2})'(0) = \widetilde{\Lambda}_1'(0) \cdot \widetilde{\Lambda}_2'(0) \text{ and } (\widetilde{\Lambda_1 * \Lambda_2})'(1) = -\widetilde{\Lambda}_1'(1) \cdot \widetilde{\Lambda}_2'(1) \ .$$

Furthermore, for any negative quadrant dependent 2-copula C, i.e. $C \leq \Pi$, it holds

$$-\widetilde{\Lambda}_1'(1) \cdot \widetilde{\Lambda}_2'(1) \leq (\widetilde{\Lambda_1 *_C \Lambda_2})'(t) \leq \widetilde{\Lambda}_1'(0) \cdot \widetilde{\Lambda}_2'(0)$$

for all $t \in [0,1]$.

Proof. Since the proof relies on techniques developed in Section 4.4, we will again defer the proof to the end of Section 4.4. $\square$



(a) $\Lambda_1 * \Lambda_3$ (b) $\Lambda_2 * \Lambda_3$

Figure 4.3: Plots of the products $(\Lambda_1 * \Lambda_2)(t, 1-t)$ and $(\Lambda_2 * \Lambda_3)(t, 1-t)$ (green line). The tail dependence functions $\Lambda_1(t, 1-t) = \min\left\{t, \frac{1-t}{2}\right\}$, $\Lambda_2(t, 1-t) = \frac{1}{2}\min\left\{t, \frac{1-t}{2}\right\}$ and $\Lambda_3(t, 1-t) = t(1-t)$ are depicted in black, the upper bound Λ^+ in grey.

We now present a general smoothing property concerning the Markov product on $\mathcal{T}_2$, which is reminiscent of the one for the Markov product on $\mathcal{C}_2$ stated in Trutschnig (2013). Note that a convex function is differentiable on an open and convex set U if and only if it is continuously differentiable on U (see Corollary 25.5.1 in Rockafellar (1997)).

Theorem 4.2.5. *The Markov product of tail dependence functions $\Lambda_1 * \Lambda_2$ is continuously differentiable whenever Λ_1 or Λ_2 is continuously differentiable.*

To prove the above theorem, we require the following lemma, which may also be of independent interest.

Lemma 4.2.6. *Suppose $f : \mathbb{R}_+^2 \to \mathbb{R}_+$ is a positive homogeneous function of degree 1, i.e. $f(s\boldsymbol{x}) = sf(\boldsymbol{x})$ for all $s > 0$. Then, for every $\boldsymbol{x} \in (0, \infty)^2$, $\partial_1 f(\boldsymbol{x})$ exists and is finite if and only if $\partial_2 f(\boldsymbol{x})$ exists and is finite.*

Proof. Without loss of generality, suppose $\partial_1 f(\boldsymbol{x})$ exists and is finite for some $\boldsymbol{x} \in (0, \infty)^2$, otherwise consider $f^\top(x_1, x_2) := f(x_2, x_1)$. Then it holds for h small enough

$$\begin{aligned}
&\frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h} \\
&= \frac{\frac{x_2+h}{x_2} f(\frac{x_2}{x_2+h} x_1, x_2) - f(x_1, x_2)}{h} \\
&= \frac{x_2 + h}{x_2} \frac{f(x_1 - \frac{h}{x_2+h} x_1, x_2) - \frac{x_2}{x_2+h} f(x_1, x_2)}{h} \\
&= \frac{x_2 + h}{x_2} \frac{f(x_1 - \frac{h}{x_2+h} x_1, x_2) - f(x_1, x_2)}{h} + \frac{x_2 + h}{x_2} \frac{f(x_1, x_2) - \frac{x_2}{x_2+h} f(x_1, x_2)}{h} \\
&= \frac{x_2 + h}{x_2} \frac{f(x_1 - \frac{h}{x_2+h} x_1, x_2) - f(x_1, x_2)}{h} + \frac{f(x_1, x_2)}{x_2}\,.
\end{aligned}$$

Now, setting $s := \frac{h}{x_2+h}x_1$ yields

$$\begin{aligned}\frac{f(x_1,x_2+h)-f(x_1,x_2)}{h} &= \frac{x_2+h}{x_2}\frac{f(x_1-\frac{h}{x_2+h}x_1,x_2)-f(x_1,x_2)}{\frac{hx_1}{x_2+h}\frac{x_2+h}{x_1}}+\frac{f(x_1,x_2)}{x_2}\\ &= \frac{x_1}{x_2}\frac{f(x_1-s,x_2)-f(x_1,x_2)}{s}+\frac{f(x_1,x_2)}{x_2}\ .\end{aligned}$$

Thus, the assertion now follows for $h \to 0$ from

$$\partial_2 f(x_1,x_2) = \lim_{h\to 0}\frac{f(x_1,x_2+h)-f(x_1,x_2)}{h} = \frac{f(x_1,x_2)-x_1\partial_1 f(x_1,x_2)}{x_2}\ . \qquad \square$$

Proof of Theorem 4.2.5. Without loss of generality, suppose Λ_1 is differentiable. Otherwise, consider $(\Lambda_1 * \Lambda_2)^\top = \Lambda_2^\top * \Lambda_1^\top$. By Theorem 25.2 in Rockafellar (1997), $\Lambda_1 * \Lambda_2$ is differentiable in $\boldsymbol{w}$ whenever the two partial derivatives $\partial_1(\Lambda_1 * \Lambda_2)(\boldsymbol{w})$ and $\partial_2(\Lambda_1 * \Lambda_2)(\boldsymbol{w})$ exist in $\boldsymbol{w}$ and are finite. Since $\Lambda_1 * \Lambda_2$ is positive homogeneous, Lemma 4.2.6 implies that we only need to show that $\partial_1(\Lambda_1 * \Lambda_2)(\boldsymbol{w})$ exists and is finite for every $\boldsymbol{w} \in (0,\infty)^2$. Using the right-hand partial derivative of Λ_2, it holds

$$\begin{aligned}\partial_1(\Lambda_1 * \Lambda_2)(\boldsymbol{w}) &= \lim_{h\to 0}\frac{(\Lambda_1 * \Lambda_2)(w_1+h,w_2)-(\Lambda_1 * \Lambda_2)(w_1,w_2)}{h}\\ &= \lim_{h\to 0}\frac{1}{h}\left(\int_0^\infty \partial_2\Lambda_1(w_1+h,t)\partial_1^+\Lambda_2(t,w_2)\ \mathrm{d}t - \int_0^\infty \partial_2\Lambda_1(w_1,t)\partial_1^+\Lambda_2(t,w_2)\ \mathrm{d}t\right)\\ &= \lim_{h\to 0}\frac{1}{h}\left(\int_0^\infty \partial_1^+\Lambda_2(t,w_2)\ \Lambda_1(w_1+h,\mathrm{d}t) - \int_0^\infty \partial_1^+\Lambda_2(t,w_2)\ \Lambda_1(w_1,\mathrm{d}t)\right)\ ,\end{aligned}$$

due to $\Lambda_1(w_1,\ \cdot\ )$ being continuously differentiable. Here, $\Lambda_1(x,\mathrm{d}t)$ for $x \in \{w_1, w_1+h\}$ denotes the Lebesgue-Stieltjes measure. An application of integration by parts (see Hewitt (1960) or Carter and Brunt (2000)) yields

$$\int_0^\infty \partial_1^+\Lambda_2(t,w_2)\ \Lambda_1(x,\mathrm{d}t) = -\int_0^\infty \Lambda_1(x,t)\ \partial_1^+\Lambda_2(\mathrm{d}t,w_2)\ ,$$

where we used that $\partial_1^+\Lambda_2(t,w_2)$ and $\Lambda_1(x,t)$ have no common discontinuities in combination with

$$\lim_{t\to\infty}\partial_1^+\Lambda_2(t,w_2) = 0\ ,\quad \left|\partial_1^+\Lambda_2(t,w_2)\right| \le 1 \text{ and } 0 \le \Lambda_1(x,w_2) \le \min\{x,w_2\}\ .$$

Thus,

$$\begin{aligned}\partial_1(\Lambda_1 * \Lambda_2)(\boldsymbol{w}) &= \lim_{h\to 0}\frac{-1}{h}\left(\int_0^\infty \Lambda_1(w_1+h,t)\ \partial_1^+\Lambda_2(\mathrm{d}t,w_2) - \int_0^\infty \Lambda_1(w_1,t)\ \partial_1^+\Lambda_2(\mathrm{d}t,w_2)\right)\\ &= \lim_{h\to 0} - \int_0^\infty \frac{\Lambda_1(w_1+h,t)-\Lambda_1(w_1,t)}{h}\ \partial_1^+\Lambda_2(\mathrm{d}t,w_2)\\ &= -\int_0^\infty \partial_1\Lambda_1(w_1,t)\ \partial_1^+\Lambda_2(\mathrm{d}t,w_2) \in [0,1]\ .\end{aligned}$$

The last equality follows from an application of the dominated convergence theorem due to

$$\left|\frac{\Lambda_1(w_1+h,t)-\Lambda_1(w_1,t)}{h}\right| \le 1 \text{ and } \int_0^\infty 1\ \left|\partial_1^+\Lambda_2(\mathrm{d}t,w_2)\right| = \partial_1^+\Lambda_2(0,w_2) \le 1\ .$$

Note that repeating the same reasoning as above for the second component only yields that $\partial_2(\Lambda_1 * \Lambda_2)(\boldsymbol{w})$ exists almost everywhere, such that we apply Lemma 4.2.6 instead. □

While the left-inverse with respect to the Markov product for 2-copulas can be used to analyse complete dependence and extremal points of $\mathcal{C}_2$, the reduction property impedes an analogy for tail dependence functions.

Theorem 4.2.7. *Suppose Λ is a bivariate tail dependence function.*

1. *If Λ is left-invertible, i.e. if there exists a bivariate tail dependence function ξ such that $\xi * \Lambda(\boldsymbol{w}) = \Lambda(\boldsymbol{w}\,;C^+)$, then $\Lambda(\boldsymbol{w}) = \Lambda(\boldsymbol{w}\,;C^+)$.*
2. *If $\partial_1\Lambda(w_1,w_2) \in \{0,1\}$ for almost all $w_1 \in \mathbb{R}_+$, then $\Lambda(w_1,w_2) = \Lambda(w_1,\alpha w_2\,;C^+)$ for some $\alpha \in [0,1]$.*

Proof. 1. If Λ is left-invertible with left-inverse ξ, then

$$\Lambda\left(\boldsymbol{w}\,;C^+\right) = (\xi * \Lambda)(\boldsymbol{w}) \le \Lambda(\boldsymbol{w}) \le \Lambda\left(\boldsymbol{w}\,;C^+\right)\ .$$

2. Assume Λ is a tail dependence function with $\partial_1\Lambda(w_1,w_2) \in \{0,1\}$ for almost all $w_1 \in \mathbb{R}_+$. Then there exists a mapping $\alpha : [0,\infty) \to [0,1]$ such that

$$\partial_1\Lambda(w_1,w_2) = \mathbb{1}_{[0,\alpha(w_2)w_2]}(w_1)\ .$$

The positive homogeneity of Λ implies that $\partial_1\Lambda$ is positive homogeneous of order 0, i.e. constant along rays. Thus, for all $s > 0$, this leads to

$$\begin{aligned}\mathbb{1}_{[0,\alpha(sw_2)w_2]}(w_1) &= \mathbb{1}_{[0,\alpha(sw_2)sw_2]}(sw_1) = \partial_1\Lambda(sw_1,sw_2)\\ &= \partial_1\Lambda(w_1,w_2) = \mathbb{1}_{[0,\alpha(w_2)w_2]}(w_1)\ .\end{aligned}$$

Consequently, $\alpha(sw_2) = \alpha(w_2) = \alpha$. □

Lastly, we derive a monotonicity property of the Markov product with respect to the pointwise order of tail dependence functions.

Proposition 4.2.8. *For $\Lambda_1, \Lambda_2 \in \mathcal{T}_2$, the following are equivalent:*

1. $\Lambda_1(\boldsymbol{w}) \leq \Lambda_2(\boldsymbol{w})$ *for all* $\boldsymbol{w} \in \mathbb{R}_+^2$.
2. $(\Lambda_1 * \Lambda)(\boldsymbol{w}) \leq (\Lambda_2 * \Lambda)(\boldsymbol{w})$ *for all* $w \in \mathbb{R}_+^2$ *and all* $\Lambda \in \mathcal{T}_2$.

Proof. The implication 2 to 1 follows immediately from the choice $\Lambda = \Lambda^+$. Conversely, assuming $\Lambda_1(\boldsymbol{w}) \leq \Lambda_2(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathbb{R}_+^2$, we have

$$\int_0^{w_2} \partial_2 \Lambda_1(w_1, t) \, \mathrm{d}t = \Lambda_1(\boldsymbol{w}) \leq \Lambda_2(\boldsymbol{w}) = \int_0^{w_2} \partial_2 \Lambda_2(w_1, t) \, \mathrm{d}t \; .$$

Since $t \mapsto \partial_1 \Lambda(t, w_2)$ is nonnegative and decreasing for any tail dependence function $\Lambda \in \mathcal{T}_2$, Proposition 2.3.6 in Bennett and Sharpley (1988) yields

$$(\Lambda_1 * \Lambda)(\boldsymbol{w}) = \int_0^{\infty} \partial_2 \Lambda_1(w_1, t) \partial_1 \Lambda(t, w_2) \, \mathrm{d}t \leq \int_0^{\infty} \partial_2 \Lambda_2(w_1, t) \partial_1 \Lambda(t, w_2) \, \mathrm{d}t = (\Lambda_2 * \Lambda)(\boldsymbol{w}) \; . \quad \square$$

4.3 Iterates of the Markov product

In the context of 2-copulas, the concepts of iterates and idempotents of the Markov product are widely investigated (see, for instance, Darsow and Olsen (2010) or Trutschnig (2013)). To investigate these concepts in the setting of tail dependence functions, we define the n-th iterate of the Markov product for 2-copulas and tail dependence functions as

$$C^{*n} := C * C^{*(n-1)} \text{ and } \Lambda^{*n} := \Lambda * \Lambda^{*(n-1)}$$

with $C^{*0} := C^+$ and $\Lambda^{*0} := \Lambda^+$, respectively. In this section, we study the asymptotic behaviour of Λ^{*n} and apply the results to idempotents of the Markov product. First, we develop a basic understanding using two simple examples.

Example 4.3.1. *Consider a copula C with*

$$\partial_1 \Lambda \left(\boldsymbol{w}\, ; C\right) = \mathbb{1}_{[0, \alpha w_2]}(w_1) \textit{ for some } \alpha \in [0,1] \; .$$

A simple calculation yields

$$\Lambda\left(\,\cdot\,; C\right)^{*2}(\boldsymbol{w}) = \int_0^{\infty} \partial_2 \Lambda\left(w_1, t\, ; C\right) \mathbb{1}_{[0, \alpha w_2]}(t) \, \mathrm{d}t = \Lambda\left(w_1, \alpha w_2\, ; C\right)$$

and iteratively

$$\Lambda\left(\,\cdot\,; C\right)^{*n}(\boldsymbol{w}) = \Lambda\left(w_1, \alpha^{n-1} w_2\, ; C\right) \to \begin{cases} 0 & \textit{for } \alpha \in [0,1) \\ \min\{w_1, w_2\} & \textit{for } \alpha = 1 \end{cases} \; .$$

Thus, in this example, the limiting behaviour of Λ^{*n} is either given by Λ^0 or Λ^+.

The next example treats a class of tail dependence functions, which will be utilized to dominate arbitrary tail dependence functions and ultimately characterize idempotents.

Example 4.3.2. *For $0 \leq p \leq \frac{1}{2}$, define the (restricted) tail dependence function*

$$\widetilde{\Lambda}(s) := \begin{cases} s & \text{for } 0 \leq s < p \\ p & \text{for } p \leq s \leq 1-p \\ 1-s & \text{for } 1-p < s \leq 1 \end{cases} . \tag{4.9}$$

Following Remark 2.4.6, $\widetilde{\Lambda}$ can be extended to a tail dependence function on $\mathbb{R}_+^2$. A straightforward calculation yields the recurrence equation

$$\Lambda^{*(n+1)}(w_1, w_2) = (1-p)\Lambda^{*n}\left(\frac{1}{q}w_1, w_2\right) + p\Lambda^{*n}\left(qw_1, w_2\right)$$

with $q := \frac{1-p}{p}$. It can be solved in two steps. First, it holds

$$\Lambda^{*(n+1)}(w_1, w_2) = \sum_{\ell=0}^{n} a_\ell^n \Lambda\left(q^{n-2\ell}w_1, w_2\right)$$

with $a_\ell^n \in \mathbb{R}_+$ such that

$$a_0^0 = 1 \ , \ a_0^{n+1} = p^n \ , \ a_{n+1}^{n+1} = (1-p)^n \text{ and } a_\ell^{n+1} = (1-p)a_{\ell-1}^n + pa_\ell^n \text{ for } 1 \leq \ell \leq n \ .$$

The general solution to multivariate recurrences of this type was derived by Neuwirth (2001) and Mansour and Shattuck (2013) and is given by

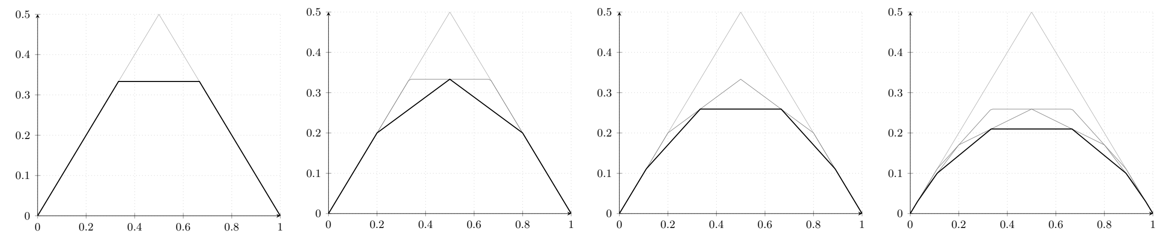
$$a_\ell^n = \binom{n}{\ell}(1-p)^\ell p^{n-\ell} \text{ for } 0 \leq \ell \leq n \ .$$

Using the positive homogeneity of Λ, we arrive at the solution

$$\Lambda^{*(n+1)}(w_1, w_2) = p^n \sum_{\ell=0}^{n} \binom{n}{\ell} \Lambda\left(q^{n-\ell}w_1, q^\ell w_2\right) \ .$$

An example of the behaviour of Λ^{*n} is shown in Figure 4.4 for different n and $p = \frac{1}{3}$. We will now derive the asymptotic behaviour of Λ^{*n} for $n \to \infty$. Due to the iterated Markov product being symmetric and due to the monotonicity of $*$, it suffices to consider $w_1 = w_2 = \frac{1}{2}$ and uneven $n = 2k+1$. Then, it holds

$$\begin{aligned} \Lambda^{*(2k+1)}\left(\frac{1}{2}, \frac{1}{2}\right) &= \frac{p^{2k}}{2} \sum_{\ell=0}^{2k} \binom{2k}{\ell} \Lambda\left(q^{2k-\ell}, q^\ell\right) \\ &= \frac{p^{2k}}{2} \sum_{\ell=0}^{2k} \binom{2k}{\ell} \left(q^{2k-\ell} + q^\ell\right) \widetilde{\Lambda}\left(\frac{q^{2k-\ell}}{q^{2k-\ell}+q^\ell}\right) \\ &\leq p^{2k} \sum_{\ell=0}^{k} \binom{2k}{\ell} q^\ell - p^{2k+1} q^k \binom{2k}{k} \\ &= \sum_{\ell=0}^{k} \binom{2k}{\ell} (1-p)^\ell p^{2k-\ell} - \binom{2k}{k} p^{k+1}(1-p)^k \ , \end{aligned}$$



(a) $\widetilde{\Lambda}(s)$ (b) $\widetilde{\Lambda^{*2}}(s)$ (c) $\widetilde{\Lambda^{*3}}(s)$ (d) $\widetilde{\Lambda^{*5}}(s)$

Figure 4.4: Plots of the tail dependence function Λ from Equation (4.9) and its iterations Λ^{*n} for $n = 2, 3$ and 5 and $p = \frac{1}{3}$.

where the inequality is due to the definition of $\widetilde{\Lambda}(s)$ and equality holds in case of $p = 1/2$. While the second part converges to zero as $n \to \infty$, the first part is a truncated binomial sum and by the weak law of large numbers, we have

$$\begin{aligned} \lim_{k\to\infty} \max_{w_1+w_2=1} \Lambda^{*(2k+1)}(w_1, w_2) &= \lim_{k\to\infty} \Lambda^{*(2k+1)}\left(\frac{1}{2}, \frac{1}{2}\right) \\ &\le \lim_{k\to\infty} \sum_{\ell=0}^{k} \binom{2k}{\ell}(1-p)^{\ell} p^{2k-\ell} - \binom{2k}{k} p^{k+1}(1-p)^{k} \\ &= \begin{cases} 0 & \text{for } p < \frac{1}{2} \\ \frac{1}{2} & \text{for } p = \frac{1}{2} \end{cases}. \end{aligned}$$

Due to $0 \le \Lambda^{*(2k+1)}$, the above inequality is in fact an equality.

Using the monotonicity property of the Markov product from Proposition 4.2.8 and the fact that the previous examples dominate any tail dependence function, we arrive at the following result.

Theorem 4.3.3. *Let Λ be a bivariate tail dependence function. Then*

$$\lim_{n\to\infty} \Lambda^{*n}(\boldsymbol{w}) = \begin{cases} \Lambda(\boldsymbol{w}\,; C^+) & \text{for } \Lambda = \Lambda(\,\cdot\,; C^+) \\ \Lambda(\boldsymbol{w}\,; \Pi) & \text{for } \Lambda \neq \Lambda(\,\cdot\,; C^+) \end{cases}.$$

This result gives another indication that the Markov product has smoothing properties, as tail independence, i.e. $\Lambda(\boldsymbol{w}\,; C) = 0$, corresponds to Fréchet-differentiability of C in zero.

Proof. If $\Lambda = \Lambda^+$, the result is immediate. Thus, consider a tail dependence function Λ with $\Lambda \neq \Lambda^+$. Define

$$p := \max_{t\in[0,1]} \widetilde{\Lambda}(t) < \frac{1}{2}$$

and set

$$\widetilde{\Lambda}_p(s) := \begin{cases} s & \text{for } 0 \le s < p \\ p & \text{for } p \le s \le 1-p \\ 1-s & \text{for } 1-p < s \le 1 \end{cases}.$$

Thus, Λ_p dominates Λ, i.e. $\Lambda \leq \Lambda_p$, and Proposition 4.2.8 yields by induction

$$\Lambda^{*n}(\boldsymbol{w}) = \Lambda * \Lambda^{*(n-1)}(\boldsymbol{w}) \leq \Lambda_p * \Lambda^{*(n-1)}(\boldsymbol{w}) \leq \Lambda_p^{*n}(\boldsymbol{w}) \to 0$$

for any $p < \frac{1}{2}$ as seen in Example 4.3.2. □

Theorem 4.3.3 has two immediate corollaries, one regarding idempotent tail dependence functions, and the other considering the tail behaviour of idempotent copulas.

Corollary 4.3.4. *A bivariate tail dependence function $\Lambda \in \mathcal{T}_2$ is idempotent, i.e. $\Lambda * \Lambda = \Lambda$, if and only if $\Lambda = \Lambda^+$ or $\Lambda = \Lambda^0$.*

Proof. If Λ is idempotent, we have

$$\Lambda(\boldsymbol{w}) = \lim_{n\to\infty} \Lambda^{*n}(\boldsymbol{w}) = \begin{cases} \Lambda\left(\boldsymbol{w}\,; C^+\right) & \text{for } \Lambda = \Lambda\left(\,\cdot\,; C^+\right) \\ \Lambda\left(\boldsymbol{w}\,; \Pi\right) & \text{for } \Lambda \neq \Lambda\left(\,\cdot\,; C^+\right) \end{cases} .$$

Conversely, Λ^0 and Λ^+ are idempotent. □

Finally, we link the previous results to the tail behaviour of idempotent 2-copulas.

Corollary 4.3.5. *Suppose C is a twice continuously differentiable idempotent 2-copula with a strict tail dependence function. Then $\Lambda\left(\boldsymbol{w}\,; C\right) = \Lambda\left(\boldsymbol{w}\,; C^+\right)$.*

Proof. As C is twice differentiable with strict tail dependence function, it holds

$$\Lambda\left(\boldsymbol{w}\,; C\right) = \Lambda\left(\boldsymbol{w}\,; C * C\right) = \Lambda\left(\,\cdot\,; C\right) * \Lambda\left(\,\cdot\,; C\right)(\boldsymbol{w}) .$$

Thus $\Lambda\left(\boldsymbol{w}\,; C\right)$ is idempotent. The assertion now follows from $\Lambda\left(\boldsymbol{w}\,; C\right) \neq \Lambda\left(\boldsymbol{w}\,; \Pi\right)$, as $\Lambda\left(\boldsymbol{w}\,; C\right)$ is a strict tail dependence function. □

4.4 Substochastic operators

We previously saw the close resemblance between the set of 2-copulas and the set of bivariate tail dependence functions endowed with their respective Markov products. For the set of 2-copulas, Olsen et al. (1996) derived an isomorphism to integral-preserving linear operators. Along those lines, we will subsequently draw a connection between a certain class of linear operators and bivariate tail dependence functions. For this, we define the underlying space

$$L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+) := \left\{ g + h \;\middle|\; g \in L^1(\mathbb{R}_+) \text{ and } h \in L^\infty(\mathbb{R}_+) \right\} ,$$

where both $L^1(\mathbb{R}_+)$ and $L^\infty(\mathbb{R}_+)$ are subsets of $L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+)$. Furthermore, we can equip $L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+)$ with a norm based on its two constituent spaces (see Bennett and Sharpley (1988))

$$\|f\|_{L^1+L^\infty} := \inf\left\{ \|g\|_1 + \|h\|_\infty \;\middle|\; f = g + h \text{ with } g \in L^1(\mathbb{R}_+) \text{ and } h \in L^\infty(\mathbb{R}_+) \right\} .$$

Definition 4.4.1. *A linear operator $T : L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+) \to L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+)$ is called doubly substochastic if*

1. *T is positive, i.e. $Tf \geq 0$ whenever $f \geq 0$.*

2. *$T(L^1(\mathbb{R}_+)) \subset L^1(\mathbb{R}_+)$ and $T(L^\infty(\mathbb{R}_+)) \subset L^\infty(\mathbb{R}_+)$.*

3. *T is nonexpansive on $L^1(\mathbb{R}_+)$ and on $L^\infty(\mathbb{R}_+)$, i.e. $\|Tg\|_1 \leq \|g\|_1$ and $\|Th\|_\infty \leq \|h\|_\infty$ for all $g \in L^1(\mathbb{R}_+)$ and all $h \in L^\infty(\mathbb{R}_+)$.*

4. *For all positive $h \in L^\infty(\mathbb{R}_+)$ and $h_n \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ such that $h_n \nearrow h$ pointwise, we have*
$$Th = \sup_{n \in \mathbb{N}} Th_n \ .$$
For arbitrary $h \in L^\infty(\mathbb{R}_+)$, we have $Th = Th^+ - Th^-$, where $h^\pm := \max\{\pm h, 0\}$.

We call T equivariant if
$$T(f \circ \sigma) = (Tf) \circ \sigma$$
holds for all dilations $\sigma(x) := \frac{x}{s}$ with $s > 0$.

Substochastic operators can be seen as a generalization of Markov operators, in the same way as doubly substochastic matrices generalize doubly stochastic matrices. Property 4 of Definition 4.4.1 is a technical requirement to ensure the unique continuation from $L^p(\mathbb{R}_+)$ to $L^\infty(\mathbb{R}_+)$ for $1 \leq p < \infty$. This property is often used in the study of (sub-)Markovian operators and semigroups (see, for example, Section 1.6 in Fukushima, Oshima and Takeda (2010)) but unnecessary in the case of 2-copulas and bounded domains. A comprehensive introduction to substochastic operators can be found in Bennett and Sharpley (1988).

In the following, we establish a one-to-one correspondence between substochastic operators and subdistribution functions. While many of the proofs work similarly to the case of compact spaces in Olsen et al. (1996), some modifications are needed due to the underlying nonfiniteness of the measure space $\mathbb{R}_+$.

Definition 4.4.2. *A function $F : \mathbb{R}_+^d \to \mathbb{R}_+$ is called a subdistribution function if it is positive, d-increasing, bounded by Λ^+ and Lipschitz continuous with Lipschitz constant 1 with respect to the ℓ_1-norm on $\mathbb{R}^d$.*

Subdistribution functions constitute a generalization of tail dependence functions, for which the condition of positive homogeneity is relaxed.

Remark 4.4.3. *The class of d-variate tail dependence functions exactly corresponds to the class of positive homogeneous subdistribution functions.*

Lemma 4.4.4. *Let T be a doubly substochastic operator. Then*
$$F_T(x, y) := \int_0^x T\mathbb{1}_{[0,y]}(s) \ \mathrm{d}s$$
is a bivariate subdistribution function. If T is additionally equivariant, then F_T is a bivariate tail dependence function, i.e. $F_T(\,\cdot\,) = \Lambda(\,\cdot\,; C)$ for some 2-copula C.

Proof. 1. Because $0 \le F_T$ is immediate for positive T, we only need to show that F_T is bounded from above by $\Lambda(\,\cdot\,; C^+)$:

$$\int_0^x T\mathbb{1}_{[0,y]}(s)\ \mathrm{d}s \le \begin{cases} \int_0^\infty T\mathbb{1}_{[0,y]}(s)\ \mathrm{d}s \le \int_0^\infty \mathbb{1}_{[0,y]}(s)\ \mathrm{d}s = y \\ \int_0^x T\mathbb{1}_{\mathbb{R}_+}(s)\ \mathrm{d}s \le x \end{cases} .$$

2. Let $R = [x_1, x_2) \times [y_1, y_2)$ with $x_1 < x_2$ and $y_1 < y_2$. Then the linearity of T yields

$$V_{F_T}(R) = \int_{x_1}^{x_2} T\mathbb{1}_{[y_1,y_2]}(s)\ \mathrm{d}s \ge 0 .$$

3. Finally, we check the Lipschitz continuity of F_T with Lipschitz constant 1. To do so, let x_1, x_2, y_1 and $y_2 \in \mathbb{R}_+$. Then

$$\begin{aligned} |F_T(x_2,y_2) - F_T(x_1,y_1)| &\le |F_T(x_2,y_2) - F_T(x_1,y_2)| + |F_T(x_1,y_2) - F_T(x_1,y_1)| \\ &\le \int_{\min\{x_1,x_2\}}^{\max\{x_1,x_2\}} T\mathbb{1}_{[0,y_2]}(s)\ \mathrm{d}s + \int_0^{x_1} T\mathbb{1}_{[\min\{y_1,y_2\},\max\{y_1,y_2\}]}(s)\ \mathrm{d}s \\ &\le \left\|T\mathbb{1}_{[0,y_2]}\right\|_\infty |x_2 - x_1| + \left\|T\mathbb{1}_{[\min\{y_1,y_2\},\max\{y_1,y_2\}]}\right\|_1 \\ &\le \left\|\mathbb{1}_{[0,y_2]}\right\|_\infty |x_2 - x_1| + \left\|\mathbb{1}_{[\min\{y_1,y_2\},\max\{y_1,y_2\}]}\right\|_1 \\ &= |x_2 - x_1| + |y_2 - y_1| \ . \end{aligned}$$

Hence, F_T is a bivariate subdistribution function. Finally, the positive homogeneity of F_T follows from

$$F_T(sx, sy) = \int_0^{sx} T\mathbb{1}_{[0,sy]}(t)\ \mathrm{d}t = \int_0^{sx} T\mathbb{1}_{[0,y]}\left(\frac{t}{s}\right)\ \mathrm{d}t = \int_0^x T\mathbb{1}_{[0,y]}(z)\, s\ \mathrm{d}z = sF_T(x,y)$$

for any $s > 0$. Thus, F_T is a positive homogeneous, bounded and 2-increasing function, and the claim follows from Proposition 2.4.4. □

Lemma 4.4.5. *Let F be a bivariate subdistribution function. Then*

$$\begin{aligned} T_F: \quad & L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+) \to L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+) \\ & T_F f(x) = \partial_x \int_0^\infty \partial_2 F(x,t) f(t)\ \mathrm{d}t \end{aligned}$$

defines a doubly substochastic operator. Moreover, if F is a bivariate tail dependence function, then T_F is equivariant.

Sketch of the proof of Lemma 4.4.5. The proof can be outlined as follows:

1. We show that
$$Tf(x) := \partial_x \int_0^\infty \partial_2 F(x,t) f(t)\ \mathrm{d}t \tag{4.10}$$
defines an operator from $L^\infty(\mathbb{R}_+)$ to $L^\infty(\mathbb{R}_+)$. Furthermore, T fulfils the (adapted) Properties 1 to 4 of Definition 4.4.1, where T is nonexpansive with respect to $\|\cdot\|_1$ for $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ only. Furthermore, T is equivariant whenever F is positive homogeneous.

2. We extend (the restricted operator) $T : L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+) \to L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ uniquely to a continuous linear operator $T_1 : L^1(\mathbb{R}_+) \to L^1(\mathbb{R}_+)$. The operator T_1 is positive, equivariant, nonexpansive with respect to $\|\cdot\|_1$ and again follows the form given in (4.10).

Combining 1 and 2, we define for $f \in L^1(\mathbb{R}_+) + L^\infty(\mathbb{R}_+)$
$$T_F f := T_1 g + Th = \partial_x \int_0^\infty \partial_2 F(x,t) f(t)\ \mathrm{d}t\ ,$$
where $f = g + h$ for some $g \in L^1(\mathbb{R}_+)$ and $h \in L^\infty(\mathbb{R}_+)$. By construction, T and T_1 agree on $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$, yielding that T_F is independent of the choice of g and h. In particular, T_F fulfils Properties 1 to 4 of Definition 4.4.1.

Part 1 of the proof of Lemma 4.4.5. Since $x \mapsto \partial_2 F(x,t)$ is increasing for all fixed $t \in \mathbb{R}_+$, we have for any $f \in L^\infty(\mathbb{R}_+)$ that $|f| \pm f \geq 0$ and thus
$$\int_0^\infty \partial_2 F(x,t)\,(|f| \pm f)\,(t)\ \mathrm{d}t$$
is again an increasing function in x and its derivative exists almost everywhere. Thus, representing f as a linear combination of $|f| + f$ and $|f| - f$ implies that Tf exists. Let us now verify the Properties 1 to 4 of Definition 4.4.1 adapted to $L^\infty(\mathbb{R}_+)$.

a) Let f be positive. As $\partial_2 F(x_2,t) - \partial_2 F(x_1,t) \geq 0$ for $x_1 \leq x_2$, we have that
$$\int_0^\infty \partial_2 F(x_2,t) f(t)\ \mathrm{d}t - \int_0^\infty \partial_2 F(x_1,t) f(t)\ \mathrm{d}t \geq 0$$
and hence $Tf \geq 0$.

b) To prove Properties 2 and 3, we note that
$$g(x) := \int_0^\infty \partial_2 F(x,t) f(t)\ \mathrm{d}t$$

is Lipschitz continuous with Lipschitz constant $L = \|f\|_\infty$: For $x_1 \le x_2$, we have

$$\begin{aligned} |g(x_2) - g(x_1)| &= \left| \int_0^\infty (\partial_2 F(x_2,t) - \partial_2 F(x_1,t))\, f(t) \, \mathrm{d}t \right| \\ &\le \|f\|_\infty \int_0^\infty |\partial_2 F(x_2,t) - \partial_2 F(x_1,t)| \, \mathrm{d}t \\ &= \|f\|_\infty \int_0^\infty (\partial_2 F(x_2,t) - \partial_2 F(x_1,t)) \, \mathrm{d}t \\ &= \|f\|_\infty \lim_{R\to\infty} [F(x_2,t) - F(x_1,t)]_0^R \le \|f\|_\infty |x_2 - x_1| \ , \end{aligned} \tag{4.11}$$

where the second equality is due to $x \mapsto \partial_2 F(x,t)$ being increasing since F is 2-increasing. The last inequality follows from the Lipschitz continuity and grounding of F. Thus, T is nonexpansive on $L^\infty(\mathbb{R}_+)$.
Now let f be in $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$. Combining the linearity and positivity of T leads to

$$|Tf| = \left|T(f^+ - f^-)\right| \le Tf^+ + Tf^- = T|f| \ .$$

Thus, without loss of generality, let f be positive. Using the absolute continuity of g, we have

$$\begin{aligned} \int_0^\infty Tf(x)\, \mathrm{d}x &= \int_0^\infty \partial_x \int_0^\infty \partial_2 F(x,t) f(t) \, \mathrm{d}t \, \mathrm{d}x = \lim_{R\to\infty} \int_0^R \partial_x \int_0^\infty \partial_2 F(x,t) f(t) \, \mathrm{d}t \, \mathrm{d}x \\ &= \lim_{R\to\infty} \int_0^\infty \partial_2 F(R,t) f(t) \, \mathrm{d}t \le \lim_{R\to\infty} \int_0^\infty f(t) \, \mathrm{d}t = \int_0^\infty f(t) \, \mathrm{d}t \ , \end{aligned}$$

due to $0 \le \partial_2 F(R,t) \le 1$ and therefore T is nonexpansive on $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ with $\|Tf\|_1 \le \|f\|_1$.

c) It remains to show Property 4. To do so, we assume $f \ge 0$, otherwise one can use the decomposition $f = f^+ - f^-$ and treat f^+ and f^- separately. We first choose $f_n := f \mathbb{1}_{[0,n]} \nearrow f$ and set $h_n := f - f_n \searrow 0$. As $Th_n \ge 0$ is a decreasing sequence due to the positivity of T, it converges towards a measurable $h \ge 0$. Moreover,

$$g_n(x) := \int_0^\infty \partial_2 F(x,t) h_n(t) \, \mathrm{d}t = \int_n^\infty \partial_2 F(x,t) f(t) \, \mathrm{d}t \to 0 \ ,$$

as $\partial_2 F(x,t) \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ and $f \in L^\infty(\mathbb{R}_+)$. Thus, for all $x \in \mathbb{R}_+$, we have by the absolute continuity of g_n

$$0 = \lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \int_0^x g_n'(t)\ \mathrm{d}t = \lim_{n\to\infty} \int_0^x Th_n(t)\ \mathrm{d}t$$
$$= \int_0^x \lim_{n\to\infty} Th_n(t)\ \mathrm{d}t = \int_0^x h(t)\ \mathrm{d}t \geq 0\ ,$$

where the exchange of the integral and limit is legitimate due to an application of the monotone convergence theorem. Therefore, $h = 0$ holds almost everywhere and

$$T(f - f_n) = Th_n \searrow 0 \implies Tf_n \nearrow Tf\ .$$

Finally, suppose $f_n, g_k \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ and $f_n \nearrow f$ and $g_k \nearrow f$. Due to $\min\{f_n, g_k\} \nearrow f_n$ as $k \to \infty$, it holds that

$$Tf_n = \sup_{k\in\mathbb{N}} T_F(\min\{f_n, g_k\}) \leq \sup_{k\in\mathbb{N}} Tg_k\ .$$

Switching the roles of f_n and g_k yields the independence from the approximating sequence.

Combining the previous three results, one sees that T indeed constitutes an operator from $L^\infty(\mathbb{R}_+)$ onto $L^\infty(\mathbb{R}_+)$ which fulfils the adapted Properties 1 to 4. If F is also positive homogeneous, then for any $s > 0$ and $\sigma(x) := \frac{x}{s}$

$$T(f\circ\sigma)(x) = \partial_x \int_0^\infty \partial_2 F(x,t) f\left(\frac{t}{s}\right)\ \mathrm{d}t$$
$$= \partial_x \int_0^\infty \partial_2 F(x, sz) f(z)\, s\ \mathrm{d}z$$
$$= \partial_x \int_0^\infty \partial_2 F\left(\frac{x}{s}, z\right) f(z)\ \mathrm{d}z = Tf\left(\frac{x}{s}\right)\ .$$

Part 2 of the proof of Lemma 4.4.5. It remains to show that Equation (4.10) also defines an operator from $L^1(\mathbb{R}_+)$ to $L^1(\mathbb{R}_+)$. Crucially, $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ is dense in $L^1(\mathbb{R}_+)$ with respect to $\|\cdot\|_1$ and T is nonexpansive on $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ with respect to $\|\cdot\|_1$. Thus, using the continuous linear extension theorem (see, e.g., Theorem 1.9.1 in Megginson (1998)), we can extend T uniquely from $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ to $L^1(\mathbb{R}_+)$ via

$$T_1 f := \lim_{n\to\infty} Tf_n\ ,$$

where $f \in L^1(\mathbb{R}_+)$ and $f_n \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$ such that $\|f - f_n\|_1 \to 0$. It follows from the properties of T that T_1 is positive, nonexpansive on $L^1(\mathbb{R}_+)$ and equivariant. An application of the representation theorem 2.3.9 in Dunford and Pettis (1940) then yields

$$T_1 f(x) = \partial_x Sf(x) := \partial_x \int_0^\infty K(x,t) f(t)\ \mathrm{d}t$$

with a kernel function $K : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ such that, among other properties, $Sf(x)$ is absolutely continuous and $K(x, \cdot\,) \in L^\infty(\mathbb{R}_+)$. Using the Lipschitz continuity of F in the first component, we have

$$\begin{aligned} F(x,y) &= \int_0^x \partial_1 F(t,y)\ \mathrm{d}t = \int_0^x T\mathbb{1}_{[0,y]}(t)\ \mathrm{d}t = \int_0^x T_1\mathbb{1}_{[0,y]}(t)\ \mathrm{d}t \\ &= \int_0^x \partial_t S\mathbb{1}_{[0,y]}(t)\ \mathrm{d}t = \int_0^\infty K(x,t)\mathbb{1}_{[0,y]}(t)\ \mathrm{d}t = \int_0^y K(x,t)\ \mathrm{d}t \end{aligned}$$

Since F is also absolutely continuous in the second component, it holds $\partial_2 F(x,y) = K(x,y)$ for all $x \in \mathbb{R}_+$ and almost all $y \in \mathbb{R}_+$. Thus, T_1 takes again the form (4.10). □

As a consequence of Lemmas 4.4.4 and 4.4.5, we obtain our main result, establishing the correspondence between subdistribution functions and substochastic operators.

Theorem 4.4.6. *Let F be a bivariate subdistribution function and T a substochastic operator, and define $\Phi(T) := F_T$ and $\Psi(F) := T_F$. Then $\Phi \circ \Psi$ and $\Psi \circ \Phi$ define identities on their respective spaces. Furthermore, F is positive homogeneous if and only if T_F is equivariant.*

Proof. Starting with a subdistribution function F, we can use its Lipschitz continuity to obtain

$$\begin{aligned} (\Phi \circ \Psi(F))(x,y) &= \int_0^x \Psi(F)\mathbb{1}_{[0,y]}(s)\ \mathrm{d}s = \int_0^x \partial_s \int_0^\infty \partial_2 F(s,t)\mathbb{1}_{[0,y]}(t)\ \mathrm{d}t\ \mathrm{d}s \\ &= \int_0^x \partial_s \int_0^y \partial_2 F(s,t)\ \mathrm{d}t\ \mathrm{d}s = \int_0^x \partial_s F(s,y)\ \mathrm{d}s = F(x,y)\ . \end{aligned}$$

Conversely, for the function $f(t) = \mathbb{1}_{[0,y]}(t)$, the absolute continuity shown in the proof of Lemma 4.4.4 yields

$$\begin{aligned} (\Psi \circ \Phi(T))f(x) &= \partial_x \int_0^\infty \partial_2 \Phi(T)(x,t)f(t)\ \mathrm{d}t = \partial_x \int_0^\infty \partial_t \int_0^x T\mathbb{1}_{[0,t]}(s)\ \mathrm{d}s f(t)\ \mathrm{d}t \\ &= \partial_x \int_0^y \partial_t \int_0^x T\mathbb{1}_{[0,t]}(s)\ \mathrm{d}s\ \mathrm{d}t = \partial_x \int_0^x T\mathbb{1}_{[0,y]}(s)\ \mathrm{d}s = T\mathbb{1}_{[0,y]}(x)\ . \end{aligned}$$

Thus, $\Psi \circ \Phi(T)$ and T are substochastic operators which agree for every $\mathbb{1}_{[0,y]}$ and therefore agree on $L^1(\mathbb{R}_+)$. Property 4 of Definition 4.4.1 now ensures that T and $\Psi \circ \Phi$ coincide on $L^1(\mathbb{R}_+)+L^\infty(\mathbb{R}_+)$. Finally, Lemmas 4.4.4 and 4.4.5 yield the equivalence between the positive homogeneity of F and the equivariance of T_F. □

The correspondence between substochastic operators and subdistribution functions is a structure-preserving isomorphism translating $*$ into $\circ$ and vice versa. To verify the structure preservation property, we need to introduce a slight generalization of the previously introduced Markov product for tail dependence functions. The Markov product of two subdistribution functions F and G is analogously defined for all $\boldsymbol{w} \in \mathbb{R}_+^2$ as

$$(F * G)(\boldsymbol{w}) = \int_0^\infty \partial_2 F(w_1, t) \cdot \partial_1 G(t, w_2)\ \mathrm{d}t\ .$$

We want to show that $F * G$ is again a subdistribution function. Applying Remark 4.1.4, it only remains to show the Lipschitz continuity of $F * G$, which follows as in Equation (4.11).

Theorem 4.4.7. *Let F and G be subdistribution functions. Then*

$$T_{F*G} = T_F \circ T_G\ .$$

Proof. In view of Theorem 4.4.6, it suffices to prove that

$$\Phi\left(T_F \circ T_G\right)(\boldsymbol{w}) = \Phi\left(T_{F*G}\right)(\boldsymbol{w}) = (F * G)(\boldsymbol{w})$$

for all $\boldsymbol{w} \in \mathbb{R}_+^2$. To do so, we use the Lipschitz continuity to obtain

$$\begin{aligned}
\Phi\left(T_F \circ T_G\right)(\boldsymbol{w}) &= \int_0^{w_1} (T_F \circ T_G)\mathbb{1}_{[0,w_2]}(s)\ \mathrm{d}s = \int_0^{w_1} \partial_s \int_0^\infty \partial_2 F(s,t) T_G \mathbb{1}_{[0,w_2]}(t)\ \mathrm{d}t\ \mathrm{d}s \\
&= \int_0^\infty \partial_2 F(w_1, t) T_G \mathbb{1}_{[0,w_2]}(t)\ \mathrm{d}t \\
&= \int_0^\infty \partial_2 F(w_1, t) \partial_t \int_0^\infty \partial_2 G(t,s) \mathbb{1}_{[0,w_2]}(s)\ \mathrm{d}s\ \mathrm{d}t \\
&= \int_0^\infty \partial_2 F(w_1, t) \partial_1 G(t, w_2)\ \mathrm{d}t = (F * G)(\boldsymbol{w})\ .
\end{aligned}$$

□

The next result establishes a connection between the doubly substochastic operator of F and the doubly substochastic operator of its transpose $F^\top$ where $F^\top(x,y) := F(y,x)$.

Proposition 4.4.8. *Let $f \in L^\infty(\mathbb{R}_+)$, $g \in L^1(\mathbb{R}_+)$ and F be a bivariate subdistribution function. Then*

$$\int_0^\infty (T_F f)(x) g(x)\ \mathrm{d}x = \int_0^\infty f(x) T_{F^\top} g(x)\ \mathrm{d}x\ .$$

Proof. As the space of compactly supported and smooth functions is dense in $L^1(\mathbb{R}_+)$, we only need to show the desired result for $g \in \mathcal{C}_0^\infty(\mathbb{R}_+)$. A calculation identical to the proof of Lemma 2.4 from Olsen et al. (1996) yields the result, except that for the integration by parts formula, we apply that g is compactly supported as well as $\partial_2 F(0,t) = 0$, which holds due to $F(0,t) \equiv 0$. □

Using this connection between the 'adjoint' of T and the transpose of F, we can establish a relation between strict subdistribution functions and Markov operators.

Definition 4.4.9. *Let F be a bivariate subdistribution function. Then we call F strict if*

$$\lim_{t\to\infty} F(w_1, t) = w_1 \quad \text{and} \quad \lim_{t\to\infty} F(t, w_2) = w_2$$

for all $(w_1, w_2) \in \mathbb{R}_+^2$.

Definition 4.4.10. *Let T be a doubly substochastic operator. T is called a doubly stochastic operator or Markov operator if*

$$T\mathbb{1}_{\mathbb{R}_+} = \mathbb{1}_{\mathbb{R}_+} \quad \text{and} \quad \int_0^\infty Tf(x)\ \mathrm{d}x = \int_0^\infty f(x)\ \mathrm{d}x$$

for all $f \in L^1(\mathbb{R}_+)$.

Proposition 4.4.11. *Let F be a bivariate subdistribution function. Then F is strict if and only if T_F and $T_{F^\top}$ are Markov operators.*

Proof. First, let F be strict. Then,

$$\begin{aligned} T_F\mathbb{1}_{\mathbb{R}_+}(x) &= \partial_x \int_0^\infty \partial_2 F(x,t)\mathbb{1}_{\mathbb{R}_+}(t)\ \mathrm{d}t = \partial_x \int_0^\infty \partial_2 F(x,t)\ \mathrm{d}t \\ &= \partial_x \Big(\lim_{t\to\infty} F(x,t) - F(x,0)\Big) = \partial_x x = \mathbb{1}_{\mathbb{R}_+}(x) \end{aligned}$$

for all $x \in \mathbb{R}_+$. Now let f be in $L^1(\mathbb{R}_+)$, then it holds

$$\begin{aligned} \int_0^\infty f(x)\ \mathrm{d}x = \int_0^\infty \mathbb{1}_{\mathbb{R}_+}(x)f(x)\ \mathrm{d}x &= \int_0^\infty (T_{F^\top}\mathbb{1}_{\mathbb{R}_+})(x)f(x)\ \mathrm{d}x \\ &= \int_0^\infty \mathbb{1}_{\mathbb{R}_+}(x)T_F f(x)\ \mathrm{d}x = \int_0^\infty T_F f(x)\ \mathrm{d}x \end{aligned}$$

using Proposition 4.4.8 and the strictness of $F^\top$. The claims for $T_{F^\top}$ can be proven analogously. Conversely, if T_F is doubly stochastic, then

$$\lim_{t\to\infty} F(t, w_2) = \lim_{t\to\infty} \int_0^t T_F\mathbb{1}_{[0,w_2]}(s)\ \mathrm{d}s = \int_0^\infty T_F\mathbb{1}_{[0,w_2]}(s)\ \mathrm{d}s = \int_0^\infty \mathbb{1}_{[0,w_2]}(s)\ \mathrm{d}s = w_2$$

and, analogously, $\lim_{t\to\infty} F(w_1, t) = w_1$ whenever $T_{F^\top}$ is doubly stochastic. □

Finally, using the theory of substochastic operators, we present an alternative proof of Theorem 4.2.1 and a proof of Proposition 4.2.4.

Proof of Theorem 4.2.1. For every substochastic operator T and every $t \geq 0$, it holds

$$\int_0^t (Tf)^*(s)\ \mathrm{d}s \leq \int_0^t f^*(s)\ \mathrm{d}s$$

or, in short, $Tf \prec f$, where f^* denotes the decreasing rearrangement of f (see Chapter 1 in Bennett and Sharpley (1988)). Thus,

$$\begin{aligned}\partial_1(\Lambda_1 * \Lambda_2)(w_1, w_2) &= \partial_1 \int_0^\infty \partial_2\Lambda_1(w_1, s)\partial_1\Lambda_2(s, w_2)\ \mathrm{d}s \\ &= T_{\Lambda_1}\partial_1\Lambda_2(\,\cdot\,, w_2)(w_1) \\ &\prec \partial_1\Lambda_2(w_1, w_2),\end{aligned}$$

and together with the concavity of the tail dependence function, the assertion follows from

$$(\Lambda_1 * \Lambda_2)(w_1, w_2) = \int_0^{w_1} \partial_1(\Lambda_1 * \Lambda_2)(t, w_2)\ \mathrm{d}t \leq \int_0^{w_1} \partial_1\Lambda_2(t, w_2)\ \mathrm{d}t = \Lambda_2(w_1, w_2)\ . \qquad \square$$

Proof of Proposition 4.2.4. The same argument as the one applied in the proof of Proposition 4.4.11 shows

$$\lim_{y\to\infty} \Lambda_2(x, y) = \widetilde{\Lambda_2}'(0)\cdot x \implies T_{\Lambda_2}\mathbb{1}_{\mathbb{R}_+}(x) = \widetilde{\Lambda_2}'(0)\mathbb{1}_{\mathbb{R}_+}(x)\ .$$

Combined with the continuity property of substochastic operators (see Property 4 in Definition 4.4.1), we have

$$\lim_{y\to\infty} \partial_1\Lambda_2(x, y) = \lim_{y\to\infty} T_{\Lambda_2}\mathbb{1}_{[0,y]}(x) = T_{\Lambda_2}\mathbb{1}_{\mathbb{R}_+}(x) = \widetilde{\Lambda_2}'(0)\mathbb{1}_{\mathbb{R}_+}(x)\ .$$

An application of the dominated convergence theorem then yields

$$\begin{aligned}(\widetilde{\Lambda_1 * \Lambda_2})'(0) &= \lim_{y\to\infty} (\Lambda_1 * \Lambda_2)(1, y) \\ &= \lim_{y\to\infty} \int_0^\infty \partial_2\Lambda_1(1, t)\partial_1\Lambda_2(t, y)\ \mathrm{d}t \\ &= \int_0^\infty \partial_2\Lambda_1(1, t)\widetilde{\Lambda_2}'(0)\mathbb{1}_{\mathbb{R}_+}(t)\ \mathrm{d}t \\ &= \widetilde{\Lambda_2}'(0) \lim_{t\to\infty} \Lambda_1(1, t) = \widetilde{\Lambda_2}'(0)\cdot\widetilde{\Lambda_1}'(0)\ .\end{aligned}$$

The second claim can be derived by observing that $\widetilde{\Lambda}'(1) = -\widetilde{\Lambda^\top}'(0)$. Finally, the last assertion stems from the fact that $\widetilde{\Lambda_1 *_C \Lambda_2}$ is concave and thus has a monotone derivative. $\square$

5 Stochastic monotonicity and the Markov product for copulas

The Markov product for tail dependence functions exhibits many analytical and dynamical properties which simply do not hold in the case of 2-copulas. From the convergence properties in Proposition 4.1.7 to the dynamical behaviour of iterates and idempotents in Section 4.3, the analytical properties of tail dependence functions play a fundamental role. But why do tail dependence functions yield such strong results when combined with the Markov product? One key aspect is that the partial derivatives $w_1 \mapsto \partial_1\Lambda(w_1, w_2)$ as well as $w_2 \mapsto \partial_2\Lambda(w_1, w_2)$ are decreasing, thereby fitting into the context of majorization theory presented in Section 2.6. In combination with the equivalence between the pointwise ordering and the majorization ordering $\preceq$ introduced in Definition 2.6.5 and adapted to the domain $\mathbb{R}_+$ (see Bennett and Sharpley (1988)) via

$$\Lambda_1 \leq \Lambda_2 \Leftrightarrow \partial_1\Lambda_1(\,\cdot\,, w_2) \preceq \partial_1\Lambda_2(\,\cdot\,, w_2) \text{ for all } w_2 \in \mathbb{R}_+ \ ,$$

many results of Chapter 4 follow quite naturally, such as $\Lambda_1 * \Lambda_2 \leq \min\{\Lambda_1, \Lambda_2\}$. A similar result cannot hold for copulas in general, consider for instance $(C^- * C^-)(\boldsymbol{u}) = C^+(\boldsymbol{u}) > C^-(\boldsymbol{u})$ for $\boldsymbol{u} \in (0,1)^2$. So, is it worthwhile to investigate these results for concave or convex copulas? Unfortunately, it is well-known that concavity is too strong a concept for copulas (see, e.g., Section 3.4.3 in Nelsen (2006)). In fact, any concave d-copula C fulfils

$$\lambda \geq C(\lambda, \ldots, \lambda) = C(\lambda\mathbf{1} + (1-\lambda)\mathbf{0}) \geq \lambda C(\mathbf{1}) + (1-\lambda)C(\mathbf{0}) = \lambda$$

with $\lambda \in [0,1]$, implying $C = C^+$. Similarly, the only convex copula is C^- in dimension 2, and none exist in dimension $d > 2$. Thus, we first have to find an appropriate notion of multivariate concavity to transfer the above ideas and results. Nelsen (2006) and Alvoni, Durante, Papini and Sempi (2007) present an extensive comparison of such concepts in the setting of copulas, including global concavity, directional concavity (that is, concavity in every component whenever the other components are held fixed), componentwise concavity (that is, concavity in only one specific component when the other components are held fixed) and Schur-concavity.

In light of the techniques applied in Chapter 4, we study the consequences of monotonicity of $u \mapsto \partial_1 C(u, v)$ for all v. Exploiting the connection between the partial derivative and the conditional expectation stated in Proposition 2.2.1, componentwise concavity captures a monotone influence of one random variable on another one. More precisely, we say X_2 is *stochastically increasing* in X_1 whenever the corresponding conditional distribution functions are pointwise decreasing, i.e.

$$x_1 \mapsto \mathbb{P}(X_2 \leq x_2 \mid X_1 = x_1)$$

is decreasing. Similarly, X_2 is called *stochastically decreasing* in X_1 if $\mathbb{P}(X_2 \leq x_2 \mid X_1 = x_1)$ is increasing in x_1. We will discuss stochastically increasing as well as stochastically decreasing pairs of random variables under the term stochastically monotone.

In applications, stochastic monotonicity is extensively used to account for directed relations between random variables as it enables, e.g., the study of the long-term behaviour of economic models such as stochastic recursions of the form $X_{n+1} = f(X_n, Z_n)$ (see Stokey, Lucas and Prescott (1989) or Foss, Shneer, Thomas and Worrall (2018) for details). Illustrative data examples which exhibit stochastic monotonicity are given by the connection between expenditures and income of a household or the income mobility from one generation to the next and can be found in Lee, Linton and Whang (2009). Furthermore, many parametric distribution families such as the multivariate Gaussian distribution, extreme-value copulas and certain Archimedean copulas are stochastically monotone, thereby providing a common framework to treat features of these seemingly disparate models (see Joe (2015)).

Aside from a stochastic introduction given in Nelsen (2006), the notion of stochastic monotonicity has mainly been considered from a geometric— and, crucially, symmetric— point of view under the term 'directional concavity'. This more restrictive concept of directional concavity requires that a copula C is stochastically monotone in every component (see, e.g., Alvoni et al. (2007), Durante and Papini (2009) and Dolati and Nezhad (2014) for an extensive treatment). One obvious drawback of this symmetric approach is the loss of a directed influence between random variables, as it suggests a circular interaction: An increase in X_1 leads to an increase in X_2, which in turn leads to higher values in X_1, and so on. We therefore focus on a directed approach, which still possesses many of the advantageous analytical properties seen in Chapter 4.

Building upon the isomorphism between 2-copulas and Markov operators given in Theorem 2.2.3, we show that stochastically monotone copulas are in one-to-one correspondence with monotonicity-preserving Markov operators. As a by-product, this implies that the set of stochastically monotone copulas is closed under the application of the Markov product, i.e. $C_1 * C_2$ is again stochastically monotone if both factors C_1 and C_2 are. We also show that for stochastically monotone copulas, many notions of convergences, such as uniform convergence and convergence of the partial derivatives, coincide. Similar results have recently been established in Kasper, Fuchs and Trutschnig (2021) for Archimedean and extreme-value copulas.

Furthermore, we characterize stochastically increasing copulas by their reduction property under the Markov product with respect to the stochastic dominance ordering. More precisely, we prove that a copula C is stochastically increasing if and only if

$$(D * C)(\boldsymbol{u}) \leq C(\boldsymbol{u})$$

holds for all $\boldsymbol{u} \in [0,1]^2$ and all 2-copulas D. Similar to our findings for tail dependence functions (see Section 4.3), this reduction property enables us to characterize idempotents and the asymptotic behaviour of limits of stochastically monotone copulas as ordinal sums of the independence copula. The greater variety of possible behaviours of the idempotents and the limits as compared to the case of tail dependence functions is due to the fact that copulas are, in general, not positive homogeneous.

This chapter is based on Siburg and Strothmann (2021b) and is structured as follows: Section 5.1 establishes a connection between stochastically monotone copulas and monotonicity-preserving Markov operators, as well as some topological closure properties of stochastically monotone copulas. Section 5.2 contains the aforementioned characterization of stochastic monotonicity in terms of a reduction property concerning the Markov product. Section 5.3 identifies idempotent, stochastically monotone copulas as ordinal sums of Π.

5.1 Stochastic monotonicity for copulas and Markov operators

Definition 5.1.1. *A 2-copula C is called stochastically increasing (decreasing) in the i-th component if $u_i \mapsto \partial_i C(u_1, u_2)$ is decreasing (increasing) for almost all $u_i \in [0,1]$.*

Whenever the meaning is clear, we will drop the specification 'in the i-th component'. If the distinction between $u_i \mapsto \partial_i C(u_1, u_2)$ being increasing or decreasing is of no concern, we simply call C stochastically monotone. Furthermore, we will state many results only with respect to the first component, from which the results in the other component follow by transposition. We denote the set of all in the first component stochastically increasing (decreasing) copulas by $\mathcal{C}^{\uparrow}$ ($\mathcal{C}^{\downarrow}$). Note that stochastically *increasing* 2-copulas have a *decreasing* partial derivative. This is due to the fact that for real-valued random variables X and Y, X is smaller than Y in the usual stochastic dominance order, $X \leq_{st} Y$ in short, if

$$F_X(t) \geq F_Y(t)$$

holds for all $t \in \mathbb{R}$.

From now on, we drop the vector notation $\boldsymbol{u} = (u_1, u_2)$ as we will be exclusively considering bivariate copulas, unless explicitly stated otherwise. Instead, we use (u, v) for a generic element of $[0,1]^2$. We begin by presenting some well-known examples of stochastically monotone 2-copulas.

Example 5.1.2. *Suppose C is an Archimedean 2-copula (see Definition 2.5.5),*

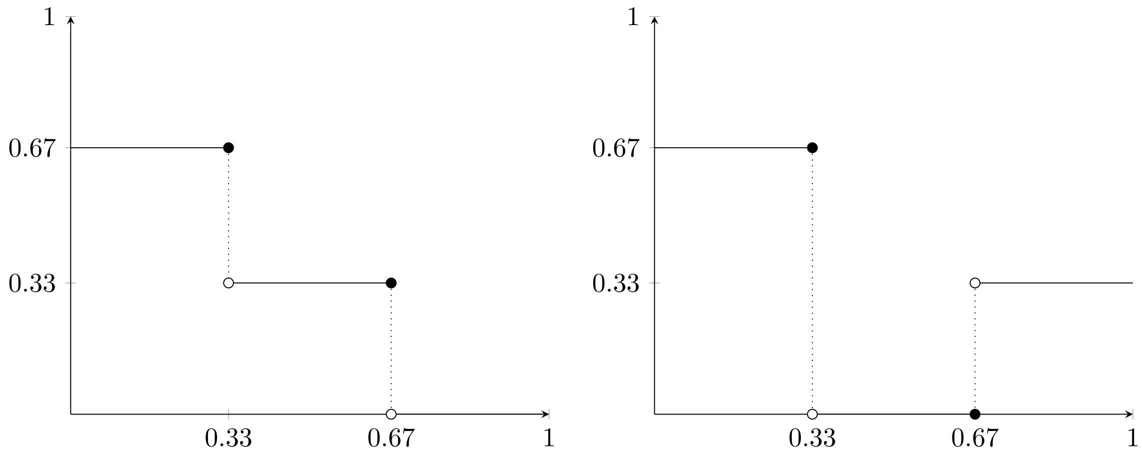
$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \ ,$$

with Archimedean generator ϕ. If the generalized inverse $\phi^{[-1]}$ is twice-differentiable, C is stochastically increasing in both components if and only if $\log\left(-\phi^{[-1]'}\right)$ is convex (see Proposition 3.3 in Capéraà and Genest (1993)). In particular, the independence copula Π with generator $\phi(t) = -\log(t)$ is stochastically increasing.

Example 5.1.3. *The class of extreme-value 2-copulas*

$$C^{EV}(u, v\,; \Lambda) := \exp\left(\log(uv)\left(1 - \widetilde{\Lambda}\left(\frac{\log(u)}{\log(uv)}\right)\right)\right) \tag{5.1}$$

given in Proposition 2.5.1, where $\widetilde{\Lambda} : [0,1] \to [0,1/2]$ is a concave function fulfilling $0 \leq \widetilde{\Lambda}(t) \leq \min\{t, 1-t\}$ for all $t \in [0,1]$, is stochastically increasing in both components (see Theorem 1 in Garralda Guillem (2000)). This class includes both Π and C^+ as examples with $\widetilde{\Lambda}(t) = 0$ and $\widetilde{\Lambda}(t) = \min(t, 1-t)$, respectively.



(a) Plot of $\partial_1 C_3^{\#}(A)(\,\cdot\,,1/3)$ (b) Plot of $\partial_2 C_3^{\#}(A)(1/3,\,\cdot\,)$

Figure 5.1: Plots depicting the partial derivatives of the checkerboard copula $C_3^{\#}(A)$ with respect to the first and second component. The doubly stochastic matrix A is given in Example 5.1.4.

All of these examples are stochastically increasing in both components, and thus do not allow for an only unidirectional positive influence between the random variables. Additionally, many common construction methods using 2-copulas as building blocks also preserve the stochastic increasing property in both components, such as convex combinations and ordinal sums of stochastically increasing 2-copulas (see Durante and Papini (2009)). A 2-copula which is stochastically increasing in the first but not in the second component is given by the following example.

Example 5.1.4. *A straightforward calculation shows that the checkerboard-copula*

$$C_3^{\#}(A)(u,v) = 3\sum_{k,\ell=1}^{3} a_{k\ell} \int_0^u \mathbb{1}_{[\frac{k-1}{n},\frac{k}{n})}(s)\ \mathrm{d}s \int_0^v \mathbb{1}_{[\frac{\ell-1}{n},\frac{\ell}{n})}(t)\ \mathrm{d}t$$

with the doubly stochastic matrix $A = (a_{k\ell})_{k,\ell=1,2,3} \in \mathbb{R}^{3\times 3}$,

$$A = \begin{pmatrix} 2/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix} ,$$

is stochastically increasing in the first but not the second component. A plot of the partial derivatives of $C_3^{\#}(A)$ *is depicted in Figure 5.1. A necessary and sufficient condition for a checkerboard copula to be stochastically increasing can be found in Section 6.2.*

To simplify subsequent proofs, we will first present a direct connection between stochastically increasing and stochastically decreasing copulas, allowing us to transfer results obtained for stochastically increasing copulas to stochastically decreasing ones and vice versa.

Lemma 5.1.5. *The mapping* $C \mapsto (C^- * C)$ *is an involution between* $\mathcal{C}^{\uparrow}$ *and* $\mathcal{C}^{\downarrow}$.

Proof. The claim follows immediately from $\partial_1(C^- * C)(u,v) = \partial_1 C(1-u,v)$ for almost all u and all $v \in [0,1]$ and any 2-copula C. □

The class of stochastically monotone copulas also provides additional structure to strengthen convergence properties since the monotonicity yields the equivalence of uniform convergence and pointwise convergence of the partial derivative.

Proposition 5.1.6. *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of stochastically monotone 2-copulas (in the first component). Then the following are equivalent:*

1. *C_n converges uniformly towards C.*

2. *$\partial_1 C_n(u,v)$ converges pointwise towards $\partial_1 C(u,v)$ for all v and almost all $u \in [0,1]$.*

Proof. Using the dominated convergence theorem, 2 implies 1. Conversely, suppose C_n converges uniformly towards C. Due to $C_n(\,\cdot\,,v)$ and $C(\,\cdot\,,v)$ being concave (convex) for all $v \in [0,1]$, Lemma 1 in Tsuji (1952) implies

$$\lim_{n\to\infty} \partial_1 C_n(u,v) = \partial_1 C(u,v)$$

for almost all $u \in [0,1]$ and all $v \in [0,1]$. This yields the implication 1 to 2. □

Remark 5.1.7. *Together with an application of the dominated convergence theorem, Proposition 5.1.6 yields the equivalence of uniform convergence and D_1-convergence (see Definition 2.3.5) for stochastically monotone copulas. More precisely, for a sequence of stochastically monotone copulas $(C_n)_{n\in\mathbb{N}}$, $\|C_n - C\|_\infty \to 0$ if and only if $D_1(C_n, C) \to 0$. Moreover, $\mathcal{C}^\uparrow$ and $\mathcal{C}^\downarrow$ are closed with respect to d_∞ and D_1.*

Remark 5.1.8. *The equivalence of 1 and 2 in Proposition 5.1.6 also holds for completely dependent copulas and their corresponding Markov operators, the so-called Markov embeddings, a proof of which can be found in Theorem 13.11 in Eisner, Farkas, Haase and Nagel (2015).*

We now give the main result of this section and characterize the behaviour of stochastically monotone 2-copulas and their corresponding Markov operators. We say $f \in L^1([0,1])$ is monotone if there exists a monotone function $g : [0,1] \to \mathbb{R}$ such that $f = g$ holds almost everywhere.

Theorem 5.1.9. *Suppose X and Y are continuous random variables with copula C_{XY}. Then the following are equivalent*

1. *Y is stochastically increasing (decreasing) in X.*

2. *C_{XY} is stochastically increasing (decreasing) in the first component.*

3. *$C_{XY}(u,v)$ is concave (convex) in u for all $v \in [0,1]$.*

4. *$T_{C_{XY}}$ is a monotonicity-preserving (monotonicity-reversing) Markov operator, i.e. it maps decreasing integrable functions onto decreasing (increasing) functions.*

5. *$\mathbb{E}(f(Y) \mid X = x)$ is decreasing (increasing) for every decreasing function f for which the expectation $\mathbb{E}(f(Y))$ exists.*

Proof of Theorem 5.1.9. We give the proof for stochastically increasing random variables; the case of stochastically decreasing random variables is analogous. The equivalence of the first three assertions is shown in Nelsen (2006). Suppose 4 holds, then $f = \mathbb{1}_{[0,v]}$ yields 2. For the implication 2 to 4, note that $T_{C_{XY}}$ maps decreasing indicator functions onto decreasing functions due to

$$T_{C_{XY}} \mathbb{1}_{[0,v]}(\,\cdot\,) = \partial_1 C_{XY}(\,\cdot\,, v)$$

being decreasing for all $v \in [0,1]$. Using the approximation of monotone functions via monotone indicator functions and applying the monotone convergence theorem, Assertion 4 follows. Similarly, 1 and 5 are equivalent due to

$$\mathbb{E}(\mathbb{1}_{[0,y]}(Y) \mid X = x) = \mathbb{P}(Y \leq y \mid X = x) \; . \qquad \square$$

5.2 The Markov product of stochastically monotone copulas

Theorem 5.1.9 guarantees that the composition of two monotonicity-preserving Markov operators is again monotonicity-preserving. Using the isomorphism between $(\mathcal{C}_2, *)$ and Markov operators equipped with the composition, we establish the following closure property of $\mathcal{C}^{\uparrow}$ and $\mathcal{C}^{\downarrow}$ with respect to the Markov product.

Corollary 5.2.1. *Suppose C_1 and $C_2 \in \mathcal{C}_2$ are stochastically monotone in the first component. Then $C_1 * C_2$ is again stochastically monotone in the first component. More precisely:*

1. *$C_1 * C_2$ is stochastically increasing if both C_1 and C_2 are either stochastically increasing or stochastically decreasing.*
2. *$C_1 * C_2$ is stochastically decreasing if one of C_1 and C_2 is stochastically increasing and the other one is stochastically decreasing.*

Proof. If C_1 and C_2 are stochastically monotone in the first component, using Theorem 5.1.9, T_{C_1} and T_{C_2} map monotone functions onto monotone functions. Their composition therefore also maps monotone functions onto monotone functions. Assertions 1 and 2 follow immediately from a case-by-case analysis using Property 4 of Theorem 5.1.9. $\square$

The above closure property is only one aspect of the interplay between stochastically monotone 2-copulas and the Markov product. The next result shows that stochastically increasing 2-copulas maximize the Markov product. Note that we observed a similar property for tail dependence functions in Theorem 4.2.1.

Theorem 5.2.2. *Let C be a 2-copula. C is stochastically increasing in the first component if and only if*

$$(D * C)(u,v) \leq C(u,v) \; .$$

holds for all 2-copulas D and all $u, v \in [0,1]$. On the other hand, C is stochastically decreasing in the first component if and only if

$$C(u,v) \leq (D * C)(u,v)$$

holds for all 2-copulas D and all $u, v \in [0,1]$.

Theorem 5.2.2 also yields that the Markov operator T_C is monotonicity-preserving if and only if

$$\int_0^u (T_D \circ T_C)f(t)\ \mathrm{d}t \leq \int_0^u T_C f(t)\ \mathrm{d}t$$

holds for all decreasing functions $f \in L^1([0,1])$, all $u \in [0,1]$ and all Markov operators T_D.

Proof. We will only show the first equivalence, the second one then follows from Lemma 5.1.5. Since

$$\int_0^v \partial_2 D(u,t)\ \mathrm{d}t = D(u,v) \leq C^+(u,v) = \int_0^v \mathbb{1}_{[0,u]}(t)\ \mathrm{d}t$$

holds for all $u, v \in [0,1]$, an application of Hardy's Lemma (see Proposition 2.3.6 in Bennett and Sharpley (1988)) yields

$$(D * C)(u,v) = \int_0^1 \partial_2 D(u,t)\partial_1 C(t,v)\ \mathrm{d}t \leq \int_0^1 \mathbb{1}_{[0,u]}(t)\partial_1 C(t,v)\ \mathrm{d}t = C(u,v)\ .$$

Now, we turn to the converse implication and assume $D * C \leq C$ holds for all 2-copulas D. Let $v \in (0,1)$ be arbitrary and set $f(u) := \partial_1 C(u,v)$. Using Proposition 2.6.3, there exists a measure-preserving transformation $\sigma : [0,1] \to [0,1]$ and a decreasing function $g : [0,1] \to [0,1]$ such that

$$\partial_1 C(u,v) = f(u) = g(\sigma(u)) = T_\sigma g(u)\ ,$$

where T_σ is a left-invertible Markov operator (commonly known as a Koopman operator). Using Theorem 2.2.3, T_σ corresponds to a left-invertible 2-copula C_σ. An application of the adjoint T'_σ together with the left-invertibility of T_σ (see Remark 2.6.4) yields

$$g(u) = T'_\sigma \partial_1 C(\,\cdot\,,v)(u) = \partial_1 (C_\sigma^\top * C)(u,v)\ .$$

Setting $D := C_\sigma^\top * C$, we have $\partial_1 D(u,v) = g(u)$ almost everywhere. Therefore $u \mapsto \partial_1 D(u,v)$ is decreasing and fulfils by assumption

$$D(u,v) = (C_\sigma^\top * C)(u,v) \leq C(u,v)\ .$$

On the other hand, the Hardy-Littlewood inequality (see Proposition 2.6.7) yields

$$C(u,v) = \int_0^u \partial_1 C(t,v)\ \mathrm{d}t \leq \int_0^u \partial_1 D(t,v)\ \mathrm{d}t = D(u,v) \leq C(u,v)\ .$$

Thus, $C(\,\cdot\,,v) = D(\,\cdot\,,v)$ and $u \mapsto \partial_1 C(u,v)$ must be decreasing. □

Remark 5.2.3. *Theorem 5.2.2 yields an alternative approach to derive the positive quadrant dependence of stochastically increasing copulas, that is, the fact that*

$$\Pi(u,v) = (\Pi * C)(u,v) \leq C(u,v)$$

holds for any stochastically increasing 2*-copula* C*.*

Remark 5.2.4. *Analogously to the previous remark, any stochastically decreasing* 2*-copula* C *is negative quadrant dependent due to*

$$C(u,v) \le (\Pi * C)(u,v) = \Pi(u,v) \ .$$

Theorem 5.2.2 also characterizes stochastically increasing, completely dependent copulas.

Remark 5.2.5. *A* 2*-copula* C *is called completely dependent if it is left-invertible, i.e. fulfils* $C^\top * C = C^+$*. Due to Theorem 5.2.2, any completely dependent and stochastically increasing copula* C *fulfils*

$$C^+ = C^\top * C \le C \le C^+ \ ,$$

so that $C = C^+$ *holds.*

5.3 Idempotents of stochastically monotone copulas

The rest of this chapter aims to characterize idempotent, stochastically monotone 2-copulas and monotonicity-preserving conditional expectations. While the idempotency of C appears to be a purely algebraic property, it translates to the fundamental stochastic property of T_C being a conditional expectation on $L^1([0,1])$ (see Proposition 2.2.7).

It is well-known that any idempotent 2-copula C is necessarily symmetric (see, for example, Darsow and Olsen (2010) or Trutschnig (2013)). Therefore, whenever it is stochastically monotone in one component, it is stochastically monotone in the same sense in the other component. Thus, it suffices to require C to be stochastically monotone in either component and we will simply call C stochastically monotone. Before stating the main result of this section, let us define the ordinal sum of copulas (see Durante and Sempi (2016)).

Definition 5.3.1. *Let* $((a_k,b_k))_{k\in\mathcal{I}}$ *be a countable family of disjoint intervals in* $(0,1)$ *and* $(C_k)_{k\in\mathcal{I}}$ *a family of* 2*-copulas. A* 2*-copula* C *is called an ordinal sum of* $(C_k)_{k\in\mathcal{I}}$ *with respect to* $((a_k,b_k))_{k\in\mathcal{I}}$ *if*

$$C(u,v) = \begin{cases} a_k + (b_k - a_k) C_k\left(\frac{u-a_k}{b_k-a_k}, \frac{v-a_k}{b_k-a_k}\right) & \text{for } (u,v) \in (a_k,b_k)^2 \text{ for some } k \in \mathcal{I} \\ C^+(u,v) & \text{else} \end{cases} .$$

We adopt the short-hand notation $C = (\langle (a_k,b_k)\,,\, C_k\rangle)_{k\in\mathcal{I}}$ *from Durante and Sempi (2016).*

Theorem 5.3.2. *Suppose* C *is a* 2*-copula.* C *is stochastically monotone and idempotent if and only if it is an ordinal sum of* Π.

We split the proof of Theorem 5.3.2 into two parts. We will begin with the result concerning stochastically decreasing copulas.

Proposition 5.3.3. *The product copula* $\Pi(u,v) = uv$ *is the only idempotent* 2*-copula which is stochastically decreasing.*

Proof. Let C be a stochastically decreasing idempotent copula. Corollary 5.2.1 together with C being idempotent yields that $C = C*C$ is stochastically increasing. Thus, $\partial_1 C(u,v) = c_v \in$

$[0,1]$ must hold for almost all $u \in [0,1]$, which, combined with the uniform margin property of copulas, leads to

$$v = \int_0^1 \partial_1 C(u,v)\, \mathrm{d}u = \int_0^1 c_v\, \mathrm{d}u = c_v \ .$$

Integrating then gives the assertion $C(u,v) = uv = \Pi(u,v)$. □

Proposition 5.3.3 states that the only idempotent copula in the class of stochastically decreasing copulas is the product copula. It is natural to ask whether the same holds true inside the larger class of negative quadrant dependent copulas (see Remark 5.2.4). Indeed, this is the case, as the following proposition shows.

Proposition 5.3.4. *The product copula $\Pi(u,v) = uv$ is the only idempotent 2-copula which is negative quadrant dependent.*

Proof. Since every idempotent copula C is symmetric (see Theorem 6.1 in Darsow and Olsen (2010)), Lemma 5.1 in Fernández Sánchez, Trutschnig and Tschimpke (2021) yields

$$(C * C)(u,u) = (C * C^\top)(u,u) \geq \Pi(u,u)$$

for all $u \in [0,1]$. Combining this with the negative quadrant dependence of C, we obtain

$$\Pi(u,u) \geq C(u,u) = (C * C)(u,u) \geq \Pi(u,u) \ .$$

Therefore, C has diagonal section $C(u,u) = \Pi(u,u) = u^2$ and Theorem 8.7 in Darsow et al. (1992) asserts that $C = \Pi$. □

Since Proposition 5.3.3 already characterizes all idempotent, stochastically decreasing copulas, it remains to analyse the behaviour of stochastically increasing copulas. The next lemma provides a crucial technical property for the proof of Theorem 5.3.2 by relating the partial derivative of a stochastically increasing copula C to the rate with which C changes from $C(v,v)$ to v.

Lemma 5.3.5. *Let C be an idempotent, stochastically increasing 2-copula. Then*

$$(v - C(v,v))\partial_2^- C(u,v) = C(u,v) - C(u, C(v,v))$$

holds for all $u, v \in (0,1)$, where $\partial_2^- C$ denotes the left-hand derivative of C with respect to the second component.

Proof. As C is stochastically increasing, the left-hand partial derivative $\partial_2^- C(u,t)$ exists everywhere on $(0,1)$ and is decreasing. Furthermore, due to $C(t,v) \le C^+(t,v) \le t$, we have

$$\begin{aligned}(C*C)(u,v) &= \int_0^1 \partial_2 C(u,t)\partial_1 C(t,v)\ \mathrm{d}t = \int_0^1 \partial_2^- C(u,t)\partial_1 C(t,v)\ \mathrm{d}t \\ &\le \int_0^1 \partial_2^- C(u,C^+(t,v))\partial_1 C(t,v)\ \mathrm{d}t \\ &\le \int_0^v \partial_2^- C(u,C(t,v))\partial_1 C(t,v)\ \mathrm{d}t + \int_v^1 \partial_2^- C(u,C^+(t,v))\partial_1 C(t,v)\ \mathrm{d}t \\ &\le \int_0^1 \partial_2^- C(u,C(t,v))\partial_1 C(t,v)\ \mathrm{d}t = \int_{C(0,v)=0}^{C(1,v)=v} \partial_2^- C(u,z)\ \mathrm{d}z = C(u,v)\ .\end{aligned}$$

The change of variables is possible due to the Riemann-integrability of $t \mapsto \partial_1 C(t,v)$ and $t \mapsto \partial_2 C(u,t)$. Now, as $(C*C)(u,v) = C(u,v)$ holds, all inequalities are in fact equalities. This yields

$$\begin{aligned}C(u,v) &= \int_0^v \partial_2^- C(u,C(t,v))\partial_1 C(t,v)\ \mathrm{d}t + \int_v^1 \partial_2^- C(u,C^+(t,v))\partial_1 C(t,v)\ \mathrm{d}t \\ &= \int_0^v \partial_2^- C(u,C(t,v))\partial_1 C(t,v)\ \mathrm{d}t + \int_v^1 \partial_2^- C(u,v)\partial_1 C(t,v)\ \mathrm{d}t \\ &= \int_{C(0,v)=0}^{C(v,v)} \partial_2^- C(u,z)\ \mathrm{d}z + \partial_2^- C(u,v)\,(v - C(v,v)) \\ &= C(u,C(v,v)) + \partial_2^- C(u,v)\,(v - C(v,v))\ .\end{aligned}$$

□

The above property of the partial derivative is the main ingredient to prove the desired characterization of idempotent, stochastically increasing copulas. The only difficulty remains in the term $(v - C(v,v)) \ge 0$. Whenever the latter is strictly positive, the following lemma characterizes the corresponding copula completely. If $(v - C(v,v)) = 0$ for some $v \in (0,1)$, we will need to consider the behaviour of C more closely in Theorem 5.3.7.

Lemma 5.3.6. *Suppose C is a 2-copula with $C(v,v) < v$ for all $v \in (0,1)$. Then C is stochastically monotone and idempotent if and only if $C(u,v) = uv = \Pi(u,v)$.*

Proof. The assertion for stochastically decreasing copulas follows from Proposition 5.3.3. Thus, applying Lemma 5.3.5 in combination with C being stochastically increasing, we obtain for arbitrary $u, v \in (0,1)$

$$\partial_2^- C(u,v) = \frac{C(u,v) - C(u, C(v,v))}{v - C(v,v)} = \frac{1}{v - C(v,v)} \int_{C(v,v)}^{v} \partial_2^- C(u,t) \ \mathrm{d}t$$

$$\geq \frac{1}{v - C(v,v)} \int_{C(v,v)}^{v} \partial_2^- C(u,v) \ \mathrm{d}t = \partial_2^- C(u,v) \ .$$

Therefore, the inequality above is actually an equality, and the partial derivative, not only the left-hand derivative, fulfils $\partial_2 C(u,t) = c_u \in [0,1]$ almost everywhere on the interval $(C(v,v), v)$. Since $C(v,v) < v$ holds for all $v \in (0,1)$, $\{(C(v,v), v)\}_{v \in (0,1)}$ is a (nondisjoint) covering of $(0,1)$ with intervals having nonempty interior. Consequently, we must have $\partial_2 C(u,t) = c_u$ for almost all $t \in (0,1)$. Hence,

$$u = C(u,1) = \int_0^1 \partial_2 C(u,t) \ \mathrm{d}t = \int_0^1 c_u \ \mathrm{d}t = c_u \ ,$$

which implies $C(u,v) = uv = \Pi(u,v)$. □

With the previous lemma, we are now able to characterize all idempotent, stochastically increasing 2-copulas. Theorem 5.3.2 then follows from a combination of Proposition 5.3.3 and Theorem 5.3.7 below.

Theorem 5.3.7. *Ordinal sums of the independence copula are the only idempotent 2-copulas which are stochastically increasing.*

Proof. The proof treats three distinct cases, depending on the behaviour along the diagonal. Suppose C is stochastically increasing and idempotent. If $C(v,v) < v$ for all $v \in (0,1)$, then $C = \Pi = \langle (0,1), \Pi \rangle$ due to Lemma 5.3.6. If on the other hand $C(v,v) = v$ holds for all $v \in (0,1)$, then $C = C^+ = \langle (a_k, b_k), \Pi \rangle_{k \in \emptyset}$. Lastly, if $C(v,v) = v$ holds for some $v \in (0,1)$ and $C \neq C^+$, Corollaries 3.2 and 3.3 from Mesiar and Sempi (2010) yield that C is the ordinal sum of ordinally irreducible 2-copulas

$$C = (\langle (a_k, b_k), C_k \rangle)_{k \in \mathcal{I}} \ .$$

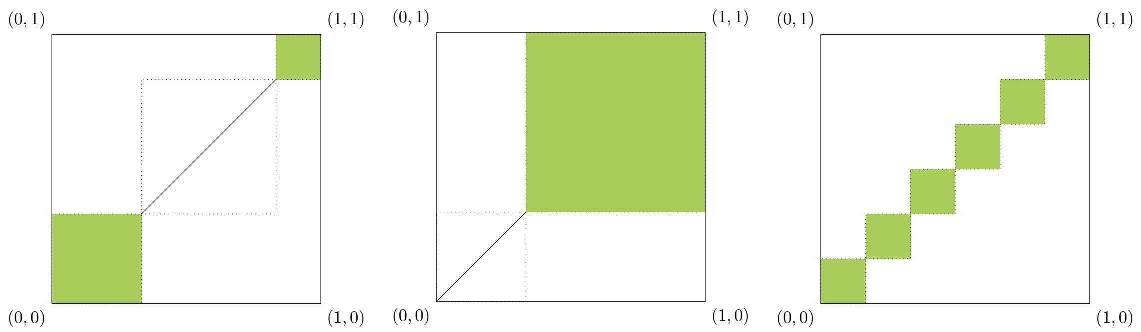
Due to Theorem 3.2.1 in Nelsen (2006), ordinally irreducible copulas C_k fulfil $C_k(v,v) < v$ for all $v \in (0,1)$. Theorem 3.1 from Albanese and Sempi (2016) then states that C is idempotent if and only if every C_k is idempotent. Moreover, the ordinal sum C is stochastically increasing if and only if every C_k is stochastically increasing. Thus, every C_k is idempotent, stochastically increasing and fulfils $C_k(v,v) < v$ on $(0,1)$. Lemma 5.3.6 then implies $C_k = \Pi$, which yields

$$C = (\langle (a_k, b_k), \Pi \rangle)_{k \in \mathcal{I}} \ .$$ □

Example 5.3.8. *Apart from the extreme cases Π and C^+, ordinal sums of Π can take various forms. Three different configurations, namely*

$$\left(\left\langle \left(0, \frac{1}{3}\right), \Pi \right\rangle, \left\langle \left(\frac{5}{6}, 1\right), \Pi \right\rangle \right), \left\langle \left(\frac{1}{3}, 1\right), \Pi \right\rangle \ \text{and} \ \left(\left\langle \left(\frac{k}{6}, \frac{k+1}{6}\right), \Pi \right\rangle \right)_{k \in \{0,\dots,5\}} ,$$

are depicted in Figure 5.2.



(a) $\left(\left\langle(0,\frac{1}{3}),\Pi\right\rangle,\left\langle(\frac{5}{6},1),\Pi\right\rangle\right)$. (b) $\left\langle(\frac{1}{3},1),\Pi\right\rangle$. (c) $\left(\left\langle(\frac{k}{6},\frac{k+1}{6}),\Pi\right\rangle\right)_{k\in\{0,\ldots,5\}}$.

Figure 5.2: Idempotent, stochastically monotone ordinal sums of Π with respect to different families of disjoint intervals.

Proposition 5.3.9. *Suppose C is stochastically increasing in the first component. Then there exists a countable family of intervals $((a_k,b_k))_{k\in\mathcal{I}}$ such that*

$$C^{*n}(u,v)\to(\langle(a_k,b_k),\Pi\rangle)_{k\in\mathcal{I}}(u,v)$$

holds pointwise.

A similar behaviour was established for Cesàro averages of iterates of quasi-constrictive Markov operators in Trutschnig and Fernández Sánchez (2015).

Remark 5.3.10. *Since the 2-copulas C^{*n} are stochastically increasing in the first component (see Corollary 5.2.1), an application of Proposition 5.1.6 yields that the pointwise convergence*

$$C^{*n}(u,v)\to(\langle(a_k,b_k),\Pi\rangle)_{k\in\mathcal{I}}(u,v)$$

is equivalent to the pointwise convergence of the partial derivatives.

Proof of Proposition 5.3.9. Due to Corollary 5.2.1, C^{*n} is stochastically increasing for all $n\in\mathbb{N}$. By Theorem 5.2.2,

$$0\leq C^{*n}(u,v)=(C*C^{*(n-1)})(u,v)\leq C^{*(n-1)}(u,v)$$

follows for all $u,v\in[0,1]$. Thus, C^{*n} is a decreasing sequence of copulas and as such, converges pointwise against some $C^*\in\mathcal{C}_2$. The pointwise limit of the concave functions $u\mapsto C^{*n}(u,v)$ is again concave, therefore C^* is stochastically increasing in the first component. Furthermore, due to the Markov product being continuous with respect to the pointwise convergence in one component, we have that

$$C*C^*=\lim_{n\to\infty}C*C^{*n}=C^*$$

holds. An inductive argument now yields

$$C^**C^*=\lim_{n\to\infty}C^{*n}*C^*=\lim_{n\to\infty}C^*=C^*\ .$$

Thus C^* is idempotent. An application of Theorem 5.3.7 then guarantees the existence of a countable family of intervals $((a_k,b_k))_{k\in\mathcal{I}}$ such that $C^*=(\langle(a_k,b_k),\Pi\rangle)_{k\in\mathcal{I}}$. ☐

Finally, we translate the results of this section into the language of Markov operators and conditional expectations. Following Proposition 2.2.7, a 2-copula C is idempotent if and only if T_C is a conditional expectation restricted to $L^1([0,1],\mathcal{B}([0,1]),\lambda)$. Thus, it follows immediately that a conditional expectation is monotonicity-preserving if and only if it is pointwise either an averaging operation or the identity.

Theorem 5.3.11. *Suppose T is a conditional expectation on $L^1([0,1],\mathcal{B}([0,1]),\lambda)$. Then T preserves or reverses monotonicity if and only if there exists a countable family of disjoint intervals $((a_k,b_k))_{k\in\mathcal{I}}$ in $(0,1)$ with $P := \cup_{k\in\mathcal{I}}(a_k,b_k)$ such that*

$$T_C f(u) = \sum_{k\in\mathcal{I}} \mathbb{1}_{(a_k,b_k)}(u)\frac{1}{b_k-a_k}\int_{a_k}^{b_k} f(t)\ \mathrm{d}t + \mathbb{1}_{P^C}(u)f(u)$$

for any $f \in L^1([0,1])$.

6 Rearranging copulas and dependence measures

When investigating the relationship between several components of a complex system, e.g., multiple financial returns, the water levels at different locations or the concentration of various air pollutants, we often have to reduce the entire information into easier to grasp statistics. In Chapters 3 and 4, we have approached dependence from an extremal point of view, where the tail dependence function was an indicator for worst-case scenarios. In contrast to this local notion of dependence, we will now focus on a global relationship between two random variables X and Y.

While independence as one extreme case of global dependence is clearly defined, its converse is not. Asymmetric dependence can be described by $Y = f \circ X$, where f is either affine linear, or strictly monotone, or continuous, or, in the most general case, measurable. The abstract question about the existence of a function f with certain smoothness properties has led to the development of various measures of dependence. Fundamentally, a measure of dependence takes values between 0 and 1 and can exactly identify one type of dependence structure, or in our context, one class of functions. For example, Pearson's correlation coefficient (see Nelsen (2006)), defined by

$$\operatorname{Corr}(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)} \cdot \sqrt{\operatorname{Var}(Y)}}$$

whenever X and Y have finite second moments, is a measure of the linear relationship between X and Y. Most importantly, $|\operatorname{Corr}(X, Y)|$ equals 1 if and only if $Y = aX + b$ holds for some $a, b \in \mathbb{R}$ with $a \neq 0$. In short, Corr detects whether $Y = f(X)$ holds for some affine function f. Similarly, the Schweizer-Wolff measures (see Schweizer and Wolff (1981))

$$\sigma_p(X, Y) := \sigma_p(C_{XY}) := \frac{\|C_{XY} - \Pi\|_p}{\|C^+ - \Pi\|_p}$$

for $p \geq 1$ attain their maximal value 1 if and only if $Y = f(X)$ holds for some strictly monotone function f. Thus, Schweizer-Wolff measures are measures of monotone relationships.

Other successfully applied concepts of— for the lack of a better term— 'global dependence' also include symmetric dependencies such as the mutual information discussed in Cover and Thomas (2006), concordance as outlined in Section 2.3 or, more recently, pure dependence introduced by Geenens and Lafaye de Micheaux (2020). But we view asymmetry as a crucial component of the dependence between random variables, a standpoint underlined by empirical evidence found in finance (e.g., the connection between commodities and exchange rates in Okimoto (2008)), hydrology (e.g., flood events considered in Bücher et al. (2017)), and even nutritional data (see Genest, Nešlehová and Quessy (2012)). A similar case for asymmetry in dependence modelling has recently been made by Chatterjee (2021) and by Junker et al. (2021).

In this chapter, we introduce a new class of measures of complete dependence, where (X, Y) is called completely dependent if $Y = f(X)$ holds almost surely for some measurable function f, while building upon well-known measures of dependence. We start by briefly recalling the key features of a 'measure of complete dependence' from the list of axioms given in Definition 2.3.1. A mapping $\mu : (X, Y) \mapsto \mu(X, Y) \in [0, 1]$ is a measure of complete dependence if the following three axioms are met:

(C1) $\mu(X, Y) = 0$ if and only if X and Y are independent.

(C2) $\mu(X, Y) = 1$ if and only if $Y = f(X)$ holds for some measurable f.

Ideally, certain types of data processing should leave the value of μ invariant. For example, a reversible affine transformation of X and Y results in the same correlation, i.e. $\mathrm{Corr}(aX + b, cY + b) = \mathrm{Corr}(X, Y)$ for $a, c \neq 0$, and strictly monotone transformations leave the Schweizer-Wolff measures σ_p invariant. Considering that our measure of complete dependence is asymmetric, we aim for the following invariance property:

(C3) $\mu(f(X), g(Y)) = \mu(X, Y)$, where g is strictly monotone and f is a measurable bijection.

Property (C3) yields that μ only depends on the copula of X and Y, i.e. $\mu(X, Y) = \mu(C_{XY})$ and, in particular, does not depend on the choice of the unit.

Note that combining (C1) with (C2) already has profound implications on the global behaviour of μ. For any random vector (X, Y) with independent components X and Y and continuous univariate marginal distributions, C_{XY} can be approximated pointwise by $C_{X_nY_n}$ for some completely dependent random variables X_n and Y_n (see Example 2.3.4). Consequently, any measure of complete dependence which is continuous with respect to pointwise convergence of copulas fails entirely to detect complete dependence since there is always a copula C corresponding to mutually completely dependent random variables with $\mu(C)$ arbitrarily close to $0 = \mu(\Pi)$. This is the case, for instance, for the Schweizer-Wolff measures and, by design, for all measures of concordance, such as Spearman's ρ or Kendall's τ. However, there exist measures of dependence that do satisfy $\mu(X, Y) = 1$ if and only if Y is a measurable (and not necessarily monotone) function of X. To the best of our knowledge, the first measure with this property was introduced in Siburg and Stoimenov (2010) to capture a symmetric variant of complete dependence. Later, unidirectional measures of complete dependence detecting measurable functional dependence $Y = f \circ X$ have been introduced in Trutschnig (2011), in Dette et al. (2013) and, in a more statistical setting, in Chatterjee (2021).

In the following, we propose a new class of complete dependence measures fulfilling (C1), (C2) and (C3), which eliminates some of the above-described shortcomings of measures such as ρ, τ and σ_p. As an essential tool in the construction of these measures, we propose a transformation of an arbitrary copula C into a stochastically increasing copula $C^\uparrow$ (a property extensively discussed in Chapter 5). Most importantly, $C^\uparrow$ contains the entire complete dependence in the sense of Trutschnig (2011), Dette et al. (2013) and Chatterjee (2021).

The so-called (SI)-rearrangement $C^\uparrow$ is constructed from the original copula C via the rearrangement for each $v \in [0, 1]$ of the conditional expectation

$$\partial_1 C(u, v) = \mathbb{E}(V \leq v \mid U = u)$$

in a decreasing order with respect to u. This rearrangement technique is known as (the Hardy-Littlewood-Pólya) majorization (see Section 2.6). Most importantly, the rearrangement of C can drastically differ from the rearrangement of $C^\top$. In the most extreme cases $C^\uparrow$ equals C^+, whereas $(C^\top)^\uparrow$ is close to Π. This corresponds to the case that Y is completely dependent on X while X is 'almost' independent of Y. Precisely this asymmetry is the key to converting symmetric measures of dependence into (directed or asymmetric) measures of complete dependence that satisfy the axioms (C1) to (C3). Loosely speaking, the underlying construction principle is as follows: The rearrangement transforms arbitrary measurable dependence into a stochastically monotone relationship, which can then be quantified using a measure of dependence μ with

$$R_\mu(X, Y) := R_\mu(C) := \mu(C^\uparrow) \ .$$

We call measures of complete dependence R_μ of the above form *rearranged dependence measures*, where choices for μ include Spearman's ρ, Kendall's τ or the Schweizer-Wolff measures σ_p with $1 \leq p < \infty$. Contrary to the properties of the underlying measure μ, R_μ constitutes a genuine measure of complete dependence detecting arbitrary functional relationships. In case the underlying measure μ is a concordance measure κ, the rearranged concordance measure even yields a consistent notion of functional dependence. That is, the general functional influence of X on Y is at least as strong as the monotone influence given by κ, i.e. $|\kappa(X, Y)| \leq R_\kappa(X, Y)$. All rearranged dependence measures R_μ also comply with the data processing inequality known from information theory, i.e.

$$R_\mu(X, Y) = \sup_g R_\mu(g(X), Y) \ ,$$

where the supremum is taken over all measurable functions g. This immediately implies an asymmetric version of the self-equitability condition recently introduced in Kinney and Atwal (2014), which in turn yields Axiom (C3).

To empirically apply the rearranged dependence measures, we construct a simple estimator $\widehat{R}_\mu$ for R_μ based on the convergence results for empirical checkerboard copulas established in Junker et al. (2021), which impose no regularity conditions on the underlying copula C. We investigate the performance of $\widehat{R}_\mu$ in a simulation study and show its fast convergence properties with Spearman's ρ as the underlying dependence measure.

This chapter is structured as follows: Section 6.1 introduces the (SI)-rearrangement of a copula, while Section 6.2 provides an approximation result for the (SI)-rearrangement. Section 6.3 presents the majorization order as an order of variability. Section 6.4 establishes the class of rearranged dependence measures and their properties and Section 6.5 constructs a consistent estimator of R_μ.

6.1 The (SI)-rearrangement of copulas

Recall from Chapter 5 that a copula is called stochastically increasing (decreasing) in the first component if $u \mapsto \partial_1 C(u, v)$ is decreasing (increasing) for all $v \in [0, 1]$. Again, we will generally drop the specification 'in the first component' and refer to such a copula simply as 'stochastically increasing.' All copulas in this section are assumed to be 2-copulas, unless explicitly stated otherwise.

In this section, we decompose an arbitrary copula C into a *unique* stochastically increasing copula $C^{\uparrow}$ and a family of shuffles $(C_v)_{v\in[0,1]}$ via

$$C(u,v) = (C_v * C^{\uparrow})(u,v) \ ,$$

where $C^{\uparrow}$ contains the entire information about the complete dependence in the sense that

$$\zeta_p(C) = \zeta_p(C^{\uparrow}) \text{ and } r(C) = r(C^{\uparrow})$$

holds, where ζ_p and r are given in Definition 2.3.7 and Equation (2.5), respectively. We would like to point out that this is in sharp contrast to the well-known decomposition $C = R * L$ of a copula C into the product of a left- and a right-invertible copula, where, in general, neither factor contains the entire information about the complete dependence, as the following proposition shows.

Proposition 6.1.1. *Let $\Pi = R * L$ be a decomposition of the independence copula Π into the product of a right-invertible copula R and a left-invertible copula L. Then neither R nor L contains the entire information about the complete dependence, i.e. neither $r(R) = r(\Pi) = 0$ nor $r(L) = r(\Pi) = 0$ holds.*

Proof. By Theorem 2.2.8, Π can be decomposed into the product of a right- and a left-invertible copula. Assume $\Pi = R * L$ with $r(R) = 0$. Then, in view of $r(C) = 0$ if and only if $C = \Pi$, we must have $R = \Pi$ which is a contradiction since Π is not right-invertible. Analogously, we see that the case $r(L) = 0$ is not possible. □

For the remainder of this chapter, $(\partial_1 C)^*(u,v)$ denotes the decreasing rearrangement of $\partial_1 C(u,v)$ with respect to u, and, analogously, $(\partial_1 C)_*(u,v)$ denotes the increasing rearrangement (see Section 2.6).

Definition 6.1.2. *The stochastically increasing rearrangement, (SI)-rearrangement in short, of a copula C is defined as*

$$C^{\uparrow}(u,v) := \int_0^u (\partial_1 C)^*(s,v) \, \mathrm{d}s$$

for all $u, v \in [0,1]$. Analogously, the stochastically decreasing rearrangement, (SD)-rearrangement in short, of C is defined as

$$C^{\downarrow}(u,v) := \int_0^u (\partial_1 C)_*(s,v) \, \mathrm{d}s \ .$$

The following central result establishes that the (SI)-rearrangement $C^{\uparrow}$ is indeed a copula and, most importantly, contains (in some sense) the entire information about the complete dependence of C.[1]

Theorem 6.1.3. *1. The (SI)-rearrangement $C^{\uparrow}$ of a copula C is a stochastically increasing copula. Analogously, the (SD)-rearrangement $C^{\downarrow}$ is a stochastically decreasing copula.*

[1]Note that Ansari and Rüschendorf (2021) independently applied the decreasing rearrangement of a copula to investigate order properties of the generalized Markov product.

2. *A copula C is stochastically increasing if and only if $C = C^{\uparrow}$.*

3. *For any copula C, there exists a (not necessarily unique) family of completely dependent copulas $(C_v)_{v\in[0,1]}$ such that*

$$C(u,v) = (C_v * C^{\uparrow})(u,v) \ .$$

4. *The (SD)-rearrangement $C^{\downarrow}$ satisfies $C^{\downarrow} = C^{-} * C^{\uparrow}$.*

5. *$C^{\uparrow}$ exhibits the same degree of complete dependence as C with regard to several measures of complete dependence. More precisely, for any copula C, we have*

$$\zeta_p(C) = \zeta_p(C^{\uparrow}) = \zeta_p(C^{\downarrow}) \text{ as well as } r(C) = r(C^{\uparrow}) = r(C^{\downarrow}) \ .$$

Proof. 1. We give the proof for $C^{\uparrow}$; the proof for $C^{\downarrow}$ follows from Lemma 5.1.5. In order to show that $C^{\uparrow}$ is a copula, we verify the Properties 1 to 3 of Example 2.1.3 for the function $C^{\uparrow}$.

a) Considering $(\partial_1 C)^*(u,0) = 0^* = 0$, we have $C^{\uparrow}(u,0) = 0$, whereas $C^{\uparrow}(0,v) = 0$ follows from the integral representation of $C^{\uparrow}$.

b) By definition, we have

$$C^{\uparrow}(u,1) = \int_0^u (\partial_1 C)^*(s,1) \,\mathrm{d}s = \int_0^u 1^* \,\mathrm{d}s = u \ ,$$

and in view of Part 4 of Proposition 2.6.3, we further obtain that

$$C^{\uparrow}(1,v) = \int_0^1 \partial_1 C^{\uparrow}(t,v) \,\mathrm{d}t = \int_0^1 \partial_1 C^{\uparrow}(\sigma(t),v) \,\mathrm{d}t = \int_0^1 \partial_1 C(t,v) \,\mathrm{d}t = v \ .$$

c) Part 5 of Proposition 2.1.8 states that $0 \le \partial_1 C(\,\cdot\,,v_1) \le \partial_1 C(\,\cdot\,,v_2)$ whenever $v_1 \le v_2$. Combining this with Part 2 of Proposition 2.6.3 yields $(\partial_1 C)^*(\,\cdot\,,v_1) \le (\partial_1 C)^*(\,\cdot\,,v_2)$. Thus, the $C^{\uparrow}$-volume of the rectangle $[u_1,u_2) \times [v_1,v_2)$ satisfies

$$\begin{aligned} V_{C^{\uparrow}}\left([u_1,u_2) \times [v_1,v_2)\right) &= C^{\uparrow}(u_2,v_2) - C^{\uparrow}(u_1,v_2) - C^{\uparrow}(u_2,v_1) + C^{\uparrow}(u_1,v_1) \\ &= \int_{u_1}^{u_2} (\partial_1 C)^*(s,v_2) - (\partial_1 C)^*(s,v_1) \,\mathrm{d}s \\ &\ge 0 \end{aligned}$$

and is indeed nonnegative.

Finally, the fact that $C^{\uparrow}$ is stochastically increasing follows from the fact that $\partial_1 C^{\uparrow}(u,v)$ equals $(\partial_1 C)^*(u,v)$, which is decreasing in u due to Proposition 2.6.3.

2. First, $C = C^{\uparrow}$ implies that $u \mapsto C(u,v)$ is concave for all $v \in [0,1]$, which is equivalent to C being stochastically increasing. Conversely, suppose C is stochastically increasing and let f denote the right-hand derivative of $u \mapsto C(u,v)$. Then f is decreasing and right-continuous due to the concavity of C. An application of Theorem 4.2 in Chong and Rice (1971) yields $f = f^*$ and therefore

$$C(u,v) = \int_0^u \partial_1 C(u,v) \, \mathrm{d}t = \int_0^u \partial_1 C^*(u,v) \, \mathrm{d}t = C^{\uparrow}(u,v) \ .$$

3. Property 4 of Proposition 2.6.3 yields for each $v \in [0,1]$ the existence of a λ-preserving map σ_v such that

$$\partial_1 C(u,v) = \partial_1 C^{\uparrow}(\sigma_v(u), v)$$

holds for almost all $u \in [0,1]$. Note that $T_v f := f \circ \sigma_v$ constitutes a Markov operator on $L^1([0,1])$ and in turn corresponds to a completely dependent copula C_v by Theorem 2.2.9. Integrating the above equality gives the pointwise relationship

$$C(u,v) = (C_v * C^{\uparrow})(u,v) \ .$$

4. This follows immediately from $\partial_1(C^- * C)(u,v) = \partial_1 C(1-u,v)$, which holds for almost all $u \in [0,1]$ and every copula C.

5. This assertion is a consequence of Proposition 2.6.3, stating that the decreasing rearrangement preserves L^p-norms, i.e.

$$\int_0^1 |\partial_1 C(u,v) - v|^p \ \mathrm{d}u = \int_0^1 \left|\partial_1 C^{\uparrow}(u,v) - v\right|^p \ \mathrm{d}u$$

for all $v \in (0,1)$. Integrating with respect to v yields the desired result. □

Let us illustrate the construction of $C^{\uparrow}$ by two examples.

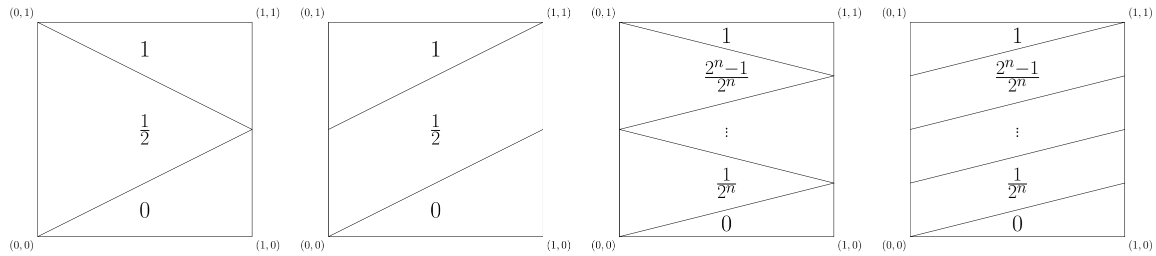
Example 6.1.4 (Shuffles of C^+)**.** *Suppose $\sigma : [0,1] \to [0,1]$ is a λ-preserving transformation. The (generalized) shuffle of C^+ with respect to σ is defined as*

$$\partial_1 C_\sigma(u,v) := \partial_1 C^+(\sigma(u), v) \in \{0,1\} \ .$$

Thus, C_σ is a rearrangement of C^+ and its (SI)-rearrangement fulfils $C_\sigma^{\uparrow} = C^+$. This property plays a central role in the characterization of complete dependence using copulas (see, for example, Lemma 10 in Trutschnig (2011)).

The next example illustrates that independence is a symmetric concept, whereas 'almost' independence is not.

Example 6.1.5 (n-fold gluing of the tent copula)**.** *Let us now construct a copula C for which $r(C)$ and $r(C^\top)$ differ dramatically, that is, a copula with a highly asymmetric dependence structure. To this end, we use the well-known tent copula (see Example 3.2.5 in Durante and*



(a) Support of $C_1^\top$. (b) Partial derivative of the (SI)-rearrangement of $C_1^\top$. (c) Support of $C_n^\top$. (d) Partial derivative of the (SI)-rearrangement of $C_n^\top$.

Figure 6.1: Figures (a) and (c) depict the support of the transposed tent copula $C_1^\top$ and of the transposed n-fold glued iteration $C_n^\top$, respectively. Additionally, the values of $\partial_1 C_1^\top(u,v)$ and $\partial_1 C_n^\top(u,v)$ on their respective triangles are shown. Figures (b) and (d) depict the values of the partial derivative for the (SI)-rearrangement of $C_1^\top$ and $C_n^\top$, respectively.

Sempi (2016)) as a building block in the iterated application of the gluing technique introduced in Siburg and Stoimenov (2008b). The tent copula is given by

$$C_1 := C^+ \circledast_{u=1/2} C^- \ ,$$

where $\circledast$ denotes the gluing of C^+ and C^- along $u = 1/2$, and its n-fold gluing by

$$C_{n+1} = C_n \circledast_{u=1/2} C_n \ .$$

Following Example 1 in Siburg and Stoimenov (2015), C_n is completely dependent and therefore fulfils $C_n^\uparrow = C^+$. In contrast, $C_n^\top$ is almost independent with $r(C_n^\top) \to 0$ as n tends to infinity. The support of the transposed n-fold gluing of the tent copula is depicted in Figure 6.1 (a) and (c). A short calculation yields

$$(\partial_1 C_n^\top)^*(u,v) = \frac{k}{2^n} + \frac{1}{2^n}\mathbb{1}_{[0,2^n(v-k2^{-n})]}(u) \tag{6.1}$$

for $v \in [k2^{-n}, (k+1)2^{-n})$. Integrating the above equation leads to $(C_n^\top)^\uparrow$, which converges towards Π uniformly and with respect to the metric D_p for $1 \le p < \infty$, due to

$$\left|\partial_1 (C_n^\top)^\uparrow(u,v) - \partial_1 \Pi(u,v)\right| \le \frac{1}{2^n}$$

for almost all $u \in [0,1]$ and all $v \in [0,1]$. Alternatively, the convergence with respect to D_p follows from Proposition 5.1.6 by exploiting the uniform convergence together with the fact that all $C_n^\uparrow$ are stochastically increasing. The values of the partial derivatives of the (SI)-rearrangement of $C_1^\top$ and $C_n^\top$ are depicted in Figure 6.1 (b) and (d), respectively.

The rearrangement of copulas can also be used to construct new copulas from existing families. This is illustrated by the following example.

Example 6.1.6 (Convex combinations). *While the (SI)-rearrangement of* $\lambda C + (1-\lambda)\Pi$ *for* $\lambda \in [0,1]$ *with an arbitrary copula* C *equals*

$$(\lambda C + (1-\lambda)\Pi)^{\uparrow} = \lambda C^{\uparrow} + (1-\lambda)\Pi \ ,$$

the (SI)-rearrangement does not, in general, commute with convex combinations. To see this, consider the convex combination

$$C_\lambda := \lambda C^{+} + (1-\lambda)C^{-} \ .$$

A short calculation yields

$$(\partial_1 C_\lambda)^*(u,v) = \begin{cases} \lambda^* \mathbb{1}_{[0,v]}(u) + (1-\lambda^*)\mathbb{1}_{(v,2v]}(u) & \text{for } v \in \left[0, \frac{1}{2}\right) \\ \lambda^* \mathbb{1}_{[0,v]}(u) + (1-\lambda^*)\mathbb{1}_{[0,2v-1]}(u) + (1-\lambda^*)\mathbb{1}_{(v,1]}(u) & \text{otherwise} \end{cases}$$

with $\lambda^* := \max\{\lambda, 1-\lambda\}$. *By integrating the above equation, we obtain the closed-form expression for the (SI)-rearrangement of* C_λ*:*

$$\lambda^* C^{+}(u,v) + (1-\lambda^*)\min\{u, (2v-1)_+\} + (1-\lambda^*)\min\{(u-v)_+, v\} \ ,$$

where $(x)_+$ *denotes* $\max\{x, 0\}$. *Therefore,* $(C_\lambda)^{\uparrow} = (\lambda C^{+} + (1-\lambda)C^{-})^{\uparrow}$ *differs from* $\lambda (C^{+})^{\uparrow} + (1-\lambda)(C^{-})^{\uparrow} = C^{+}$ *whenever* $\lambda \in (0,1)$.

While the exact behaviour of convex combinations under the rearrangement can generally not be derived, the convex combination of the (SI)-rearrangements establishes an upper bound.

Proposition 6.1.7. *Let* C_1 *and* C_2 *be copulas. Then it holds for all* $\lambda \in [0,1]$

$$(\lambda C_1 + (1-\lambda)C_2)^{\uparrow} \le \lambda C_1^{\uparrow} + (1-\lambda)C_2^{\uparrow} \ .$$

Proof. First, a simple rephrasing of the (SI)-rearrangement yields

$$C_1^{\uparrow} = \int_0^u \partial_1 C_1^*(s,v) \, \mathrm{d}s = u \cdot \frac{1}{u} \int_0^u \partial_1 C_1^*(s,v) \, \mathrm{d}s =: u(\partial_1 C_1)^{**}(u,v) \ ,$$

where $(\partial_1 C_1)^{**}(u,v)$ is the so-called maximal function of $\partial_1 C_1$. Contrary to the decreasing rearrangement, the maximal function is subadditive (see Theorem 2.3.4 in Bennett and Sharpley (1988)), i.e. fulfils

$$(\partial_1(\lambda C_1 + (1-\lambda)C_2))^{**}(u,v) \le \lambda(\partial_1 C_1)^{**}(u,v) + (1-\lambda)(\partial_1 C_2)^{**}(u,v)$$

for all $u, v \in [0,1]$, which yields the assertion. □

It turns out that the rearrangements $C^{\downarrow}$ and $C^{\uparrow}$ yield lower and upper bounds for C, improving the Fréchet-Hoeffding bounds C^{-} and C^{+}.[2]

Theorem 6.1.8. *For any copula* C*, we have for all* $u, v \in [0,1]$,

$$C^{\downarrow}(u,v) \le C(u,v) \le C^{\uparrow}(u,v) \ .$$

[2] Of course, the Fréchet-Hoeffding bounds are uniform bounds independent of C, in contrast to $C^{\downarrow}$ and $C^{\uparrow}$.

Proof. The Hardy-Littlewood inequality (see Property 1 of Proposition 2.6.7) yields

$$\begin{aligned} C(u,v) = (C^+ * C)(u,v) &= \int_0^1 \partial_2 C^+(u,t)\partial_1 C(t,v)\ \mathrm{d}t \\ &\leq \int_0^1 \partial_2 C^+(u,t)(\partial_1 C)^*(t,v)\ \mathrm{d}t \\ &= (C^+ * C^\uparrow)(u,v) = C^\uparrow(u,v)\ . \end{aligned}$$

The lower bound follows analogously using

$$(\partial_1 C)^*(1-t,v) = \partial_1(C^- * C^\uparrow)(t,v) = \partial_1 C^\downarrow(t,v)\ . \qquad \square$$

These pointwise bounds immediately extend to bivariate concordance measures κ, such as Spearman's ρ and Kendall's τ.[3]

Corollary 6.1.9. *For every copula C and every concordance measure κ, we have the bounds*

$$-\kappa(C^\uparrow) = \kappa(C^\downarrow) \leq \kappa(C) \leq \kappa(C^\uparrow)\ .$$

Considering that the (SI)-rearrangement $C^\uparrow$ of C is a rearrangement of the partial derivatives, the pointwise bounds given in Theorem 6.1.8 appear rather crude. Therefore, we will now investigate the behaviour of the rearrangement with regard to D_p, i.e. the L^p-distance of the partial derivatives.

Theorem 6.1.10. *Let C and D be copulas. Then, for $1 \leq p < \infty$,*

$$D_p(C^\uparrow, D^\uparrow) \leq D_p\left(C, D\right) \leq D_p(C^\uparrow, D^\downarrow) = D_p(C^\uparrow, C^- * D^\uparrow)\ .$$

Remark 6.1.11. *Of course, the lower and upper bounds in Theorem 6.1.10 are also attained by any other copula pair $C_\sigma * C^\uparrow$ and $C_\sigma * D^\downarrow$ instead of $C^\uparrow$ and $D^\downarrow$, where C_σ is a left-invertible copula. This follows immediately from Proposition 2.6.3 due to*

$$\begin{aligned} D_p^p(C_\sigma * A, C_\sigma * B) &= \int_{[0,1]^2} |\partial_1 A(\sigma(u),v) - \partial_1 B(\sigma(u),v)|^p\ \mathrm{d}\lambda(u,v) \\ &= \int_{[0,1]^2} |\partial_1 A(u,v) - \partial_1 B(u,v)|^p\ \mathrm{d}\lambda(u,v) \\ &= D_p^p(A,B) \end{aligned}$$

for all copulas A and B.

Proof of Theorem 6.1.10. Property 2 of Proposition 2.6.7 yields

$$\partial_1 C^\uparrow(\,\cdot\,,v) - \partial_1 D^\uparrow(\,\cdot\,,v) \preceq \partial_1 C(\,\cdot\,,v) - \partial_1 D(\,\cdot\,,v)$$

[3] Concordance measures are monotone with respect to the pointwise ordering $C_1 \leq C_2$ (see Section 2.3).

for all $v \in [0,1]$, where $\preceq$ denotes the majorization order given in Definition 2.6.5. Thus, due to Proposition 2.6.6, we have for all $v \in [0,1]$ and any $1 \leq p < \infty$ that

$$\int_0^1 \left|\partial_1 C^{\uparrow}(u,v) - \partial_1 D^{\uparrow}(u,v)\right|^p \, \mathrm{d}u \leq \int_0^1 |\partial_1 C(u,v) - \partial_1 D(u,v)|^p \, \mathrm{d}u$$

and integrating with respect to v yields the desired result. The second inequality follows in the same way. □

Theorem 6.1.10 implies the continuity of the mapping $C \mapsto C^{\uparrow}$ with respect to D_p. This property will play a crucial role in approximating the (SI)-rearrangement via checkerboard copulas in Section 6.2.

Corollary 6.1.12. *The mapping $C \mapsto C^{\uparrow}$ is continuous with respect to D_p for all $1 \leq p \leq \infty$.*

Proof. First, suppose $(C_n)_{n\in\mathbb{N}}$ converges towards C with respect to D_p for some $1 \leq p < \infty$. Then an application of Theorem 6.1.10 yields

$$0 \leq D_p(C_n^{\uparrow}, C^{\uparrow}) \leq D_p(C_n, C) \to 0 \ .$$

The continuity with respect to D_∞ follows from the equivalence of D_1 and D_∞. □

Example 6.1.13. *Consider two completely dependent copulas C_1 and C_2. Then their (SI)-rearrangement is C^+, while their (SD)-rearrangement is C^-. Thus,*

$$0 = D_p(C^+, C^+) \leq D_p(C_1, C_2) \leq D_p(C_1^{\uparrow}, C_2^{\downarrow}) = D_p(C^+, C^-) = \frac{1}{2^{1/p}} \ .$$

This result is reminiscent of the diameter of $(\mathcal{C}_2, \langle \cdot \ , \ \cdot \rangle_S)$ presented in Corollary 15 of Siburg and Stoimenov (2008a).

The final theorem of this section gives a geometric characterization quite similar to the findings of Siburg and Stoimenov (2011) in the setting of inner products.

Theorem 6.1.14. *Let C_1, C_2 be copulas with $r(C_1) = \rho_1$ and $r(C_2) = \rho_2$. Then the following assertions are equivalent:*

1. $D_2(C_1, C_2) = \min\{D_2(C, D) \mid r(C) = \rho_1, r(D) = \rho_2\}$.
2. $\langle \partial_1 C_1 \, , \, \partial_1 C_2 \rangle_{L^2([0,1]^2)} = \max\left\{\langle \partial_1 C \, , \, \partial_1 D \rangle_{L^2([0,1]^2)} \mid r(C) = \rho_1, r(D) = \rho_2\right\}$.

Moreover, in each of the two cases, C_1 and C_2 can be chosen to be stochastically increasing.

Proof. A straightforward calculation yields

$$\begin{aligned} D_2^2(C,D) &= \int_{[0,1]^2} \partial_1 C(u,v)^2 - 2\partial_1 C(u,v)\partial_1 D(u,v) + \partial_1 D(u,v)^2 \, \mathrm{d}\lambda(u,v) \\ &= \frac{r(C)}{6} + \frac{r(D)}{6} + \frac{2}{3} - 2 \int_{[0,1]^2} \partial_1 C(u,v)\partial_1 D(u,v) \, \mathrm{d}\lambda(u,v) \\ &= \frac{\rho_1}{6} + \frac{\rho_2}{6} + \frac{2}{3} - 2 \langle \partial_1 C \, , \, \partial_1 D \rangle_{L^2([0,1]^2)} \ , \end{aligned}$$

where we have used the identity

$$r(C) = 6 \int_{[0,1]^2} \partial_1 C(u,v)^2 \,\mathrm{d}\lambda(u,v) - 2 \,.$$

Thus, $D_2(C, D)$ is minimal if and only if $\langle \partial_1 C \,,\, \partial_1 D\rangle$ is maximal. Due to Theorem 6.1.3 and Theorem 6.1.10, C_1 and C_2 can be chosen to be stochastically increasing. □

6.2 Approximating the (SI)-rearrangement

In general, the computation of the rearrangement of a function, and hence the computation of $C^{\uparrow}$, may be a difficult task. In this section, we propose a simple approximation scheme for $C^{\uparrow}$ based on the concept of checkerboard copulas, thereby circumventing the need to treat partial derivatives. Checkerboard copulas are an important tool in statistical applications and are constructed from doubly stochastic matrices. Recall from Definitions 2.5.15 and 2.5.16, that for a doubly stochastic matrix A, the copula

$$C_n^{\#}(A)(u,v) := n \sum_{k,\ell=1}^{n} a_{k\ell} \int_0^u \mathbb{1}_{\left[\frac{k-1}{n}, \frac{k}{n}\right)}(s) \,\mathrm{d}s \int_0^v \mathbb{1}_{\left[\frac{\ell-1}{n}, \frac{\ell}{n}\right)}(t) \,\mathrm{d}t$$

is called checkerboard copula.[4] For a copula C, the induced checkerboard copula is defined as

$$C_n^{\#}(C) := C_n^{\#}(A_n)$$

for the doubly stochastic matrix A_n with

$$(A_n)_{k\ell} := n \cdot V_C\left(\left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{\ell-1}{n}, \frac{\ell}{n}\right]\right) \,.$$

We point out that the partial derivatives of $C^{\#}(A)$ are piecewise constant. Moreover, a direct calculation shows that $C^{\#}(A)^{\uparrow} = C^{\#}(A)$ if and only if

$$\sum_{j=1}^{\ell} a_{k_2 j} \le \sum_{j=1}^{\ell} a_{k_1 j} \tag{6.2}$$

for all $\ell \in \{1, \ldots, n\}$ and all $k_1 \le k_2$, that is, if and only if the rows of A are ordered with respect to the majorization ordering of vectors (see Marshall et al. (2011)). This suggests the following algorithm to calculate the (SI)-rearrangement of an arbitrary checkerboard copula:

Theorem 6.2.1. *For a doubly stochastic matrix $A \in \mathbb{R}^{n\times n}$, it holds*

$$C_n^{\#}(A)^{\uparrow} = C_n^{\#}(A^{\uparrow}) \,,$$

where $A^{\uparrow} = (a_{k\ell}^{\uparrow})$ is the doubly stochastic matrix constructed via the following algorithm:

[4]Whenever it is unambiguous, we will drop the index n and simply write $C^{\#}(A)$.

1. *Calculate* $B_k^\ell := \partial_1 C_n^\#(A)\left(\frac{k}{n}, \frac{\ell}{n}\right) = \sum_{j=1}^{\ell} a_{kj}$ *and set* $B_k^0 := 0$.

2. *For each fixed* $\ell = 0, \ldots, n$, *sort* B^ℓ *in a decreasing manner and denote the result by* $\widetilde{B}^\ell$. *Then* $\widetilde{B}_k^\ell = \partial_1 C_n^\#(A)^\uparrow\left(\frac{k}{n}, \frac{\ell}{n}\right)$ *holds.*

3. *Calculate* $a_{k\ell}^\uparrow$ *iteratively using*

$$a_{k\ell}^\uparrow := \widetilde{B}_k^\ell - \widetilde{B}_k^{\ell-1} \geq 0 \ .$$

Proof. The equality $C_n^\#(A)^\uparrow = C^\#(A^\uparrow)$ follows directly from the definition of the algorithm. It remains to show that $A^\uparrow$ is indeed doubly stochastic. To do so, we simply calculate

$$\sum_{\ell=1}^{n} a_{k\ell}^\uparrow = \sum_{\ell=1}^{n} \widetilde{B}_k^\ell - \widetilde{B}_k^{\ell-1} = \widetilde{B}_k^n - \widetilde{B}_k^0 = \widetilde{B}_k^n = 1$$

as well as

$$\begin{aligned}\sum_{k=1}^{n} a_{k\ell}^\uparrow &= \sum_{k=1}^{n} \widetilde{B}_k^\ell - \widetilde{B}_k^{\ell-1} = \sum_{k=1}^{n} B_k^\ell - B_k^{\ell-1} \\ &= \sum_{j=1}^{\ell}\sum_{k=1}^{n} a_{kj} - \sum_{j=1}^{\ell-1}\sum_{k=1}^{n} a_{kj} = \ell - (\ell-1) = 1 \ .\end{aligned}$$

The nonnegativity of $a_{k\ell}^\uparrow$ follows from the algorithm. □

Example 6.2.2. *For the doubly stochastic matrix* A *and its rearrangement* $A^\uparrow$ *given by*

$$A := \begin{pmatrix} 3/8 & 1/8 & 1/6 & 2/6 \\ 1/8 & 3/8 & 2/6 & 1/6 \\ 1/8 & 3/8 & 2/6 & 1/6 \\ 3/8 & 1/8 & 1/6 & 2/6 \end{pmatrix} \quad \text{and } A^\uparrow = \begin{pmatrix} 3/8 & 1/8 & 2/6 & 1/6 \\ 3/8 & 1/8 & 2/6 & 1/6 \\ 1/8 & 3/8 & 1/6 & 2/6 \\ 1/8 & 3/8 & 1/6 & 2/6 \end{pmatrix} ,$$

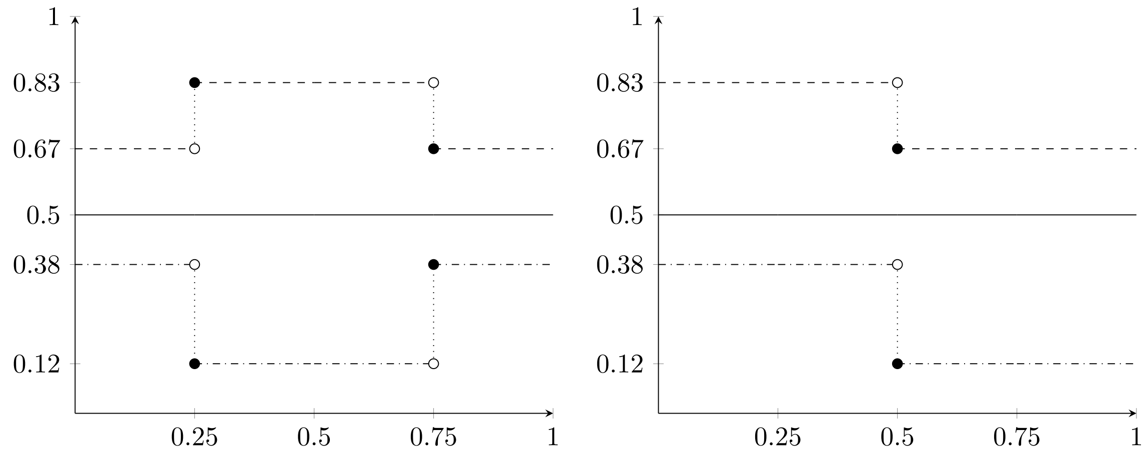
the partial derivatives of the corresponding checkerboard copulas are depicted in Figure 6.2. A corresponding λ*-preserving transformation is, for example,*

$$\sigma_v(u) = \begin{cases} (u + 1/4) \bmod 1 & \text{for } v \in \left[0, \frac{1}{2}\right) \\ (u + 3/4) \bmod 1 & \text{for } v \in \left[\frac{1}{2}, 1\right] \end{cases} .$$

Note that the choice of σ_v *is not unique since the same rearrangement is also attained by*

$$\sigma_v(u) = \begin{cases} (u + 1/2\mathbb{1}_{(1/4,1/2]\cup(3/4,1]}(u)) \bmod 1 & \text{for } v \in \left[0, \frac{1}{2}\right) \\ (u + 1/2\mathbb{1}_{(0,1/4]\cup(1/2,3/4]}(u)) \bmod 1 & \text{for } v \in \left[\frac{1}{2}, 1\right] \end{cases} .$$

While the decreasing rearrangement is well-defined for all copulas, the partial derivative may not always be tractable. Nevertheless, we will be able to approximate the (SI)-rearrangement using checkerboard copulas, as the next result shows.



(a) Plot of $u \mapsto \partial_1 C^{\#}(A)(u,v)$.

(b) Plot of $u \mapsto \partial_1 C^{\#}(A^{\uparrow})(u,v)$.

Figure 6.2: Plots of the partial derivative of the checkerboard copulas $C^{\#}(A)$ and $C^{\#}(A^{\uparrow})$ for fixed v. The plots for $v = 1/4, 1/2$ and $3/4$ are depicted by the dash-dotted, solid and dashed line, respectively.

Theorem 6.2.3. *For any copula C, we have $D_p(C_n^{\#}(A_n^{\uparrow}), C^{\uparrow}) \to 0$, where $A_n^{\uparrow}$ is the doubly stochastic matrix associated with $C_n^{\#}(C)^{\uparrow}$.*

Due to Theorem 6.2.3, $C^{\uparrow}$ can be approximated using the algorithm presented in Theorem 6.2.1.

Proof. This follows from the continuity result in Corollary 6.1.12 and Proposition 2.5.18, due to

$$D_p(C^{\#}(A_n^{\uparrow}), C^{\uparrow}) = D_p(C_n^{\#}(C)^{\uparrow}, C^{\uparrow}) \leq D_p(C_n^{\#}(C), C) \to 0 \ . \qquad \square$$

6.3 Majorization order

In this section, we introduce the majorization order for copulas, which measures the variability of conditional distributions with respect to the conditioning variable in a similar way to the convex order of random variables. Recall that the majorization order $\preceq$ for functions in Definition 2.6.5 uses the concept of a decreasing rearrangement and that we always rearrange a bivariate function with respect to its first variable.

Definition 6.3.1. *A copula C_1 is said to be smaller than a copula C_2 in the majorization order, written $C_1 \preceq_m C_2$, if and only if $\partial_1 C_1(\,\cdot\,, v) \preceq \partial_1 C_2(\,\cdot\,, v)$ holds for all $v \in [0,1]$.*

Note that in view of Definition 2.6.5, $C_1 \preceq_m C_2$ is equivalent to the condition

$$\int_0^u (\partial_1 C_1)^*(s,v)\ \mathrm{d}s \leq \int_0^u (\partial_1 C_2)^*(s,v)\ \mathrm{d}s$$

for all $u, v \in [0,1]$.[5]

On the subset of stochastically increasing copulas, the majorization order is equivalent to the usual stochastic order as the following proposition shows.

Proposition 6.3.2. *If C_1 and C_2 are stochastically increasing copulas, then $C_1 \preceq_m C_2$ is equivalent to $C_1 \le C_2$.*

Proof. The proof is a simple application of the identity

$$C(u,v) = \int_0^u \partial_1 C(s,v)\,\mathrm{d}s = \int_0^u (\partial_1 C)^*(s,v)\,\mathrm{d}s\ ,$$

which is valid for stochastically increasing copulas C since then $\partial_1 C = (\partial_1 C)^*$. Furthermore, equality holds in case of $u = 1$ due to the uniform margin property of copulas. □

The following result collects various equivalent descriptions of two copulas being ordered with respect to $\preceq_m$.

Theorem 6.3.3. *For any two copulas C_1 and C_2, the following assertions are equivalent:*

1. $C_1 \preceq_m C_2$.
2. $C_1^{\uparrow} \le C_2^{\uparrow}$.
3. *There exists a family $(C_v)_{v\in[0,1]}$ of copulas such that*

$$C_1^{\uparrow}(u,v) = \left(C_v * C_2^{\uparrow}\right)(u,v)\ .$$

4. *There exists a family $(C_v)_{v\in[0,1]}$ of copulas such that*

$$C_1(u,v) = (C_v * C_2)(u,v)\ .$$

Proof. The equivalences follow from the respective definitions in combination with Proposition 6.3.2 and Part 3 of Proposition 2.6.6. □

The next proposition states that the Markov product preserves the ordering with respect to the majorization order.

Proposition 6.3.4. *For two copulas C_1 and C_2, we have*

$$(C_1 * C_2) \preceq_m C_2$$

as well as $\Pi \preceq_m C_1 \preceq_m C^+$. If C_1 is additionally stochastically increasing, $\Pi \le C_1 \le C^+$ holds.

[5] The majorization order is called Schur order in Definition 2.15 of Ansari (2019), where it was defined as a technical notion.

Proof. The first result is a direct consequence of Proposition 2.6.6 and the fact that

$$\partial_1(C_1 * C_2)(u,v) = \partial_u \int_0^1 \partial_2 C_1(u,t)\partial_1 C_2(t,v)\ \mathrm{d}t = T_{C_1}\partial_1 C_2(\,\cdot\,,v)(u)\ .$$

In turn, the first result implies the second one since $(\Pi * C_1) = \Pi$ and $C_1 = (C_1 * C^+)$. The third result follows from a combination of the second result and Proposition 6.3.2. □

Recall the two measures of complete dependence r and ζ_p as well as the measure of mutual complete dependence ω from Section 2.3. It turns out that all these measures are monotone with respect to the majorization order.

Theorem 6.3.5. *The measures of complete dependence r and ζ_p for $1 \le p < \infty$ are monotone with respect to the majorization order. Likewise, on the subset of symmetric copulas, the measure of complete dependence ω is monotone with respect to the majorization order.*

Proof. If $C_1 \preceq_m C_2$, then $r(C_1) \le r(C_2)$ as well as $\zeta_p(C_1) \le \zeta_p(C_2)$ due to the definitions of the respective measures together with Theorem 2.6.6. The same is true for the symmetrized measure ω if both C_1 and C_2 are symmetric themselves. □

Theorem 6.3.6. *Let C_1 and C_2 be stochastically increasing copulas with $C_1 \le C_2$. Then either $r(C_1) < r(C_2)$ or $C_1 = C_2$.*

Proof. Since $C_1 \preceq_m C_2$ and both C_1 and C_2 are stochastically increasing, an application of Hardy's Lemma (see Proposition 2.3.6 in Bennett and Sharpley (1988)) yields

$$\int_0^1 \partial_1 C_1(u,v)\partial_1 C_1(u,v)\ \mathrm{d}u \le \int_0^1 \partial_1 C_1(u,v)\partial_1 C_2(u,v)\ \mathrm{d}u$$

for all $v \in [0,1]$. Integrating leads to

$$\int_0^1\int_0^1 \partial_1 C_1(u,v)\partial_1 C_1(u,v)\ \mathrm{d}u\ \mathrm{d}v \le \int_0^1\int_0^1 \partial_1 C_1(u,v)\partial_1 C_2(u,v)\ \mathrm{d}u\ \mathrm{d}v\ ,$$

which is the first component of the inner product structure introduced in Siburg and Stoimenov (2008a). A straightforward calculation similar to that of Theorem 6.1.14 then yields

$$\begin{aligned}
0 &\le D_2^2(C_1,C_2)\\
&= \frac{r(C_2)}{6} + \frac{r(C_1)}{6} + \frac{2}{3} - 2\int_0^1\int_0^1 \partial_1 C_1(u,v)\partial_1 C_2(u,v)\ \mathrm{d}u\ \mathrm{d}v\\
&\le \frac{r(C_2)}{6} + \frac{r(C_1)}{6} + \frac{2}{3} - 2\int_0^1\int_0^1 \partial_1 C_1(u,v)\partial_1 C_1(u,v)\ \mathrm{d}u\ \mathrm{d}v\\
&= \frac{r(C_2)}{6} + \frac{r(C_1)}{6} - 2\frac{r(C_1)}{6}\\
&= \frac{r(C_2) - r(C_1)}{6}
\end{aligned}$$

and the statement follows from Theorem 6.3.5. □

Corollary 6.3.7. *For arbitrary copulas C_1 and C_2 with $C_1 \preceq_m C_2$, we have either $r(C_1) < r(C_2)$ or $C_1^\uparrow = C_2^\uparrow$.*

In other words, any two ordered copulas C_1 and C_2 with $C_1 \preceq_m C_2$ are either rearrangements of each other, i.e. $C_1 = (C_v * C_2)$ for a family of shuffles $(C_v)_{v\in[0,1]}$, or possess strictly different degrees of complete dependence as measured by r or ζ_2.

6.4 Rearranged dependence measures

According to Section 6.1, the (SI)-rearrangement contains the entire information about the (directed) complete dependence between two random variables. This property lays the foundation for our new class of complete dependence measures henceforth called rearranged dependence measures. The underlying construction principle for these rearranged dependence measures can be described as follows: The (SI)-rearrangement transforms arbitrary functional dependence into a stochastically increasing relationship, which we can quantify using other measures of dependence, such as concordance measures. The rearranged dependence measures are shown to be monotone with respect to the majorization order $\preceq_m$, which immediately implies the so-called data processing inequality, and they are shown to satisfy axioms similar to those of a measure of regression dependence stated in Definition 2 of Dette et al. (2013).

Recall that we denote the set of all stochastically increasing copulas by $\mathcal{C}^\uparrow$.

Definition 6.4.1. *We call $\mu : \mathcal{C}^\uparrow \to [0,1]$ a measure of (SI)-dependence if μ satisfies the following conditions:*

1. $\mu(C) = 0$ *if and only if* $C = \Pi$.
2. $\mu(C) = 1$ *if and only if* $C = C^+$.
3. $\mu(C_1) \le \mu(C_2)$ *whenever* $C_1 \le C_2$.
4. $\mu(\widehat{C}) = \mu(C)$, *where* $\widehat{C}(u,v) := u + v - 1 + C(1-u, 1-v)$ *denotes the survival copula.*
5. μ *is continuous with respect to pointwise convergence.*

Due to the fact that μ is defined on $\mathcal{C}^\uparrow$ instead of $\mathcal{C}$, Properties 1 to 5 of Definition 6.4.1 are significantly weaker than the properties of general concordance or dependence measures. For example, Spearman's ρ, when restricted to $\mathcal{C}^\uparrow$, constitutes a measure of (SI)-dependence, even though Property 1 does not hold for arbitrary copulas since

$$\rho(\Pi) = 0 = \rho\left(\frac{C^+ + C^-}{2}\right) .$$

The following definition introduces the main concept of this section.

Definition 6.4.2. *Let X and Y be continuous random variables with copula C_{XY} and μ a measure of (SI)-dependence. Then the associated rearranged dependence measure R_μ is defined as*

$$R_\mu(X,Y) := R_\mu(C_{XY}) := \mu(C_{XY}^\uparrow) .$$

Whenever we subsequently write $R_\mu(X,Y)$, we implicitly assume X and Y to be continuous random variables. Note that if Y is stochastically increasing in X, we have $C^\uparrow = C$ and the rearranged dependence measure reduces to the underlying measure, i.e. $R_\mu(C) = \mu(C)$. This observation will simplify the calculation of $R_\mu(C)$ for many common copula families later on.

Before investigating various properties of rearranged dependence measures, we introduce two specific classes, R_p and R_κ, based on Schweizer-Wolff measures σ_p and measures of concordance κ, respectively. The ability of R_p and R_κ to measure complete dependence is in stark contrast to the behaviour of the underlying L^p-distances and measures of concordance, which are only able to detect strictly monotone and monotone relationships between the two random variables, respectively.

Example 6.4.3. *Each L^p-norm with $1 \le p < \infty$ can be used to define the so-called Schweizer-Wolff measure (see Schweizer and Wolff (1981))*

$$\sigma_p(C) := \frac{\|C - \Pi\|_p}{\|C^+ - \Pi\|_p} \ .$$

We now show that each σ_p is a measure of (SI)-dependence by checking the five conditions stated in Definition 6.4.1:

1. *$\sigma_p(\Pi) = 0$ follows from $\Pi^\uparrow = \Pi$. Moreover, $0 = \sigma_p(C) = \|C^\uparrow - \Pi\|_p$ implies $C^\uparrow = \Pi$. Thus, $C^\uparrow$ is convex and concave in the first component, i.e. $u \mapsto C(u,v)$ is linear for all $v \in [0,1]$, which is equivalent to $C = \Pi$.*

2. *Now suppose C is completely dependent. Then $C^\uparrow = C^+$, which yields $\sigma_p(C^\uparrow) = 1$. On the other hand, if C is not completely dependent, then $C^\uparrow < C^+$ holds on a set of positive measure. Thus,*

$$\sigma_p(C) = \frac{\left\|C^\uparrow - \Pi\right\|_p}{\|C^+ - \Pi\|_p} < \frac{\|C^+ - \Pi\|_p}{\|C^+ - \Pi\|_p} = 1 \ .$$

3. *The assertion follows from $|C_1 - \Pi|^p \le |C_2 - \Pi|^p$ due to $\Pi \le C_1 \le C_2$.*

4. *This follows from $\Pi(u,v) = \widehat{\Pi}(u,v)$.*

5. *The continuity property of σ_p is a consequence of the corresponding property of the L^p-norm.*

The rearranged Schweizer-Wolff measure R_p associated with σ_p is defined as

$$R_p(C) := R_{\sigma_p}(C) = \frac{\left\|C^\uparrow - \Pi\right\|_p}{\|C^+ - \Pi\|_p} \ .$$

For simplicity, we denote the rearranged Schweizer-Wolff measure R_1 by R.

Surprisingly, it turns out that R allows a representation in terms of Spearman's ρ.

Proposition 6.4.4. *The rearranged Schweizer-Wolff measure R can be written in terms of Spearman's ρ as*

$$R(C) = \rho(C^\uparrow) = \frac{\rho(C^\uparrow) - \rho(C^\downarrow)}{2} \ .$$

Proof. In view of $\Pi \leq C^{\uparrow}$, we have

$$R(C) = 12 \int_{[0,1]^2} C^{\uparrow}(u,v) - \Pi(u,v) \, \mathrm{d}\lambda(u,v)$$
$$= 12 \int_{[0,1]^2} C^{\uparrow}(u,v) \, \mathrm{d}\lambda(u,v) - 3 = \rho(C^{\uparrow}) = \frac{\rho(C^{\uparrow}) - \rho(C^{\downarrow})}{2} \, .$$

The last equality follows from $\rho(C^{\downarrow}) = \rho(C^{-} * C^{\uparrow}) = -\rho(C^{\uparrow})$, using the identity for reflected copulas (see Property 5 of Definition 2.3.9). □

We now turn to the aforementioned second class of rearranged dependence measures, those based on measures of concordance.

Example 6.4.5. *Any concordance measure κ fulfils Properties 3, 4 and 5 of Definition 6.4.1 by virtue of the axioms stated in Definition 2.3.9. In particular, it also holds that $-1 \leq \kappa(C) \leq 1$ with $\kappa(\Pi) = 0$ and $\kappa(C^{+}) = 1$. Thus, in order to obtain measures of (SI)-dependence, we only have to impose the following nondegeneracy condition: Assume that for every stochastically increasing copula C, we have*

$$\kappa(C) = 0 \iff C = \Pi$$

as well as

$$\kappa(C) = 1 \iff C = C^{+} \, .$$

This condition can be verified for Spearman's ρ, Kendall's τ and Gini's γ using their representations via the concordance functional Q (see Section 5.1 in Nelsen (2006)). In contrast, Blomqvist's β (see Nelsen (2006)) does not satisfy the nondegeneracy condition since the ordinal sum

$$C(u,v) = \begin{cases} 2\Pi(u,v) & \text{if } (u,v) \in (0,1/2)^2 \\ C^{+} & \text{else} \end{cases}$$

is stochastically increasing with $C \neq C^{+}$, yet $\beta(C) = 4C(1/2,1/2) - 1 = 1 = \beta(C^{+})$.

Any measure of concordance κ satisfying the nondegeneracy condition is, by definition, a measure of (SI)-dependence, and the corresponding rearranged concordance measure R_{κ} is given by

$$R_{\kappa}(C) = \kappa(C^{\uparrow}) \, .$$

We point out that $R = R_1 = R_{\kappa}$ is both a rearranged Schweizer-Wolff measure (for $p = 1$) and a rearranged concordance measure (for $\kappa = \rho$).

Remark 6.4.6. *For the sake of completeness, let us discuss the rearranged dependence measures constructed from measures of complete dependence such as r. By definition, r fulfils Properties 1 and 2 of Definition 6.4.1, whereas Properties 3 and 5 follow from Theorem 6.3.3 and Proposition 5.1.6, respectively. A short calculation also yields Property 4, such that r is a measure of (SI)-dependence. With Theorem 6.1.3, the associated rearranged dependence measure reduces to*

$$R_r(C) = r(C^{\uparrow}) = r(C) \, .$$

The following key result shows that rearranged dependence measures are genuine measures of complete dependence. In particular, they characterize the two extreme cases of independence and general functional dependence.

Theorem 6.4.7. *Let X and Y be continuous random variables and let μ be a measure of (SI)-dependence. Then the associated rearranged dependence measure R_μ satisfies the following properties:*

1. $R_\mu(X,Y) \in [0,1]$.
2. $R_\mu(X,Y) = 0$ *if and only if X and Y are independent.*
3. $R_\mu(X,Y) = 1$ *if and only if Y is a measurable function of X.*
4. R_μ *is monotone with respect to $\preceq_m$, i.e. $C_{XY} \preceq_m C_{X'Y'}$ implies $R_\mu(X,Y) \le R_\mu(X',Y')$.*
5. R_μ *is an increasing function of $|r|$ for bivariate jointly normal distributed random vectors (X,Y) with correlation coefficient r. Moreover, if μ is strictly increasing in $|r|$, so is R_μ.*
6. R_μ *is continuous with respect to the metric D_p.*

In many situations, R_μ is actually a *strictly* increasing function of $|r|$ for Gaussian copulas. For instance, this is the case for R_p and R_κ since the underlying measures σ_p and κ already possess this property.

Proof. 1. Since $\Pi \le C^\uparrow \le C^+$ holds, we obtain 1.

2. Using that $C = \Pi$ if and only if $C^\uparrow = \Pi$, the assertion follows from the properties of μ.
3. Since C is completely dependent if and only if $C^\uparrow = C^+$, the result follows from the properties of μ.
4. Suppose $C_{XY} \preceq_m C_{X'Y'}$ holds, then Theorem 6.3.3 yields $C^\uparrow_{XY} \le C^\uparrow_{X'Y'}$. The assertion then follows from the monotonicity of μ with respect to $\le$.
5. We denote the Gaussian copula with correlation coefficient $r \in [-1,1]$ by C_r. The assertion then follows from $C_{r_1} \le C_{r_2}$ for $r_1 \le r_2$ and $C^\uparrow_r = C^\uparrow_{|r|} = C_{|r|}$. In case μ is strictly increasing in $|r|$, the assertion follows immediately from the definition.
6. $D_p(C_n, C) \to 0$ implies that $C^\uparrow_n$ converges pointwise towards $C^\uparrow$, from which the assertion follows.

□

6.4.1 Data processing inequality and self-equitability

We now turn towards the significance of the monotonicity of R_μ with respect to $\preceq_m$ and its connection to the so-called data processing inequality. Informally, the data processing inequality states that a (functional) modification of the input data cannot increase the information contained in the data. We refer to Cover and Thomas (2006) for an in-depth treatment of the data processing inequality in the context of information theory.

Proposition 6.4.8 (Data processing inequality). *Suppose X, Y and Z are continuous random variables such that Y and Z are conditionally independent given X. Then the data processing inequality*

$$R_\mu(Z, Y) \leq R_\mu(X, Y)$$

holds. In particular, $R_\mu(f(X), Y) \leq R_\mu(X, Y)$ holds for all measurable functions f.[6]

Proof. By assumption, Y and Z are independent given X, and by Theorem 3.1 of Darsow et al. (1992), $C_{ZY} = C_{ZX} * C_{XY}$ holds. Thus, Proposition 6.3.4 yields

$$C_{ZY} = C_{ZX} * C_{XY} \preceq_m C_{XY}$$

and the data processing inequality $R_\mu(C_{ZY}) \leq R_\mu(C_{XY})$ follows immediately from Theorem 6.4.7. Lastly, setting $Z = f(X)$ for a measurable function f, Y and $f(X)$ are conditionally independent given X and the second assertion follows. □

Following Proposition 2.1 of Geenens and Lafaye de Micheaux (2020), our proof of the data processing inequality also immediately yields an asymmetric version of the so-called self-equitability introduced in Kinney and Atwal (2014).

Corollary 6.4.9. *Suppose f is a measurable function such that X and Y are conditionally independent given $f(X)$. Then*

$$R_\mu(f(X), Y) = R_\mu(X, Y) \ .$$

Intuitively, self-equitability states that, for example, under the regression model $Y = f(X) + \varepsilon$ with independent continuous random variables X and ε, the dependence measure R_μ should depend on the strength of the signal-to-noise ratio of the model and not on the knowledge of the underlying function f. A similar idea is captured in Figures 3 and 4 of Junker et al. (2021). Corollary 6.4.9 also implies the invariance of R_μ under strictly monotone transformations of the random variables.

Proposition 6.4.10. *Suppose X and Y are continuous random variables, then*

$$R_\mu(f(X), g(Y)) = R_\mu(X, Y)$$

holds for any measurable bijective function f and any strictly monotone function g.

Proof. First, for any measurable bijection f, we have

$$R_\mu(f(X), g(Y)) = R_\mu(X, g(Y)) \ ,$$

using that X and Y are conditionally independent given $f(X)$ in combination with Corollary 6.4.9. If g is strictly increasing, the assertion follows immediately from the rank invariance of copulas. Otherwise, a straightforward calculation yields $C^{\uparrow}_{Xg(Y)} = \widehat{C^{\uparrow}}_{XY}$ for any strictly decreasing function g. Together with Property 4 of μ, $R_\mu(f(X), g(Y)) = R_\mu(X, Y)$ follows. □

[6]Note that for $R_\mu(f(X), Y)$ to be well-defined, $f(X)$ needs to be a continuous random variable.

6.4.2 Rearranged concordance measures

Let κ be a measure of concordance satisfying the nondegeneracy condition stated in Example 6.4.5. While κ measures the strength of the monotone relationship between two random variables X and Y, the corresponding rearranged concordance measure R_κ measures the strength of their (directed) functional relationship. Thus, intuitively, κ should always attain smaller values than R_κ, which is the consistency result given in the next theorem.

Theorem 6.4.11. *For all continuous random variables X and Y, we have*

$$|\kappa(X,Y)| \leq R_\kappa(X,Y) \ .$$

Proof. For $C := C_{XY}$, the assertion follows immediately from Corollary 6.1.9, due to

$$-\kappa(C^\uparrow) = \kappa(C^\downarrow) \leq \kappa(C) \leq \kappa(C^\uparrow) \implies |\kappa(C)| \leq \kappa(C^\uparrow) = R_\kappa(C) \ . \qquad \square$$

The next corollary treats the case of equality, i.e. $|\kappa(X,Y)| = R_\kappa(X,Y)$, which allows for a simpler calculation of the regression dependence measure R_κ.

Corollary 6.4.12. *If Y is stochastically monotone in X, then $R_\kappa(X,Y) = |\kappa(X,Y)|$.*

Proof. The assertion follows from the fact that Y is stochastically monotone in X if and only if C is stochastically monotone in the first component. $\square$

Note that the symmetry of κ does *not* imply $R_\kappa(X,Y) = |\kappa(X,Y)| = R_\kappa(Y,X)$ as X is, in general, not stochastically monotone in Y.

Many of the frequently used copula families are either stochastically increasing or stochastically decreasing, which simplifies the (possibly complex) calculation of $R_\kappa(C)$ to that of $\kappa(C)$. This property is similar to that of the Schweizer-Wolff measure σ_1, as observed in Nelsen (2006).

Example 6.4.13 (Gaussian copulas). *Suppose (X,Y) is a jointly normal distributed random vector with correlation coefficient $r \in [-1,1]$. Then Y is stochastically increasing in X if $r \geq 0$ and stochastically decreasing if $r \leq 0$. Together with the well-known formula for Spearman's ρ, we have*

$$R(X,Y) = R(Y,X) = \frac{6}{\pi} \arcsin\left(\frac{|r|}{2}\right) \ .$$

Thus, R is a strictly increasing function of $|r|$.

Example 6.4.14 (Extreme-value copulas). *The class of extreme value 2-copulas*

$$C^{EV}(u,v\,;\Lambda) = \exp\left(\log(uv)\left(1 - \widetilde{\Lambda}\left(\frac{\log(u)}{\log(uv)}\right)\right)\right)$$

given in Proposition 2.5.1 is stochastically increasing in both components (see Theorem 1 in Garralda Guillem (2000)). The rearranged Spearman's ρ for extreme value copulas is then given by

$$R(C) = \rho(C^\uparrow) = \rho(C) = 12 \int_0^1 \frac{1}{(2-\widetilde{\Lambda}(t))^2} \,\mathrm{d}t - 3 \ .$$

Most notably, although C is not necessarily symmetric, the degree of complete dependence between X and Y is identical, i.e.

$$R_\mu(X,Y) = R(C) = \rho(C) = \rho(C^\top) = R(C^\top) = R_\mu(Y,X) \ .$$

Example 6.4.15 (Archimedean copulas)**.** *Suppose C is an Archimedean 2-copula,*

$$C(u,v) = \phi^{[-1]}(\phi(u)+\phi(v)) \ ,$$

with Archimedean generator ϕ (see Definition 2.5.5). If the generalized inverse $\phi^{[-1]}$ is twice-differentiable, C is stochastically increasing in both components if and only if $\log\left(-\phi^{[-1]\prime}\right)$ is convex (see Proposition 3.3 in Capéraà and Genest (1993)). Then, we have for the rearranged Kendall's τ

$$R_\tau(C) = 4\int_0^1 \frac{\phi(s)}{\phi'(s)}\,\mathrm{d}s + 1 \ .$$

Moreover, the class of Archimax copulas, which combines the convex generator $\widetilde{\Lambda}$ and the Archimedean generator ϕ, is again stochastically increasing if and only if $\log\left(-\phi^{[-1]\prime}\right)$ is convex.

Example 6.4.16 (Shuffles of C^+)**.** *Let C_σ be a shuffle of C^+ as defined in Example 6.1.4. Since $C_\sigma^\uparrow = C^+$ holds for any shuffle C_σ of C^+ with some λ-preserving transformation σ, we obtain*

$$R(C_\sigma) = \rho(C^+) = 1 \ .$$

Example 6.4.17 (n-fold gluing of the tent copula)**.** *Suppose C_n is the n-fold gluing of the tent copula as presented in Example 6.1.5. Following the previous example, we have $R(C_n) = 1$ for all n. In contrast, using the computation of $(\partial_1 C_n^\top)^*$ from Equation (6.1) in Example 6.1.5, we obtain*

$$R(C_n^\top) = 12\int_{[0,1]^2} (C_n^\top)^\uparrow(u,v)\,\mathrm{d}\lambda(u,v) - 3 = 12\left(\frac{1}{4} + 2^{-(n+1)}\frac{1}{6}\right) - 3 = 2^{-n} \ .$$

Therefore, $R(C_n^\top) \to 0$ as $n \to \infty$.

The combination of $R(C_n) = 1$ and $R(C_n^\top) \to 0$ reflects the heuristic idea that the random variable Y is completely dependent on X for every n, whereas X becomes 'increasingly independent' of Y as n grows larger (see also Siburg and Stoimenov (2015) and Trutschnig (2011)).

Finally, we remark that the scope of Corollary 6.4.12 is not limited to parametric families. Many construction methods for copulas, such as ordinal sums (see Definition 5.3.1), convex combinations and the gluing construction (see Siburg and Stoimenov (2008b)), create stochastically increasing copulas if all input copulas are stochastically increasing.

6.5 Estimation of R_μ

If Y is not stochastically monotone in X, the measure $R_\mu(C)$ may be quite hard to calculate analytically. Using the induced checkerboard copulas $C_n^\#$ of C, however, $R_\mu(C)$ can be approximated by $R_\mu(C_n^\#)$ due to the uniform convergence of $C_n^{\#\uparrow}$ towards $C^\uparrow$. This motivates the construction of the estimator $\widehat{R}_\mu$ of $R_\mu(C)$.

The main ingredients for the construction of $\widehat{R}_\mu$ are the approximation result for checkerboard copulas in Theorem 6.2.3, and the convergence of the empirical checkerboard copulas established in Junker et al. (2021). For a sample $(x_1, y_1), \dots, (x_n, y_n)$ of a random vector (X, Y) with continuous univariate margins, the empirical checkerboard copula with bandwidth N is defined as

$$C_{N,n}^\#(\widehat{A}_n) := C_N^\#\big(C_n^\#(\widehat{A}_n)\big) \ ,$$

where $\widehat{A}_n = (\hat{a}_{ij})_{1 \le i,j \le n}$ is the $n \times n$ permutation matrix defined by

$$\hat{a}_{ij} := \begin{cases} 1 & \text{if there exists some } k \in \{1, \dots, n\} \text{ with } \operatorname{rank}(x_k) = i \text{ and } \operatorname{rank}(y_k) = j \\ 0 & \text{else} \end{cases} ,$$

where $\operatorname{rank}(x_k)$ denotes the position of x_k in the ordered sample $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$.

Theorem 6.5.1. *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from (X, Y), where (X, Y) is defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and has continuous univariate marginal distributions and copula C. For the empirical checkerboard copula $C_{N(n),n}^\#(\widehat{A}_n)$ with bandwidth $N(n) := \lfloor n^s \rfloor$ for $s \in \big(0, \frac{1}{2}\big)$, we then have that*

$$\widehat{R}_\mu := \mu(C_{N(n),n}^\#(\widehat{A}_n^\uparrow)) \to R_\mu(C)$$

holds $\mathbb{P}$-almost surely as $n \to \infty$.

Note that $C_{N(n),n}^\#(\widehat{A}_n^\uparrow)$ can be explicitly calculated via the algorithm described in Theorem 6.2.1.

Proof. By Theorem 3.12 in Junker et al. (2021), there exists a set $M \in \mathcal{A}$ with $\mathbb{P}(M) = 1$ such that $D_1(C_{N(n),n}^\#(\widehat{A}_n(\omega)), C)$ converges towards 0 for all $\omega \in M$. Note that $C_{N(n),n}^\#(\widehat{A}_n(\omega))$ is a genuine copula for all $\omega \in M$. Thus, an application of the continuity property given in Theorem 6.2.3 yields

$$0 \le D_1(C_{N(n),n}^\#(\widehat{A}_n^\uparrow(\omega)), C^\uparrow) \le D_1(C_{N(n),n}^\#(\widehat{A}_n(\omega)), C) \to 0 \ .$$

Since $C_{N(n),n}^\#(\widehat{A}_n^\uparrow(\omega))$ converges towards $C^\uparrow$ with respect to D_1 if and only if $C_{N(n),n}^\#(\widehat{A}_n^\uparrow(\omega))$ converges pointwise towards $C^\uparrow$,

$$\mu(C_{N(n),n}^\#(\widehat{A}_n^\uparrow)(\omega)) \to \mu(C^\uparrow) = R_\mu(C)$$

for all $\omega \in M$ now follows from Property 5 of Theorem 6.4.1. □

As many copula families are stochastically monotone, Corollary 6.4.12 suggests a simpler estimator for R_κ. Provided that the given data set displays stochastic monotonicity, which can be tested, for example, via Lee et al. (2009), the estimation of $R_\kappa(C)$ reduces to the well-known problem of estimating $|\kappa(C)|$ for which, in many cases, fast and reliable estimators already exist.

In the final subsection, we will practically explore the behaviour of the estimator $\widehat{R}_\mu$ and, if applicable, the behaviour of $|\hat{\kappa}|$ for the copula families discussed throughout this chapter. Our aim is to gain an overview over the statistical properties of $\widehat{R}_\mu$, while a more detailed investigation (considering, e.g., the influence of the bandwidth choice $N(n)$ or the behaviour with respect to various underlying (SI)-measures) will be addressed in future work.

6.5.1 Simulation study

In the subsequent simulation study, we choose Spearman's ρ as the underlying measure of (SI)-dependence since it induces the rearranged Schweizer-Wolff measure σ_1 as well as the rearranged Spearman's ρ. We generate samples of size $n \in \{50, 100, 500, 1000, 5000, 10000\}$ and calculate $\widehat{R}_\rho$ and additionally $|\hat{\rho}|$ whenever the underlying copula is stochastically monotone. The presented results are based on 1000 replications each.

All simulations have been conducted using the statistical software 'R' (see R Core Team (2021)). With the exception of the n-fold tent copula, the samples have been generated using the package 'copula' (see Hofert, Kojadinovic, Mächler and Yan (2020)). The package 'qad' (see Griessenberger, Junker, Petzel and Trutschnig (2021)) was used to calculate the doubly stochastic matrix $\widehat{A}_n$ of the the empirical checkerboard copula. Figures 6.3 to 6.7 have been created using 'ggplot2' (see Wickham (2016)).

Remark 6.5.2. *As discussed in Junker et al. (2021), the optimal choice of the bandwidth* $N(n) = \lfloor n^s \rfloor$ *with* $s \in \left(0, \frac{1}{2}\right)$ *for the empirical checkerboard copula is determined by the underlying (but unknown) copula* C*. Generally speaking, smaller values of* s *are more advantageous in case of an 'almost independent' copula* C*, whereas larger values of* s *improve the convergence rate of* $\widehat{R}_\mu$ *for a completely dependent copula* C*. Thus, motivated by Theorem 6.2.3, we propose a simple adaptive choice of* s *based on the threshold value* $|\hat{\rho}(C)|$*. In additional simulations not reported here, we found the following rule of thumb for the bandwidth choice to yield good results:*

$$s = \begin{cases} 0.5 & \text{if } |\hat{\rho}(C)| > \frac{2}{3} \\ 0.4 & \text{if } |\hat{\rho}(C)| \leq \frac{2}{3} \end{cases} .$$

Example 6.5.3 (Gaussian copulas)**.** *A jointly normal distributed random vector* (X, Y) *with correlation coefficient* $\mathrm{Corr}(X, Y) = r \in [-1, 1]$ *and copula* C_r *is stochastically monotone. Therefore, the rearranged Spearman's* ρ *equals*

$$R(C_r) = R(X, Y) = |\rho(X, Y)| = \frac{6}{\pi} \arcsin\left(\frac{|r|}{2}\right) \approx r \ .$$

The estimated values of $\widehat{R}_\rho$ *and* $|\hat{\rho}|$ *for the 1000 replications are presented as boxplots for* $r = 0.25$ *and* 0.75 *in Figure 6.3 and 6.4, respectively, next to a plot of a sample of size* $n = 1000$ *generated from the underlying copula. For ease of comparison, we use a similar visualization of the results as in Junker et al. (2021).*

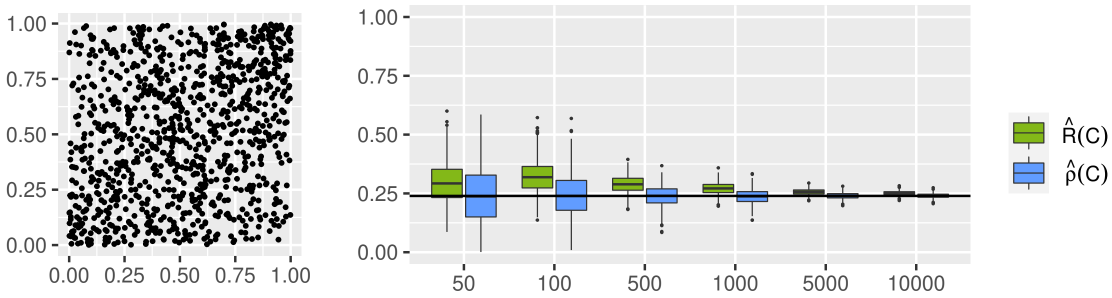


Figure 6.3: The left panel depicts a sample of 1000 datapoints generated from the Gaussian copula with $r = 0.25$, while the right panel depicts the boxplots of the 1000 estimates of $\widehat{R}_\rho(C_r)$ and $|\hat{\rho}(C_r)|$ in green and blue, respectively, for growing sample sizes n. The horizontal black line marks the true value of $R(C_r)$.

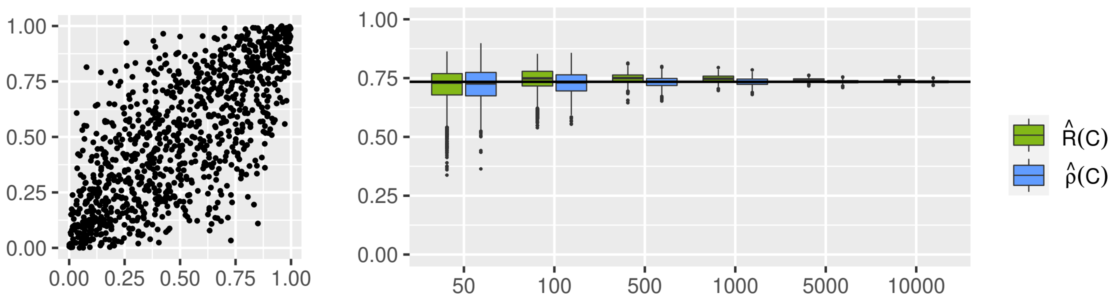


Figure 6.4: The left panel depicts a sample of 1000 datapoints generated from the Gaussian copula with $r = 0.75$, while the right panel depicts the boxplots of the 1000 estimates of $\widehat{R}_\rho(C_r)$ and $|\hat{\rho}(C_r)|$ in green and blue, respectively, for growing sample sizes n. The horizontal black line marks the true value of $R(C_r)$.

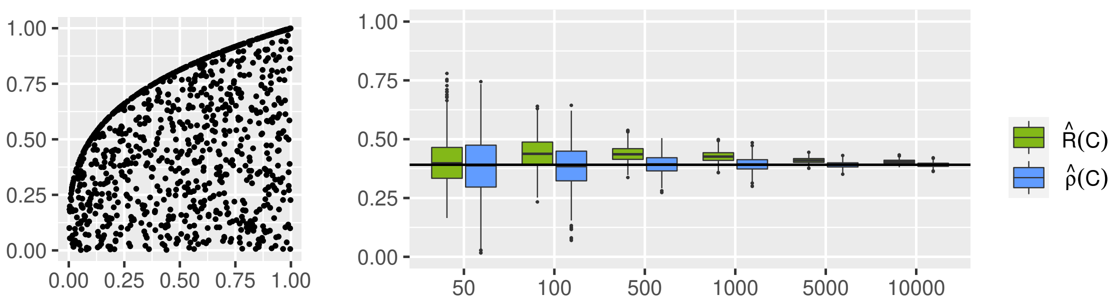


Figure 6.5: The left panel depicts a sample of 1000 datapoints generated from the Marshall-Olkin copula with $\alpha = 0.3$ and $\beta = 1$, while the right panel depicts the boxplots of the 1000 estimates of $\widehat{R}_\rho(C^{MO}_{\alpha,\beta})$ and $|\hat{\rho}(C^{MO}_{\alpha,\beta})|$ in green and blue, respectively, for growing sample sizes n. The horizontal black line marks the true value of $R(C^{MO}_{\alpha,\beta})$.

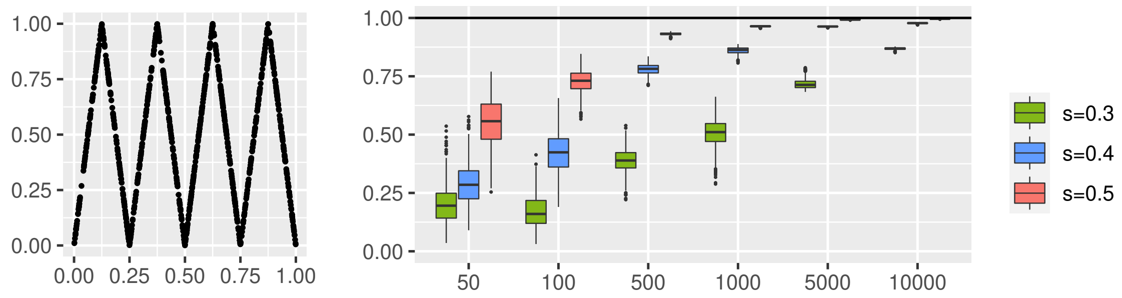


Figure 6.6: The left panel depicts a sample of 1000 datapoints generated from the 3-fold gluing of the tent copula C_3, while the right panel depicts the boxplots of the 1000 estimates of $\widehat{R}_\rho(C_3)$ with bandwidth $\lfloor n^s \rfloor$ for $s = 0.3, 0.4$ and 0.5 and for growing sample sizes n. The horizontal black line marks the true value of $R(C_3)$.

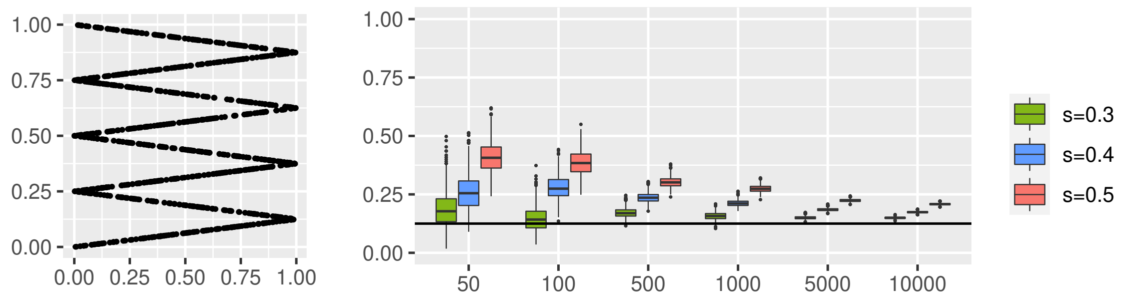


Figure 6.7: The left panel depicts a sample of 1000 datapoints generated from the transposed 3-fold gluing of the tent copula $C_3^\top$, while the right panel depicts the boxplots of the 1000 estimates of $\widehat{R}_\rho(C_3^\top)$ with bandwidth $\lfloor n^s \rfloor$ for $s = 0.3, 0.4$ and 0.5 and for growing sample sizes n. The horizontal black line marks the true value of $R(C_3^\top)$.

Example 6.5.4 (Marshall-Olkin copulas)**.** *In contrast to Gaussian copulas, the Marshall-Olkin copulas*

$$C^{MO}_{\alpha,\beta}(u,v) = \min\left\{u^{1-\alpha}v, uv^{1-\beta}\right\}$$

with $\alpha, \beta \in [0,1]$ *can exhibit asymmetric dependence and can contain a singular component. As* $C^{MO}_{\alpha,\beta}$ *is an extreme-value* 2*-copula, it is stochastically increasing for all choices of* α *and* β*. A short calculation (see Nelsen (2006)) yields*

$$R(C^{MO}_{\alpha,\beta}) = \rho(C^{MO}_{\alpha,\beta}) = \frac{3\alpha\beta}{2\alpha - \alpha\beta + 2\beta} \, .$$

The results for the Marshall-Olkin copula with $\alpha = 0.3$ *and* $\beta = 1$ *are given in Figure 6.5.*

Example 6.5.5 (Completely dependent copulas)**.** *The, in some sense, most pronounced asymmetric relationship between two random variables can be achieved by completely dependent copulas. We consider the n-fold gluing of the tent copula C_n from Examples 6.1.5 and 6.4.17. As seen in these examples, C_n exhibits a strongly asymmetric dependence structure with $R(C_n) = 1$ and $R(C_n^\top) \to 0$. Since neither C_n nor $C_n^\top$ are stochastically increasing, $|\hat{\rho}|$ does not yield an estimator for $R(C_n)$. This fact is underlined by the fact that $\rho(C_n) = 0$ for all $n \in \mathbb{N}$. Figures 6.6 and 6.7 therefore present the estimates of $\widehat{R}_\rho$ for the copulas C_3 and $C_3^\top$, respectively, and for the three bandwidth choices $s = 0.3, 0.4$ and 0.5.*

Bibliography

[1] A. Albanese and C. Sempi (2016). ‘Idempotent copulæ: Ordinal sums and Archimedean copulæ’. *J. Math. Anal. Appl.* **438**, pp. 1055–1065.

[2] E. Alvoni, F. Durante, P. L. Papini and C. Sempi (2007). ‘Different types of convexity and concavity for copulas’. In: *New dimensions in Fuzzy Logic and related Technologies - Proceedings of the 5th EUSFLAT Conference* (Ostrava, 11th–14th Sept. 2007). Ed. by M. Štěpnička, V. Novák and U. Bodenhofer. Vol. 1. Ostrava: University of Ostrava, pp. 185–189.

[3] J. Ansari (2019). ‘Ordering risk bounds in partially specified factor models’. Dissertation. University of Freiburg. DOI: 10.6094/UNIFR/149351.

[4] J. Ansari and L. Rüschendorf (2021). ‘Sklar’s theorem, copula products, and ordering results in factor models’. *Dependence Modeling* **9**, pp. 267–306.

[5] C. Bennett and R. C. Sharpley (1988). *Interpolation of Operators*. Boston: Academic Press.

[6] N. H. Bingham, C. M. Goldie and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

[7] A. Bücher, H. Dette and S. Volgushev (2012). ‘A test for Archimedeanity in bivariate copula models’. *J. Multivariate Anal.* **110**, pp. 121–132.

[8] A. Bücher, F. Irresberger and G. N. F. Weiss (2017). ‘Testing asymmetry in dependence with copula-coskewness’. *N. Am. Actuar. J.* **21**, pp. 267–280.

[9] P. Capéraà and C. Genest (1993). ‘Spearman’s ρ is larger than Kendall’s τ for positively dependent random variables’. *J. Nonparametr. Stat.* **2**, pp. 183–194.

[10] M. Carter and B. van Brunt (2000). *The Lebesgue-Stieltjes Integral*. New York: Springer.

[11] F. Chabi-Yo, S. Ruenzi and F. Weigert (2018). ‘Crash sensitivity and the cross section of expected stock returns’. *J. Financ. Quant. Anal.* **53**, pp. 1059–1100.

[12] A. Charpentier and J. Segers (2007). ‘Lower tail dependence for Archimedean copulas: Characterizations and pitfalls’. *Insurance Math. Econom.* **40**, pp. 525–532.

[13] A. Charpentier and J. Segers (2009). ‘Tails of multivariate Archimedean copulas’. *J. Multivariate Anal.* **100**, pp. 1521–1537.

[14] S. Chatterjee (2021). ‘A new coefficient of correlation’. *J. Amer. Statist. Assoc.* **116**, pp. 2009–2022.

[15] K. M. Chong (1974). ‘Some extensions of a theorem of Hardy, Littlewood and Pólya and their applications’. *Canad. J. Math.* **26**, pp. 1321–1340.

[16] K. M. Chong and N. M. Rice (1971). *Equimeasurable Rearrangements of Functions.* Kingston: Queen's University.

[17] P. Christoffersen, V. Errunza, K. Jacobs and H. Langlois (2012). 'Is the potential for international diversification disappearing? A dynamic copula approach'. *Rev. Financ. Stud.* **25**, pp. 3711–3751.

[18] T. M. Cover and J. A. Thomas (2006). *Elements of Information Theory.* 2nd edition. Hoboken: Wiley-Interscience.

[19] C. Czado (2019). *Analyzing Dependent Data with Vine Copulas.* Cham: Springer International Publishing.

[20] D. Dabrowska (1981). 'Regression-based orderings and measures of stochastic depence'. *Ser. Statistics* **12**, pp. 317–325.

[21] W. F. Darsow, B. Nguyen and E. T. Olsen (1992). 'Copulas and Markov processes'. *Illinois J. Math.* **36**, pp. 600–642.

[22] W. F. Darsow and E. T. Olsen (2010). 'Characterization of idempotent 2-copulas'. *Note Mat.* **30**, pp. 147–177.

[23] P. W. Day (1972). 'Rearrangement inequalities'. *Canad. J. Math.* **24**, pp. 930–943.

[24] P. W. Day (1973). 'Decreasing rearrangements and doubly stochastic operators'. *Trans. Amer. Math. Soc.* **178**, pp. 383–392.

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

[26] A. Dolati and A. D. Nezhad (2014). 'Some results on convexity and concavity of multivariate copulas'. *Iran. J. Math. Sci. Inform.* **9**, pp. 87–100.

[27] N. Dunford and B. J. Pettis (1940). 'Linear operations on summable functions'. *Trans. Amer. Math. Soc.* **47**, pp. 323–392.

[28] F. Durante, J. Fernández Sánchez and C. Sempi (2013). 'Multivariate patchwork copulas: A unified approach with applications to partial comonotonicity'. *Insurance Math. Econom.* **53**, pp. 897–905.

[29] F. Durante, E. P. Klement and J. J. Quesada-Molina (2008). 'Bounds for trivariate copulas with given bivariate marginals'. *J. Inequal. Appl.*, 161537.

[30] F. Durante and P. L. Papini (2009). 'Componentwise concave copulas and their asymmetry'. *Kybernetika* **45**, pp. 1003–1011.

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

[32] J. H. J. Einmahl, A. Krajina and J. Segers (2008). 'A method of moments estimator of tail dependence'. *Bernoulli* **14**.

[33] T. Eisner, B. Farkas, M. Haase and R. Nagel (2015). *Operator Theoretic Aspects of Ergodic Theory.* Cham: Springer International Publishing.

[34] P. Embrechts (2009). 'Copulas: A personal view'. *J. Risk Insur.* **76**, pp. 639–650.

[35] A. Feidt, C. Genest and J. Nešlehová (2010). 'Asymptotics of joint maxima for discontinuous random variables'. *Extremes* **13**, pp. 35–53.

[36] J. Fernández Sánchez, W. Trutschnig and M. Tschimpke (2021). 'Markov product invariance in classes of bivariate copulas characterized by univariate functions'. *J. Math. Anal. Appl.* **501**, 125184.

[37] S. Foss, V. Shneer, J. P. Thomas and T. Worrall (2018). 'Stochastic stability of monotone economies in regenerative environments'. *J. Econ. Theory* **173**, pp. 334–360.

[38] G. Frahm (2006). 'On the extremal dependence coefficient of multivariate distributions'. *Statist. Probab. Lett.* **76**, pp. 1470–1481.

[39] C. Fruciano, P. Colangelo, R. Castiglia and P. Franchini (2020). 'Does divergence from normal patterns of integration increase as chromosomal fusions increase in number? A test on a house mouse hybrid zone'. *Curr. Zool.* **66**, pp. 527–538.

[40] M. Fukushima, Y. Oshima and M. Takeda (2010). *Dirichlet Forms and Symmetric Markov Processes.* 2nd revised and extended edition. Berlin: De Gruyter.

[41] A. I. Garralda Guillem (2000). 'Structure de dépendance des lois de valeurs extrêmes bivariées'. *C. R. Acad. Sci. Paris* **330**, pp. 593–596.

[42] G. Geenens and P. Lafaye de Micheaux (2020). 'The Hellinger correlation'. *J. Amer. Statist. Assoc.*, pp. 1–15.

[43] C. Genest, J. Nešlehová and J. F. Quessy (2012). 'Tests of symmetry for bivariate copulas'. *Ann. Inst. Statist. Math.* **64**, pp. 811–834.

[44] F. Griessenberger, R. R. Junker, V. Petzel and W. Trutschnig (2021). *qad: Quantification of asymmetric dependence.* R package version 1.0.0. URL: `https://CRAN.R-project.org/package=qad`.

[45] G. Gudendorf and J. Segers (2010). 'Extreme-value copulas'. In: *Copula Theory and Its Applications* (Warsaw, 25th–26th Sept. 2009). Ed. by P. Jaworski, F. Durante, W. K. Härdle and T. Rychlik. Heidelberg: Springer, pp. 127–145.

[46] J. Harder (2013). 'Eine Präordnung für Randabhängigkeit von Copula-Funktionen'. Diploma thesis. TU Dortmund.

[47] G. H. Hardy, J. E. Littlewood and G. Pólya (1952). *Inequalities.* 2nd edition. Cambridge: Cambridge University Press.

[48] E. Hewitt (1960). 'Integration by parts for Stieltjes integrals'. *Am. Math. Mon.* **67**, pp. 419–423.

[49] W. Hoeffding (1940). 'Maßstabinvariante Korrelationstheorie'. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* **5**, pp. 181–233.

[50] M. Hofert, I. Kojadinovic, M. Mächler and J. Yan (2020). *copula: Multivariate dependence with copulas.* R package version 1.0-1. URL: `https://CRAN.R-project.org/package=copula`.

[51] L. Hua (2012). ‘Multivariate extremal dependence and risk measures’. PhD thesis. University of British Columbia.

[52] P. Jaworski (2006). ‘On uniform tail expansions of multivariate copulas and wide convergence of measures’. *Appl. Math.* **33**, pp. 159–184.

[53] P. Jaworski (2010). ‘Tail behaviour of copulas’. In: *Copula Theory and Its Applications* (Warsaw, 25th–26th Sept. 2009). Ed. by P. Jaworski, F. Durante, W. K. Härdle and T. Rychlik. Heidelberg: Springer, pp. 161–186.

[54] P. Jaworski (2015). ‘Univariate conditioning of vine copulas’. *J. Multivariate Anal.* **138**, pp. 89–103.

[55] H. Joe (1990). ‘Multivariate concordance’. *J. Multivariate Anal.* **35**, pp. 12–30.

[56] H. Joe (1997). *Multivariate Models and Multivariate Dependence Concepts.* London: Chapman and Hall.

[57] H. Joe (2015). *Dependence Modeling with Copulas.* Boca Raton: CRC Press.

[58] H. Joe, H. Li and A. K. Nikoloulopoulos (2010). ‘Tail dependence functions and vine copulas’. *J. Multivariate Anal.* **101**, pp. 252–270.

[59] R. R. Junker, F. Griessenberger and W. Trutschnig (2021). ‘Estimating scale-invariant directed dependence of bivariate distributions’. *Comput. Statist. Data Anal.* **153**, 107058.

[60] T. M. Kasper, S. Fuchs and W. Trutschnig (2021). ‘On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation’. *Bernoulli* **27**, pp. 2217–2240.

[61] J. B. Kinney and G. S. Atwal (2014). ‘Equitability, mutual information, and the maximal information coefficient’. *Proc. Natl. Acad. Sci. USA* **111**, pp. 3354–3359.

[62] H. König (2003). ‘The (sub/super)additivity assertion of Choquet’. *Studia Math.* **157**, pp. 171–197.

[63] S. Lee, O. Linton and Y. J. Whang (2009). ‘Testing for stochastic monotonicity’. *Econometrica* **77**, pp. 585–602.

[64] Haijun Li (2013). ‘Dependence comparison of multivariate extremes via stochastic tail orders’. In: *Stochastic Orders in Reliability and Risk.* Ed. by Haijun Li and X. Li. New York: Springer, pp. 363–387.

[65] Hui Li (2015). ‘On nonsymmetric nonparametric measures of dependence’. arXiv: `1502.03850 [stat.ME]`.

[66] X. Li, P. Mikusiński, H. Sherwood and M. D. Taylor (1997). ‘On approximation of copulas’. In: *Distributions with given Marginals and Moment Problems.* Ed. by Viktor Beneš and Josef Štěpán. Dordrecht: Springer, pp. 107–116.

[67] T. Mansour and M. Shattuck (2013). ‘A combinatorial approach to a general two-term recurrence’. *Discrete Appl. Math.* **161**, pp. 2084–2094.

[68] A. W. Marshall, I. Olkin and B. C. Arnold (2011). *Inequalities: Theory of Majorization and Its Applications.* 2nd edition. New York: Springer.

[69] A. J. McNeil (2008). 'Sampling nested Archimedean copulas'. *J. Stat. Comput. Simul.* **78**, pp. 567–581.

[70] A. J. McNeil and J. Nešlehová (2009). 'Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions'. *Ann. Statist.* **37**, pp. 3059–3097.

[71] R. E. Megginson (1998). *An Introduction to Banach Space Theory.* New York: Springer.

[72] R. Mesiar and C. Sempi (2010). 'Ordinal sums and idempotents of copulas'. *Aequationes Math.* **79**, pp. 39–52.

[73] T. Mikosch (2006). 'Copulas: Tales and facts'. *Extremes* **9**, pp. 3–20.

[74] R. B. Nelsen (2006). *An Introduction to Copulas.* 2nd edition. New York: Springer.

[75] E. Neuwirth (2001). 'Recursively defined combinatorial functions: Extending Galton's board'. *Discrete Math.* **239**, pp. 33–51.

[76] W. P. Novinger (1972). 'Mean convergence in L^p spaces'. *Proc. Amer. Math. Soc.* **34**, pp. 627–628.

[77] T. Okimoto (2008). 'New evidence of asymmetric dependence structures in international equity markets'. *J. Financ. Quant. Anal.* **43**, pp. 787–815.

[78] E. T. Olsen, W. F. Darsow and B. Nguyen (1996). 'Copulas and Markov operators'. In: *Distributions with Fixed Marginals and Related Topics.* Hayward: Institute of Mathematical Statistics, pp. 244–259.

[79] R Core Team (2021). *R: A Language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna.

[80] A. Rényi (1959). 'On measures of dependence'. *Acta Math. Acad. Sci. Hungar.* **10**, pp. 441–451.

[81] S. I. Resnick (1987). *Extreme Values, Regular Variation and Point Processes.* New York: Springer.

[82] R. T. Rockafellar (1997). *Convex Analysis.* Princeton: Princeton University Press.

[83] J. V. Ryff (1963). 'On the representation of doubly stochastic operators'. *Pacific J. Math.* **13**, pp. 1379–1386.

[84] J. V. Ryff (1965). 'Orbits of L^1-functions under doubly stochastic transformations'. *Trans. Amer. Math. Soc.* **117**, pp. 92–100.

[85] J. V. Ryff (1970). 'Measure preserving transformations and rearrangements'. *J. Math. Anal. Appl.* **31**, pp. 449–458.

[86] F. Salmon (2012). 'The formula that killed Wall Street'. *Significance* **9**, pp. 16–20.

[87] M. Scarsini (1984). 'On measures of concordance.' *Stochastica* **8**, pp. 201–218.

[88] F. Schmid and R. Schmidt (2007). 'Multivariate conditional versions of Spearman's rho and related measures of tail dependence'. *J. Multivariate Anal.* **98**, pp. 1123–1140.

[89] S. Scholtes (2012). *Introduction to Piecewise Differentiable Equations.* New York: Springer.

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

[91] M. Shaked and J. G. Shanthikumar (2007). *Stochastic Orders.* New York: Springer.

[92] K. F. Siburg and P. A. Stoimenov (2008a). 'A scalar product for copulas'. *J. Math. Anal. Appl.* **344**, pp. 429–439.

[93] K. F. Siburg and P. A. Stoimenov (2008b). 'Gluing copulas'. *Commun. Statist. Theory Methods* **37**, pp. 3124–3134.

[94] K. F. Siburg and P. A. Stoimenov (2010). 'A measure of mutual complete dependence'. *Metrika* **71**, pp. 239–251.

[95] K. F. Siburg and P. A. Stoimenov (2011). 'Symmetry of functions and exchangeability of random variables'. *Statist. Papers* **52**, pp. 1–15.

[96] K. F. Siburg and P. A. Stoimenov (2015). 'Almost opposite regression dependence in bivariate distributions'. *Statist. Papers* **56**, pp. 1033–1039.

[97] K. F. Siburg and C. Strothmann (2021a). 'A Markov product for tail dependence functions'. *J. Math. Anal. Appl.* **498**, 124942.

[98] K. F. Siburg and C. Strothmann (2021b). 'Stochastic monotonicity and the Markov product for copulas'. *J. Math. Anal. Appl.* **503**, 125348.

[99] M. Sibuya (1960). 'Bivariate extreme statistics, I'. *Ann. Inst. Stat. Math. Tokyo* **11**, pp. 195–210.

[100] N. L. Stokey, R. E. Lucas and E. C. Prescott (1989). *Recursive Methods in Economic Dynamics.* Cambridge: Harvard University Press.

[101] M. Trede and C. Savu (2013). 'Do stock returns have an Archimedean copula?' *J. Appl. Stat.* **40**, pp. 1764–1778.

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

[103] W. Trutschnig (2013). 'On Cesáro convergence of iterates of the star product of copulas'. *Statist. Probab. Lett.* **83**, pp. 357–365.

[104] W. Trutschnig and J. Fernández Sánchez (2015). 'Some consequences of the Markov kernel perspective of copulas'. In: *Innovations in Quantitative Risk Management* (TU München, 9th–13th Sept. 2013). Ed. by K. Glau, M. Scherer and R. Zagst. Cham: Springer International Publishing, pp. 393–409.

[105] M. Tsuji (1952). 'On F. Riesz' fundamental theorem on subharmonic functions'. *Tohoku Math. J.* **4**, pp. 131–140.

[106] H. Wickham (2016). *ggplot2: Elegant graphics for data analysis.* New York: Springer.