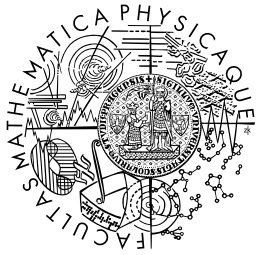

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER'S THESIS

Matvei Slavenko

Copula Based Models for Multivariate Data

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: doc. Ing. Marek Omelka, Ph.D.

Study programme: Probability, Mathematical Statistics and Econometrics

Prague 2024

I declare that I carried out this master thesis on my own, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

First and foremost, I thank Jakub Grulich, my partner (in crime) for always being there for me.

I would like to thank Heorhi Kolas for being an (almost) ideal room-mate over the years. Special thanks go to Iurii Galiaminskii; among other things, I appreciate the hours of co-working.

Thanks to Jack Colmer Gale for all his support.

I also express my feelings of gratitude to Kristina Mozyreva, my friends and my family for the trust they put in me and their continuous support.

Last but not least, my sincere gratitude goes to my supervisor, Marek Omelka. Thank you for your support, patience and guidance.

Title: Copula Based Models for Multivariate Data

Author: Matvei Slavenko

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Ing. Marek Omelka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis investigates probabilistic copula models for bivariate random vectors and their applications to dependency modelling and Scarsini measures of concordance. The first part focuses on bivariate distributions with non-atomic (continuous) marginals, establishing key principles of classic copula theory. It demonstrates that copula functions preserve all scale-invariant characteristics of random vectors, fully describing their dependency structures. Key concepts such as counter- and co-monotonicity, as well as concordance ordering, are introduced. The study also examines Scarsini measures of concordance, establishing that copulae accurately capture these measures. Particular instances, including the population versions of Kendall's tau and Spearman's rho, are explored in depth. The second part focuses on arbitrary bivariate distributions. Issues arising from the presence of atoms in the marginals are outlined; ways to address the non-uniqueness of associated copulae are discussed. We introduce the concept of the standard extension copula, which generalises the unique copula of the non-atomic case and validates both analytical and synthetic copula models for arbitrary margins. The behaviour of Kendall's tau a Spearman's rho is explored and various re-normalisations for these coefficients are considered. A number of examples illustrate the theoretical concepts.

Keywords: concordance ordering, dependency modelling, Kendall's tau, scale-invariant characteristics, Scarsini measures of concordance, Spearman's rho, standard extension copula

Contents

Introduction

Students of probability and statistics encounter the notion of *data dependencies* early in their studies, usually beginning with the concept of *independent random events*. Consider a probability space $(\Omega, \mathcal{A}, P)$. Recall that random events $A, B \in \mathcal{A}$ are called independent if and only if

$$\mathsf{P}[A \cap B] = \mathsf{P}[A]\ \mathsf{P}[B].$$

This foundational idea is then extended to the independence of random variables and, further, to the independence of σ-algebras. Regardless of the level of abstraction, the concept remains dichotomous: either independence holds, allowing (under some additional assumptions) for the application of numerous theorems and thus enabling us to tell quite a lot about the behaviour of the objects at hand; or it does not, and then we end up stripped of the majority of our analytical tools.

The introduction of *correlation coefficients or indices*, such as Pearson's r and Spearman's ρ, is the next step. Such indices to a certain extent quantify the *strength* as well as its *direction* of dependence between random variables X and Y. Although useful, these measures have their limitations. While providing certain insights into the relationships between random variables, they fail to capture the full complexity of their dependencies. For instance, it is well known that uncorrelated random variables being are not necessarily independent. Unless strict assumptions, such as bivariate normality, are imposed, correlation coefficients do not suffice to reconstruct the joint cumulative distribution function from the marginals.

In this thesis, we explore the so-called 'copulae' that allow us to capture the entire dependency structure of a random vector (X, Y). We demonstrate that copulae contain the essential information required to reconstruct the joint distribution function from its marginals. We then use copula theory to develop the theory of population rank-based correlation coefficients, also known as *measures of concordance*. The thesis is divided into four chapters and organised as follows.

In Part I, comprising the first two chapters, we focus on the special case of bivariate distributions with non-atomic (continuous) marginals. This is a standard and common setting for studying copulae used in the classic texts, such as [1] or [2].

In Chapter 1, we introduce copula models and outline their basic properties. In particular, Section 1.1 briefly sketches out the key topics explored and studied in this thesis. We introduce the concept of a copula associated with a random vector; we present Sklar's theorem showing that under the non-atomicity assumption, the associated copula is unique and we formally argue that it indeed captures the random vector's dependency structure in its entirety. We distinguish between the so-called *analytical* and *synthetic* copula models. The concept of counter- and co-monotonicity is introduced and its relation to copulae is pinpointed.

In Chapter 2, we apply copula theory in order to explore certain dependency concepts. We begin with introducing the *concordance partial ordering*. We then proceed by defining Scarcini's abstract measures of concordance; one general technique for constructing such measures is described. We then use it to introduce the population versions of Kendall's τ and Spearman's ρ. Certain more sophisticated

dependency concepts that are mentioned in literature are briefly introduced. On each step, we demonstrate that these concepts are fully captured and described by the associated copula.

Part II addresses arbitrary distributions. The entry-level and overviewing literature on copula models for arbitrary distributions is rather limited and these models are not considered in classic literature; hence it is especially valuable that the key facts are brought together.

Chapter 3 examines copula models for arbitrary distributions. The key differences with the case of non-atomic marginals, which was covered in Chap. 1, are highlighted and the challenges stemming from the presence of atoms in the marginal distributions. In particular, we demonstrate that the associated copula is no longer unique and outline the consequences to which ignoring this fact may lead. We also introduce the standard extension copula and argue that, to a great extent, it takes the role of the unique associated copula that we worked with under the non-atomicity assumption. Finally, the properties of synthetic copula models are summarised.

The first part of Chapter 4 mirrors Chap. 2, focusing on arbitrary distributions. We use the standard extension copula theory in order to re-establish and generalise the concepts of concordance partial order, concordance function and Scarsini abstract measures of concordance as well as Spearman's ρ and Kendall's τ. In the second part, we examine how the presence of atoms in the marginals affects the population Spearman's ρ and Kendall's τ. We study various possible renormalisations for ρ and τ and outline their key properties.

Part I

Distributions with Non-Atomic Marginals

1 Copulae for Distributions with Non-Atomic Marginals

Over the course of this chapter, we consider (X, Y) to be a random vector with the joint cumulative distribution function (CDF) denoted by H. Unless explicitly stated otherwise, the marginals X and Y are assumed to be non-atomic (in the sense of Definition 40) and their CDFs are denoted as F and G, respectively. For further details regarding the basic notation and conventions used in this work, please refer to Appendix A.

This chapter is organised as follows. In Section 1.1, we derive the notion of a copula naturally and with minimal effort using basic statistical concepts. We briefly show how they can be used to capture the dependency structure and how they are related to rank-based correlation coefficients. In Section 1.2, we formally define copulae as a family of joint distribution functions whose marginals are uniform on the interval $(0; 1)$ and explore its basic properties. In Section 1.3, we formally show that copulae indeed embody the multivariate structure of a random vector. We also introduce various notions of co- and counter-monotonicity and show how they are related to the so-called Fréchet-Hoeffding bounds copulae.

1.1 Setting the Stage

We begin by reviewing basic facts related to Pearson's and Spearman's correlation coefficients. Subsequently, we present an alternative representation of the sample Spearman's rank correlation coefficient, which we use to define its population counterpart. Examining the proposed definition, in a straightforward and natural manner, we derive a structure that will be referred to as a 'copula' . Additionally, in this section, we sketch out some key topics the thesis focuses on, namely copulae, correlation coefficients (also known as *measures of association or concordance*), as well as the relationships and connections between them. In this section, our intention is to present the key ideas, so the rigorous proofs are left to the subsequent part of the text.

1.1.1 Pearson's r and Spearman's ρ

We begin with defining population *Pearson's linear correlation coefficient* r, also known as *product-moment correlation coefficient* as well as its sample counterpart.

Definition 1 (Population Pearson's r)**.** Let $X, Y \in \mathcal{L}^2_+$. We define population Pearson's correlation coefficient as:

$$r(X, Y) := r(H_{X,Y}) := \frac{\mathsf{E}[(X - \mathsf{E}\,X)(Y - \mathsf{E}\,Y)]}{\sqrt{\operatorname{var} X \operatorname{var} Y}} = \frac{\mathsf{cov}(X, Y)}{\sqrt{\operatorname{var} X \operatorname{var} Y}}. \tag{1.1}$$

Definition 2 (Sample Pearson's r)**.** Let $\{(X_i, Y_i)\}_{i=1}^n$ be a random sample.

The sample Pearson's correlation coefficient is defined as follows:

$$\widehat{r}_n(X_i, Y_i) := \frac{\sum_{i=1}^n \left[(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)\right]}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sum_{i=1}^n (Y_i - \overline{Y}_n)^2}}. \tag{1.2}$$

Pearson's r is widely used as a measure of association between two random variables, yet it possesses certain limitations. The population correlation is defined only for $X, Y \in \mathcal{L}^2_+$. Additionally, its reliance on expected values renders the sample correlation coefficient sensitive to outliers in the data. Another unwilling property of Pearson's r is that it is not invariant under strictly increasing transformations of the marginals, and its possible values generally depend on the marginals (see, for instance, [3, p. 12, Ex. 2]).

One of the attempts to address the issues is grounded in using rank methods, which leads to *Spearman's rank correlation coefficient*. The idea behind the sample Spearman's ρ is simple: instead of working with the raw observations, we apply the known procedure to their ranks. In our case, we calculate sample Pearson's r from the ranks of the observations.

Definition 3 (Sample Spearman's ρ)**.**
Consider a random sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution (X, Y). Let R_i denote the rank (Def. 49) of X_i in the sequence $\{X_i\}_{i=1}^n$. Let S_j denote the rank of Y_j. We define the sample Spearman's ρ as follows:

$$\widehat{\rho}_n(X_i, Y_i) := \widehat{r}_n(R_i, S_i). \tag{1.3}$$

Notice that since X and Y are assumed to be non-atomic, with probability one there are no ties in the samples, and the ranks are defined uniquely.

While the definition of the sample Spearman's ρ can be readily found in existing literature, obtaining the definition of the *population* Spearman's ρ presents certain challenges. In the following we rewrite the sample ρ in a manner that provides insights into the potential definition of the population ρ.

We begin by employing Theorem 58, which allows us to express the sample ρ as follows:

$$\widehat{\rho}_n(X_i, Y_i) := \widehat{r}_n(R_i, S_i) = \widehat{r}_n\left(n \cdot \widehat{F}_n(X_i), n \cdot \widehat{G}_n(Y_i)\right). \tag{1.4}$$

In order to proceed, we require the subsequent theorem, which states that Pearson's r is invariant up to the sign under non-trivial affine transformations of random variables. We leverage the theorem to finish the calculations we started above, formalising them as Theorem 2.

Theorem 1 (Invariance of Pearson's r under affine transformations)**.**
Let $a, b \neq 0$, and $c, d \in \mathbb{R}$. Then it holds that

$$r(aX + c, bY + d) = \operatorname{sgn}(a) \cdot \operatorname{sgn}(b) \cdot r(X, Y). \tag{1.5}$$

Similarly, on the sample level it holds that

$$\widehat{r}_n(aX_i + c, bY_i + d) = \operatorname{sgn}(a) \cdot \operatorname{sgn}(b) \cdot \widehat{r}_n(X_i, Y_i). \tag{1.6}$$

Theorem 2 (Alternative representation of sample Spearman's ρ)**.**
Consider a random sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution (X, Y). Assume X and Y are non-atomic. Let $\widehat{F}_n$ and $\widehat{G}_n$ be the marginal empirical CDFs for the samples $\{X_i\}$ and $\{Y_i\}$, respectively. Then the following holds:

$$\widehat{\rho}_n(X_i, Y_i) = \widehat{r}_n\left(\widehat{F}_n(X_i), \widehat{G}_n(Y_i)\right). \tag{1.7}$$

Proof. We start with the right-hand side:

$$\widehat{r}_n\left(\widehat{F}_n(X_i), \widehat{G}_n(Y_i)\right) = \widehat{r}_n\left(\frac{1}{n}\cdot R_i, \frac{1}{n}\cdot S_i\right) = \\ = \operatorname{sgn}^2\left(\frac{1}{n}\right)\cdot \widehat{r}_n(R_i, S_i) = \widehat{r}_n(R_i, S_i) =: \widehat{\rho}_n(X_i, Y_i). \quad (1.8)$$

□

Theorem 2 states that sample Spearman's ρ is essentially the sample Spearman's r calculated from the observations, after being transformed by their empirical cumulative distribution functions (ECDFs). Given the consistency of the sample estimator $\widehat{r}_n$ in estimating the population value r, along with the uniform consistency of the ECDF as an estimator of the true cumulative distribution function (Theorem 59), it is reasonable to hope that the population Spearman's ρ could be defined by simply 'taking off the hats' in the alternative representation of the sample ρ. Based on this rationale, the following definition is proposed.

Definition 4 (Population Spearman's ρ)**.** We define population Spearman's rank correlation coefficient as follows:

$$\rho(X,Y) := \rho(H_{X,Y}) := r\left(F(X), G(Y)\right) \\ = \frac{\mathsf{E}[(F(X) - \mathsf{E}\,F(X))(G(Y) - \mathsf{E}\,G(Y))]}{\sqrt{\operatorname{var} F(X)\operatorname{var} G(Y)}}. \quad (1.9)$$

For proper reasoning why Definition 4 is valid and a deeper scrutiny of its properties, please refer to Section 2.4.

The value $F_X(x_i)$, where x_i is an observation of a random variable X, is called *observations's grade* and is the population counter-part of its rank. Definition 4 hence explains why Spearman's ρ is sometimes called *grade correlation coefficient*, see [1, pp. 169–170].

1.1.2 Deriving a copula

Let us examine the population Spearman's ρ closer. Notice that expression (1.9) could be simplified employing the so-called probability integral transform (Theorem 56). Exploiting the fact that $F_X(X)$, $G_Y(Y) \sim Uni(0;1)$, we obtain the result formalised as the following theorem.

Theorem 3 (Alternative representation of population Spearman's ρ)**.**
Denote $U := F(X)$ and $V := G(Y)$. Let $C_{U,V}$ be the joint CDF of the random vector (U,V). The following holds:

$$\rho(X,Y) = 12\,\mathsf{E}[F(X)\,G(Y)] - 3 \quad (1.10)$$

$$= 12\int_{\mathbb{R}^2} uv\,dC_{U,V}(u,v) - 3 \quad (1.11)$$

$$= 12\,\mathsf{E}[UV] - 3. \quad (1.12)$$

Theorem 3, specifically equation (1.11) is the touchstone of the whole section. It demonstrates, that population Spearman's ρ could be expressed solely as a

functional of $C_{U,V}$. It means, that the value of ρ does not depend on the marginal distributions; specifically it is invariant under strictly increasing transformations.

In fact, the $C_{U,V}$ object we have just obtained is a so-called *copula* and the whole thesis is dedicated to such objects. The following theorem formulates and formalises several key properties of what copulas, at least for the case of non-atomic marginals.

Theorem 4 (Key copula properties)**.**
Let (X, Y) be a random vector with the joint CDF H. Let F and G be the marginal distribution functions of X and Y, respectively. Let X and Y be non-atomic. Denote $U := F(X)$ and $V := G(Y)$. Let $C_{U,V}$ be the joint CDF of the random vector (U, V). Then the following holds:

- *Univariate marginals property:* $U, V \sim Uni(0; 1)$. (1.13)
- *Closed form expression:* $C_{U,V}(u, v) = H\left(F^{-1}(u), G^{-1}(v)\right)$. (1.14)
- *Coupling property:* $H(x, y) = C_{U,V}\left(F(x), G(y)\right)$. (1.15)

Let us elaborate on the meaning of each point in Theorem 4. Equation (1.13) stresses that $C_{U,V}$ has uniform marginals; this property is often used to define copulae axiomatically (see Definition 5 and the corresponding remarks). Section 1.2 is dedicated to exploring that way of thinking about copulae.

Equation (1.14) once again highlights that (in case of the non-atomic marginals) the object $C_{U,V}$ is derived in a closed form from the joint CDF $H_{X,Y}$ and thus could be treated as a certain trait or property of random vector's joint distribution; i.e., we may write $C_{X,Y}$, Theorem 9 justifies that notation.

Equation (1.15) is known as *Sklar's representation* (see [4, p. 2]). It explicates the mechanism how $C_{U,V}$ encompasses the dependence structure of $H_{X,Y}$. It is well known the joint distribution defines the marginals, yet it does not work the other way around: the marginals generally do not allow to reconstruct the joint distribution. In this context, $C_{U,V}$ is exactly the element that binds (or 'couples'—hence *copula*) the marginal distribution functions to produce the joint CDF. Hence we can perceive $C_{U,V}$ as the embodiment of a particular way of how X and Y depend on each other within the (X, Y) random vector (as a 'multivariate' or 'dependence' structure, see [5, Sec. 1.6, first paragraph]). Section 1.3 focuses on that side of copulae.

1.2 Copulae as Distribution Functions

1.2.1 Defining copulae

Definition 5 (2-copula)**.** Consider a 2-CDF C. Let (U, V) be a random vector such that its joint distribution function is given by C. It is said that C is a *2-copula* (or simply *copula*) if and only if it holds that $U, V \sim Uni(0; 1)$.

Notice that our interest in copulae stems from the Sklar's representation (Equation (1.15)): we are interested in studying functions with the coupling property. For a distribution function F, it holds that $\operatorname{Ran} F \subseteq [0; 1]$. Hence we

are primarily interested in the unit square $[0;1]^2$, while the functions' behaviour outside the unit square is a nuisance and merely a question of standardisation.

The standardisation 'a 2-CDF with uniform marginals' that we chose in Definition 5 is but one of many possible. Our approach follows the definition in [5, p. 12, last paragraph] and is one of the most common ways to introduce copulae. Another source using the same approach is [4, first sentence].

Another approach is taken by Nelsen; the main and only difference is the *domain*. In [1, p. 10, Definitions 2.2.1–2.2.2], he defines copulae as functions with the domain being equal to the unit square $[0;1]^2$ fulfilling certain additional properties. The same approach is employed in [6, p. 347].

Definition 6 (Nelsen copula)**.** A function $C : [0;1]^2 \to \mathbb{R}$ is called a (Nelsen) copula if and only if the following holds:

1. $\forall u, v \in [0;1] : C(0,v) = 0 = C(u,0)$.
2. $\forall\ u_1 < u_2,\ v_1 < v_2 \in [0;1] : C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0$.
3. $\forall u, v \in [0;1] : C(u,1) = 1; C(1,v) = 1$.

Although strictly speaking Definitions 5 and 6 are not equivalent, it bears no consequences for our purposes for the differences are merely technical. To clarify this point, we first introduce the so-called '0-1 collar operator'.

Definition 7 (0-1 collar operator)**.** Let $x \in \mathbb{R}$. Define $\check{x}$ as follows:

$$\check{x} := \begin{cases} 0, & x < 0, \\ x, & x \in [0;1], \\ 1, & x > 1. \end{cases} \tag{1.16}$$

Notice that for a random variable $U \sim Uni(0;1)$ it holds that $F_U(u) = \check{u}$.

Now, let C be a 2-copula in the sense of Definition 5. Then a restriction $C|_{[0;1]^2}$ is a Nelsen copula. Conversely, consider a Nelsen copula C'. The extension $C(x,y) := C'(\check{x},\check{y})$ is then a 2-copula. As we can see, the differences are merely technical. In the rest of the text we shall rely solely on Definition 5.

1.2.2 Properties of copulae as a family of functions

A nice property of copulae that is useful in theoretical considerations is the fact that copulae are uniformly continuous.

Theorem 5. *Let C be a copula. Then C is uniformly continuous.*

Proof. See [1, p. 1, Theorem 2.2.4] □

It is also worth observing that convex combinations of copulae are again copulae (see [1, p. 14, Ex. 2.3]).

Theorem 6. *Let B and C be copulae. Then for $\alpha \in (0;1)$, the convex combination $\alpha\, B + (1-\alpha)\, C$ defined pointwisely is a copula, too.*

All cumulative distribution functions are trivially bounded by 0 from below and by 1 from above (see Defs. 44 and 47). Since 2-copulae are CDFs by definition, the same boundaries trivially hold for them as well. An important fact about 2-copulae is that the boundaries for them can be refined. Definition 8 introduces functions that work as the bounds for 2-copulae, Theorems 7 and 8 then state that the functions actually bound 2-copulae and that these boundaries are attainable.

Definition 8 (Upper and lower Fréchet-Hoeffding bounds)**.**
Consider functions $M, W : \mathbb{R}^2 \to \mathbb{R}$, defined as

1. $M(u, v) := \min(\check{u}, \check{v})$;
2. $W(u, v) := \max(\check{u} + \check{v} - 1, 0)$.

We call M and W the *upper and lower FH-bounds*, respectively.

Definition 8 corresponds to the definition from [1, p. 11] while accommodating it to our definition of copulae. When considering $u, v \in [0; 1]^2$, the collar operator may be disregarded and we obtain exactly Nelsen's definition. The collar operator is then required to extend the function to the whole $\mathbb{R}^2$ consistently with Definition 5.

Theorem 7 (Fréchet-Hoeffding)**.** *Let C be a 2-copula. Then the following holds:*

$$W(u, v) \leq C(u, v) \leq M(u, v).$$

Proof. For proof, see [1, p. 10, Theorem 2.2.3] or [2, p. 27, Th. 1.7.3]. □

Theorem 8 (Upper and lower FH-bounds are copulae)**.**
Both M and W from Definition 8 are 2-copulae.

Proof. See [2, p. 11, Exs. 1.3.3 and 1.3.5]. □

Notice that both M and W induce distributions singular with respect to Lebesgue measure. The upper FH bound corresponds to a random vector (U, U), while the lower FH-bound is a 2-CDF of the random vector $(U, 1 - U)$; in both cases $U \sim Uni(0; 1)$. For both random vectors, the mass is concentrated in one of the unit square's diagonals.

Another example of copulae that is worth mentioning is the product copula.

Definition 9 (Product copula)**.** Define a mapping $\Pi : \mathbb{R}^2 \to [0; 1]$ so that $\Pi(u, v) := \check{u}\check{v}$. Π is then the *product copula*.

Observe that the product copula Π is the 2-CDF corresponding to a random vector $(U, V) \sim Uni(0; 1)^2$.

1.3 Copulae as Dependence Structures

1.3.1 Sklar's theorem

As we have already stated, our interest in copulae is driven by the Sklar's representation (Equation (1.15)) and our consequent intent to interpret copulae as (marginal-free) embodiments of dependence structure. This interest is two-fold.

On the one hand, one might want to extract the copula from a random vector or its 2-CDF in order to analyse solely the dependence structure; we may perceive this is an *analytic model*. It is then convenient to introduce the definition of copulae *associated* or *compatible* with a random vector and its 2-CDF (see [7, p. 477, Definition 1.A]).

Definition 10 (Copulae associated with a joint CDF)**.**
We say that a 2-copula C is *associated* or *compatible* with an arbitrary random vector (X, Y) and its 2-CDF H if and only if it holds that

$$H(x, y) = C\left(F(x), G(y)\right). \tag{1.17}$$

As we will see in this section, it turns out that for any random vector there exists at least one compatible copula. Moreover, if the marginals are non-atomic, the associated copula is unique; hence we can denote it as $C_{X,Y}$ or C_H.

On the other hand, an analyst might be interested in finding a joint distribution with the pre-specified marginals and, at the same time, having some desired dependency properties (which could be perceived as a *synthetic model*). A laboratory example of such a task is, for instance, a problem often offered to university students: finding a joint distribution such that the marginals are normal, yet the joint distribution itself is not bivariate normal. In real-world applications, the tasks might be similar in spirit, yet more complex in details.

The Sklar's representation (Eq. (1.15)) seems to be a convenient tool for solving problems like that: one may choose the marginals and then specify the dependency structure separately via selecting a suitable copula. It turns out that this approach is valid: substituting arbitrary marginal distribution functions and copulae to the Sklar's representation, we always obtain a valid 2-CDF.

These facts together are known as Sklar's theorem (see, for instance, [1, p. 18, Theorem 2.3.3]).

Theorem 9 (Sklar)**.** *Consider an arbitrary random vector (X, Y) with the joint distribution function $H_{X,Y}$. Let F_X and G_Y be the marginal distribution functions. Then the following holds:*

1. *There exists at least one copula C associated with (X, Y).*
2. *If X and Y are non-atomic, then the associated copula C is unique and thus can be denoted as $C_{X,Y}$ or C_H.*
3. *Conversely, if C is a copula and F and G are univariate CDFs, then $C(F(x), G(y))$ defines a valid 2-CDF.*

Proof. See Theorem 24. □

A trivial yet important observation is that copulae are invariant under strictly increasing transformations. As Durante and Sempi [2, p. 58, Rem. 2.4.2] point out, this corresponds to the fact, that the change of physical units, conversion to other currencies or considering a logarithm of the studied values (if it makes sense) do not affect the dependency structure. This also suggests that copulae and rank-based procedures are closely interrelated.

Theorem 10 (Copula invariance under increasing transformations)**.**
Consider a random vector (X, Y) and its associated copula C. Let f, g be strictly increasing transformations.
Then C is also compatible with $(f(X), g(Y))$.

Proof. The proof is straight-forward, see [1, p. 25, Theorem 2.4.3] or [2, pp. 57–60]. □

1.3.2 FH-bound and co-monotonicity

We are now going to explore the dependency structures induced by FH-bounds. In a nutshell, FH-bounds correspond to random variables almost surely being in a strictly monotone (increasing or decreasing) dependence.

To clarify that, we introduce the concept of co-monotonicity. One of the earlier attempts to introduce the notion was made in 1987 by Menahem Yaari [8, p. 103].

Definition 11 (Yaari co-monotonicity)**.** Consider random variables X and Y defined on the probability space $(\Omega, \mathcal{A}, P)$. We say, that X and Y are co-monotonic if and only if the following holds:

$$\forall \omega, \omega' \in \Omega : (X(\omega) - X(\omega'))\,(Y(\omega) - Y(\omega')) \geq 0. \tag{1.18}$$

Let us elaborate on the meaning of that definition. The definition may be rewritten as follows:

$$\forall \omega, \omega' \in \Omega : [X(\omega) \geq X(\omega') \wedge Y(\omega) \geq Y(\omega')] \vee [X(\omega) \leq X(\omega') \wedge Y(\omega) \leq Y(\omega')]\,. \tag{1.19}$$

If we reformulate it in terms of mappings from $\mathbb{R}$ to $\mathbb{R}$, we will obtain something like that. Let $f, g : \mathbb{R} \to \mathbb{R}$. Then f and g are co-monotonic if and only if the following holds:

$$\forall a, b \in \mathbb{R}, a < b : [f(a) \geq f(b) \wedge g(a) \geq g(b)] \vee [f(a) \leq f(b) \wedge g(a) \leq g(b)]\,. \tag{1.20}$$

We utilised the fact that $\mathbb{R}$ is a well-ordered set; without loss of generality we put $a < b$ in the expression above. The fact that f and g are Yaari co-monotonic is then equivalent to both f and g having the same type of monotonicity (non-increasing or non-decreasing) on each given interval; it thus generalises the definition of *concordant* functions introduced by Lehmann [9, p. 1138]. Since Ω is usually not well-ordered, the more general approach reflected in Definition 11 is required.

Yaari ([8, p. 104]) introduces the definition as a part of his economic theory of risk, ascribing to it the economic interpretation of 'bets on the same event': *'It is, in fact, an analogue of perfect correlation for this distribution-free setting. When two random variables are co-monotonic, then it can be said that neither of them is a hedge against the other.'*

Notice that Definition 11 is, as Yaari says, 'distribution-free' in the sense that it does not rely in any way on the measure P; it is therefore restrictive and clumsy in our current approach to the probability theory and statistics. We hence will explore other ways of thinking about co-monotonicity.

Definition 12 (Weak co-monotonicity)**.**
Consider an arbitrary random vector (X, Y). We say that X and Y are *weakly co-monotonic* if and only if there exists a random variable $U \sim Uni(0; 1)$ such that $X = f(U)$ and $Y = g(U)$ almost surely, where f and g are functions non-decreasing on $(0; 1)$.

Definition 13 (Strong co-monotonicity)**.**
Consider an arbitrary random vector (X, Y). Let f be a transformation strictly increasing on $\operatorname{Ran} Y$. We say that X and Y are *strongly co-monotonic* if and only if $X = f(Y)$ almost surely.

Both ways to define co-monotonicity can be found, for instance, in [4, p. 5]. It is easy to observe that for X and Y non-atomic the definitions are equivalent.

Now we are in position to interpret the meaning the upper FH-bound in terms of the dependency structure it imposes. The proof for the case of non-atomic marginals is more or less obvious.

Theorem 11 (Upper FH-bound induces co-monotonicity)**.** *Consider a random vector (X, Y) with non-atomic marginals. Then the following statements are equivalent:*

1. *X and Y are weakly co-monotonic;*
2. *(X, Y) is strongly co-monotonic;*
3. *The upper FH-bound M is compatible with (X, Y).*

Proof. Follows from Theorems 10 and 56. □

In the same manner we may construct the concept of random variables being *counter-monotonic*; it then turns out that counter-monotonicity is induced by the lower FH-bound W. The details are however out of the scope of this text.

Another way to characterise the dependency structure induced by M and W is based on examining the support of the random vector (X, Y). Details could be found in [1, pp. 30–32].

1.3.3 Independence and product copula

Another particularly interesting instance of multivariate structures is *independence*, which is induced by the product copula.

Theorem 12. *Let Π be a copula associated with the random vector (X, Y). Then X and Y are independent.*

2 Measures of Concordance for Distributions with Non-Atomic Marginals

Over the course of this chapter, we consider (X, Y) to be a random vector with the joint cumulative distribution function (CDF) denoted by H. Unless explicitly stated otherwise, the marginals X and Y are assumed to be non-atomic (in the sense of Definition 40) and their CDFs are denoted as F and G, respectively. For further details regarding the basic notation and conventions used in this work, please refer to Appendix A.

2.1 Concordance Order

2.1.1 Intuition behind concordance

In the previous chapter, we studied the margin-free dependence structure of a random vector embodied in the unique associated copula, and we focused on the multivariate structure as a whole. In certain cases we were able to interpret the structure in a meaningful way; e.g., the product copula corresponds to the random variables being independent. Instead of describing the whole dependency structure we might be interested in some narrower dependency properties. For instance, we may focus on the *concordance* or, as it is often called nowadays, *correlation* of the random vector's elements.

As Scarsini [10, p. 201] puts it, the intuitive notion of concordance or correlation between the random variables X and Y may be perceived as the tendency of large values of X to go along with large values of Y. Following this idea, perfectly concordant (discordant) random variables almost surely are strictly increasing (decreasing) transformations of each other. In other words, perfectly concordant (discordant) random variables are strongly co-monotonic (counter-monotonic). We may thus perceive the notion of concordance as some generalisation of co- and counter-monotonicity to a spectrum.

Pearson's r discussed in Section 1.1 is one way to gauge the concordance, yet as it was shown, among its other disadvantages, it is not invariant under strictly increasing transformations, so generally X and Y being co-monotonic does not imply that $r(X, Y) = 1$. We thus need to develop other tools to capture the idea of concordance that will respect the two poles: counter- and co-monotonicity. One way to do that was suggested by Marco Scarsini in 1984 (see [10]).

Consider a fixed point $(a, b) \in \mathbb{R}^2$ dividing the $\mathbb{R}^2$ plain to four quadrants:

$$\begin{aligned} Q_1(a,b) &:= \{x, y \in \mathbb{R} : x > a, y > b\}; \\ Q_2(a,b) &:= \{x, y \in \mathbb{R} : x \leq a, y > b\}; \\ Q_3(a,b) &:= \{x, y \in \mathbb{R} : x \leq a, y \leq b\}; \\ Q_4(a,b) &:= \{x, y \in \mathbb{R} : x > a, y \leq b\}. \end{aligned} \tag{2.1}$$

If we perceive *large* values as those being larger than a on the horizontal axis and those larger than b for the vertical axis, we obtain the following picture.

In Q_1 (Q_3), large (small) values of x go with large (small) values of y. Conversely, the quadrant Q_2 (Q_4) contains the small (large) values of x paired with large (small) values of y. The odd quadrants contain the values that are *concordant*, while the even quadrants contain values that are *discordant* with respect to the reference point (a, b).

2.1.2 Concordance partial order

Using this partitioning, we may compare the concordance of two random vectors, assuming they share the same marginals. Before we proceed, it is useful to introduce the concept of *Fréchet classes*. Although it is a broad notion (see, for instance, [11, p. 1]), in our work the following simplified version that was used by Scarsini [10, p. 202] will suffice:

Definition 14 (Fréchet class)**.** Let F and G be univariate CDFs. Let $\Gamma(F, G)$ be the set of all 2-CDFs whose marginals equal to F and G, respectively. The set is then called the *Fréchet class defined by or corresponding to the marginals F and G*.

In other words, two-variate distributions are said to be in the same *Fréchet class* if their corresponding marginal distributions coincide.

Remark 1. In line with Remark 11, we shall also write:

- $(X, Y) \in \Gamma(F, G)$, meaning that the joint CDF and the distribution of the random vector (X, Y) have the marginals F and G.
- $H^* \in \Gamma(X, Y)$, meaning that the random vector (X, Y) and the 2-CDF H^* share the same marginals.
- $(X^*, Y^*) \in \Gamma(X, Y)$, meaning $X^* \sim X$ and $Y^* \sim Y$.
- etc.

Although this is not entirely correct from the technical point and may be qualified as notation abuse, it is convenient, accurate and transparent enough for our purposes.

Scarsini [10, pp. 202–203, Definition 1] introduces the following definition:

Definition 15 (Concordance partial order on a given Fréchet class)**.**
Consider two random vectors (X, Y) and (X^*, Y^*) belonging to the same Fréchet class. We say that (X, Y) is more concordant than (X^*, Y^*) if and only if it holds that:

$$\forall a, b \in \mathbb{R} : \mathsf{P}\left[(X, Y) \in Q_1(a, b) \cup Q_3(a, b)\right] \geq \mathsf{P}\left[(X^*, Y^*) \in Q_1(a, b) \cup Q_3(a, b)\right]. \quad (2.2)$$

We use the following notation to reflect that fact:

$$H \succeq H^*, \qquad (X, Y) \succeq (X^*, Y^*). \quad (2.3)$$

Remark 2. Notice that this definition works for arbitrary Fréchet classes. Specifically, it does not require the margins to be non-atomic. More on that in Section 4.1.

It is easy to observe that the definition could be equivalently reformulated in terms of the joint 2-CDFs.

Theorem 13 (Partial concordance order characterisation)**.**
Consider an arbitrary 2-CDF H. Let $H^* \in \Gamma(H)$. The following statements are then equivalent:

1. $H \succeq H^*$;
2. $\forall a, b \in \mathbb{R} : H(a, b) \geq H^*(a, b)$.

Notice that this establishes a partial order on the set of bivariate distributions sharing the same marginals or, in other words, on the Fréchet class $\Gamma(X, Y)$.

The reason why we (for now) require the compared vectors to share the marginal distributions is the following. Assume

$$(X, Y) \sim Uni([1; 2]^2) \quad \text{and} \quad (X^*, Y^*) \sim Uni([0; 1]^2). \tag{2.4}$$

Then clearly $H \geq H^*$ pointwisely, yet intuitively there seems to be no reason to claim that H is more concordant than H^*. There are two ways to resolve this problem.

The first approach is to limit ourselves to comparing only two-variate distributions belonging to the same Fréchet class. This approach is powerful enough to satisfy one's needs in many cases. If, say, an analyst has pre-specified marginals and selects the proper dependence structure to produce the joint distribution, then they naturally stick to comparing various random vectors sharing the very same pre-specified marginals that were chosen at the beginning; there seems to be no need to seek something more general: see, e.g., [12, p. 546].

Another approach is to demand that the strictly increasing transformations do not affect the order. Indeed, in the example above $(X - 1, Y - 1) \sim (X^*, Y^*)$, so these two vectors should be equivalent in terms of concordance. In light of the fact that it is exactly the associated copulae that are invariant under such transformations (Theorem 10), this naturally leads us to using copulae to generalise the concordance partial order to any bivariate distribution with non-atomic marginals, at least in our current setting of the marginals being non-atomic. Scarsini [10, p. 205, Def. 4] then proposes the following definition:

Definition 16 (Concordance order for distribution with non-atomic margins)**.**
Consider two random vectors (X, Y) and (X^*, Y^*) with non-atomic marginals. Let C and C^*, respectively, be their unique associated copulae.
We then say that (X, Y) is *more concordant* than (X^*, Y^*) if and only if the following holds:

$$\forall u, v \in \mathbb{R}^2 : C(u, v) \geq C^*(u, v). \tag{2.5}$$

Observe that the partial order introduced in Definition 16 indeed extends the order introduced in Definition 15 which relies on the fact that the compared distributions share the same marginals. We thus will use the same notation $H \succeq H^*$ to denote it.

Some people refer to the concordance partial order that we have just introduced as *positive quadrant dependence ordering* (see [6, p. 350] and [7, pp. 497–498]), while in other sources it is called *correlation order* (see [13, p. 203, Def. 2 and p. 204, Th. 1]). Nelsen also discusses concordance order, see [1, pp. 38–39].

The concordance partial ordering is indeed only a *partial* order on the set of bivariate distributions with non-atomic marginals. For instance, it is easy to see that the product copula Π and the copula produced as an average of FH-bounds $\frac{W+M}{2}$ are not comparable.

Another important observation is that the concordance order extends the order suggested by the Fréchet-Hoeffding theorem (Th. 7). Indeed, the theorem implies that the upper FH-bound M is the largest and maximal element under the concordance ordering, while the lower bound W is the smallest and minimal element. This fact justifies the claim that the multivariate structure associated with M (W) is called *perfect concordance* (*perfect discordance*).

2.1.3 Positive and negative quadrant dependency

The concordance order defines the W–M axis, with the product copula Π being an important point between the two poles. It seems natural to divide all distributions according to their position with respect to the product copula Π. The following definitions were proposed by Lehmann in 1966 [9, p. 1137], see also [7, p. 496, Def. 5].

Definition 17 (Positive and negative quadrant dependency)**.** We say that H is *positively quadrant dependent (PQD)* if $H \succeq \Pi$, or, in other words, if it holds that

$$\forall x, y \in \mathbb{R} : H(x, y) \geq F(x)\, G(y). \tag{2.6}$$

We denote this as $DEP_1(X, Y)$ or $DEP_1(H)$.
We say that H is in *negative quadrant dependency (NQD)* if $\Pi \succeq H$, or, in other words, if it holds that

$$\forall x, y \in \mathbb{R} : H(x, y) \leq F(x)\, G(y). \tag{2.7}$$

These notions may be used by analysts as broad reference points when deciding on the dependency structure suitable for their task. Besides that, distributions in these classes have some interesting properties. For instance, the following proposition is due to Lehmann [9, p. 1140, Lemma 3].

Theorem 14. *Let X and Y be positively quadrant dependent. Provided the expectations below exist, it holds that:*

$$\mathsf{E}[XY] \geq \mathsf{E}[X]\, \mathsf{E}[Y], \tag{2.8}$$

with equality holding if and only if X and Y are independent.

2.2 Abstract Measures of Concordance

In the previous section we established a partial ordering of the set of two-variate distributions with non-atomic marginals. Naturally, the next step is to consistently

extend the partial ordering to a complete one. One technique that could be used to extend the partial ordering $\preceq$ on some set A is constructing a so-called *measure*: a mapping $m\colon A \to \mathbb{R}$ such that $\forall a, b \in A:\ a \preceq b \implies m(a) \leq m(b)$. We then use the order induced by the mapping m to compare the elements that are not comparable by $\preceq$.

The established concordance ordering is based on the pointwise comparison of the unique associated copulae, so without loss of generality we may limit our debate to the set of copulae. To ensure that the extended order is meaningful and of use, Scarsini [10, pp.205–206] proposes an axiomatic definition for the so-called *measures of concordance*. We present the axioms as per Nešlehová [12, p. 546, Def. 1].

Definition 18 (Scarsini's abstract measure of concordance)**.**
Consider a mapping $\kappa : \Gamma \to \mathbb{R}$ where Γ stands for the set of all two-variate random vectors with non-atomic marginals. We say κ is a *measure of concordance* if and only if the following holds:

1. *(Symmetry)* $\kappa(X, Y) = \kappa(Y, X)$.
2. *(Bounds)* $\kappa(X, Y) \in [-1; 1]$.
3. *(Normalisation)* If $X = f(Y)$ a.s. with f strictly *increasing* on the range of Y, then $\kappa(X, Y) = 1$. If $X = f(Y)$ a.s. with f strictly *decreasing* on the range of Y, then $\kappa(X, Y) = -1$.
4. *(Independence)* If X and Y are independent, then $\kappa(X, Y) = 0$.
5. *(Change of sign)* If f is strictly monotone on the range of X, then
$$\kappa(f(X), Y) = \begin{cases} \kappa(X, Y), & \text{for } f \text{ strictly increasing,} \\ -\kappa(X, Y), & \text{for } f \text{ strictly decreasing.} \end{cases}$$
6. *(Continuity)* Consider a sequence of random vectors $(X_n, Y_n) \in \Gamma$ such that $(X_n, Y_n) \xrightarrow{D} (X, Y) \in \Gamma$. Then $\kappa(X_n, Y_n) \underset{n\to\infty}{\longrightarrow} \kappa(X, Y)$.
7. *(Coherence)* For $(X, Y) \succeq (X^*, Y^*)$ it holds that $\kappa(X, Y) \geq \kappa(X^*, Y^*)$.

It is worth observing that the list of axioms is not minimal. Moreover, various authors suggest slightly different, yet equivalent axioms. For instance, Nelsen [1, pp. 168–169, Def. 5.1.7] formulates the *Normalisation* axiom as $\kappa(X, X) = 1$ and $\kappa(X, -X) = -1$.

Let us comment on the choice of the axioms. The *Coherence* axiom ensures that the measure we construct to extend the ordering is in accordance with the underlying partial ordering, namely the concordance ordering.

Symmetry seems to be a reasonable demand since there is no reason to believe that the concordance of (X, Y) is somehow different from (Y, X). The *Symmetry* axiom thus guarantees that these two vectors are equivalent in terms of their concordance.

Normalisation and *Bounds* axioms are consistent with the idea of the W–M axis that we mentioned in Section 2.1.3; they in fact ensure that the lowest

value that κ can attain is ascribed to W, while the largest corresponds to the upper FH-bound M, which in our setting is equivalent to both weak and strong co-monotonicity (Thm. 11). The *Independence* axiom then codifies the role of the product copula Π as an important reference point between the poles, requiring that $\kappa(\Pi) = 0$. It then immediately follows that (X, Y) being in positively quadrant dependence implies $\kappa(X, Y) \geq 0$, while $\kappa(X, Y) \leq 0$ is a necessary condition for NQD.

Intuitively it seems that if X and Y are 'highly concordant', then X and $-Y$ should be 'highly discordant'; this is in line with Pearson's r behaviour, see Theorem 1. This intuition is formalised as the *Change of sign axiom*.

Finally, the *Continuity* axiom requires that κ is continuous with respect to the weak convergence of measures. It can be reformulated as that $P_n \xrightarrow{w} P$ implies that $\kappa(P_n) \to \kappa(P)$ whenever the latter expression makes sense. We may perceive that as a requirement that if distributions P and P^* are 'close', then $\kappa(P)$ should be 'close' to $\kappa(P^*)$.

2.3 Concordance Function

Now that we have introduced the concept of measures of concordance, it is pertinent to ask whether the definition is productive. Do such measures exist? If so, how can we construct one?

One generic technique for constructing measures of concordance relies on the so-called *concordance function*, see [12, p. 547] and [1, p. 159, Th. 5.1.1]. We first introduce the concept of a *reference random vector*.

Definition 19 (Reference random vector)**.**
Consider an arbitrary random vector (X, Y). A *reference vector* for (X, Y) is a random vector $(X^*, Y^*) \in \Gamma(X, Y)$ such that $(X^*, Y^*) \perp\!\!\!\perp (X, Y)$. The vector (X, Y) is then called a *gauged* random vector.

In other words, the gauged and reference vectors share common marginals and are independent from each other, yet their multivariate structure may differ. It is also in place to observe, that the terms *gauged* and *reference* are symmetric: if (X^*, Y^*) is a reference for (X, Y), then (X, Y) is a reference for (X^*, Y^*).

Definition 20 (Concordance function, probabilistic definition)**.**
Let (X, Y) be an arbitrary *gauged* random vector with joint CDF H. Consider a *reference* random vector (X^*, Y^*) with joint CDF H^*. The concordance function is then defined as follows:

$$Q_{X^*,Y^*}(X, Y) := \mathsf{P}\left[(X - X^*)(Y - Y^*) > 0\right] - \mathsf{P}\left[(X - X^*)(Y - Y^*) < 0\right]. \quad (2.9)$$

In line with Remark 11, we put

$$Q_{H^*}(H) := Q_{X^*,Y^*}(X, Y). \quad (2.10)$$

Definition 20 may be interpreted as follows. We calculate the difference of probabilities that the gauged vector is (sharply) concordant with the reference one minus the probability of them being (sharply) discordant. Compare it with Yaari's definition of co-monotonicity (Def. 11).

We may also look at the concordance function in terms of the setting we explored in Section 2.1. The reference vector defines the reference point for the four quadrants. Concordance functions then expresses the difference between the probability of the gauged vector falling in the interiors of $Q_1(X^*, Y^*)$ and $Q_3(X^*, Y^*)$, minus the probability of it being in the interiors of the even quadrants.

In our current setting where (X, Y) is assumed to have non-atomic marginals, the concordance function can be expressed in terms of the associated copulae, see [12, p. 547]. Besides that, Nešlehová [12, p. 553, Note 2] gives a *stochastic* representation, which may be useful for computation purposes.

Theorem 15 (Q is a function of the associated copula)**.**
Let (X, Y) be a random vector with continuous marginals and joint distribution function H. Let (X^*, Y^*) be a reference vector. Let C and C^* be the associated copulae of the gauged and reference random vectors, respectively.
Let $(U, V) \sim C$ and $(U^*, V^*) \sim C^*$. Then it holds that:

$$Q_{X^*,Y^*}(X, Y) = Q_{U^*,V^*}(U, V) = Q_{C^*}(C). \tag{2.11}$$

Besides that, $Q_{C^*}(C)$ can be expressed analytically, yielding:

$$Q_{C^*}(C) = 4 \int_{\mathbb{R}^2} C(u, v)\, dC^*(u, v) - 1. \tag{2.12}$$

Proof. See [1, p. 159, Th. 5.1.1]. □

Remark 3. The expression (2.12) is sometimes used to define the concordance function. In contrast with the *probabilistic definition* presented in Def. 20, the expression given in Eq. (2.12) is then called *analytical definition* of the concordance function.

Theorem 16 (Stochastic representation of Q)**.**
Assume the setting of Theorem 15 holds. Then the concordance function has a stochastic representation:

$$Q_{X^*,Y^*}(X, Y) = 4\ \mathsf{E}[H(X^*, Y^*)] - 1. \tag{2.13}$$

Eq. (2.12) gives us a deeper understanding of the underlying mechanics of the Q function. Specifying the reference copula C^*, we choose which parts of the gauged copula C we want to focus on and what weights are put on them. Setting $C^* := \Pi$, we assess the whole C uniformly (recall that Π corresponds to the distribution uniform on the unit square $[0; 1]^2$). Putting $C^* := M$, we limit ourselves to studying the behaviour of C on the diagonal $\{(t, t) \mid t \in [0; 1]\}$ with weights spread uniformly along it (recall that M is a 2-CDF of a random vector (U, U), where $U \sim Uni(0; 1)$). Nelsen [1, Sec. 5.1, especially pp. 160–162] scrutinises the consequences of choosing different C^*.

Besides that, it is also worth observing some other convenient properties of the concordance function. For instance, it turns out that Q is symmetric in terms of the gauged and reference distributions, see [1, p. 160, Cor. 5.1.2].

Theorem 17 (Properties of concordance function)**.**

1. *Symmetry.* $Q_{C^*}(C) = Q_C(C^*)$.

2. *Non-decreasing in reference. Let* $C^* \succeq C'$. *Then* $Q_{C^*}(C) \geq Q_{C'}(C)$.

3. *Non-decreasing in the gauged vector. Let* $C \succeq C'$. *Then* $Q_{C^*}(C) \geq Q_{C^*}(C')$.

The property of being non-decreasing in the gauged distribution hints that Q can be used as a basis to construct Scarsini's measures of concordance.

2.4 Examples of Measures of Concordance

Choosing different references for the concordance function Q and applying suitable normalisations, we obtain different indices focusing on different aspects of the gauged dependence structure C. Consider the following instances.

Spearman's ρ

If we put $C^* := \Pi$, which is equivalent to setting the elements of the reference random vector independent, we obtain the probabilistic expression for Spearman's ρ, see [7, p. 490], [12, p. 547] and [1, p. 167]. This measure gauges the whole copula C uniformly.

Definition 21 (Population Spearman's ρ, probabilistic definition)**.**
Consider an arbitrary random vector (X, Y). Let (X^*, Y^*) be a reference vector such that $X^* \perp\!\!\!\perp Y^*$. We define Spearman's ρ as follows:

$$\rho(X, Y) := 3\, Q_{(X^*, Y^*)}(X, Y). \tag{2.14}$$

Let H be the joint cumulative distribution function of the random vector (X, Y). In line with Remark 11, we put:

$$\rho(H) := \rho(X, Y) \tag{2.15}$$

Using Theorems 15 and 16 we obtain a *stochastic* representation and some *analytical* expressions for Spearman's ρ which may be useful for better understanding of its mechanics as well as for computational purposes. Regarding the stochastic representation, see [7, p. 491, Prop. 9].

Theorem 18 (Stochastic representation of Spearman's ρ)**.**
Let (X, Y) *be a random vector with non-atomic marginals. Let* H *be the joint CDF and assume* F *and* G *for the respective marginal distribution functions. Let* (X^*, Y^*) *be a reference vector such that* $X^* \perp\!\!\!\perp Y^*$. *Then it holds that*

$$\rho(X, Y) = 12\ \mathsf{E}[H(X^*, Y^*)] - 3 = 12\, \mathsf{E}[F(X)\, G(Y)] - 3. \tag{2.16}$$

Theorem 19 (Alternative forms of Spearman's ρ)**.**
Assume (X, Y) *has non-atomic marginals and let* C *be its unique associated copula. Then it holds that:*

$$\begin{aligned}
\rho(X, Y) &= 3\, Q_{\Pi}(C) && (2.17)\\
&= 12 \int_{[0;1]^2} C(u, v)\ du\, dv - 3 && (2.18)\\
&= 12 \int_{\mathbb{R}^2} uv\, dC(u, v) - 3 && (2.19)\\
&= r\left(F(X), G(Y)\right) && (2.20)\\
&= 12 \int_{[0;1]^2} C(u, v) - \Pi(u, v)\ du\, dv. && (2.21)
\end{aligned}$$

Proof.

Eqs. (2.18) and (2.17). Using Theorem 15, we get:

$$\begin{aligned}\rho(X,Y) = 3\,Q_\Pi(C) &= 12 \int_{\mathbb{R}^2} C(u,v)\,d\Pi(u,v) - 3 \\ &= 12 \int_{\mathbb{R}^2} C(u,v)\ d\check{u}\,d\check{v} - 3 \\ &= 12 \int_{[0;1]^2} C(u,v)\ du\,dv - 3.\end{aligned}$$

Eq. (2.19). Theorem 17 ensures that $Q_\Pi(C) = Q_C(\Pi)$. Observing that C concentrates all mass inside the unit square $[0;1]^2$, we get:

$$\begin{aligned}\rho(X,Y) = 3\,Q_\Pi(C) = 3\,Q_C(\Pi) &= 12 \int_{\mathbb{R}^2} \Pi(u,v)\ dC(u,v) - 3 \\ &= 12 \int_{\mathbb{R}^2} \check{u}\,\check{v}\ dC(u,v) - 3 \\ &= 12 \int_{\mathbb{R}^2} uv\ dC(u,v) - 3.\end{aligned}$$

Eq. (2.20) follow from the previous equality and Theorem 3.

Eq. (2.21). Using (2.19) and the linearity of integral, we get:

$$\begin{aligned}\rho(X,Y) &= 12 \int_{[0;1]^2} C(u,v)\ du\,dv - 3 \\ &= 12 \int_{[0;1]^2} C(u,v)\ du\,dv - 12 \int_{[0;1]^2} uv\ du\,dv \\ &= 12 \int_{[0;1]^2} C(u,v) - uv\ du\,dv \\ &= 12 \int_{[0;1]^2} C(u,v) - \Pi(u,v)\ du\,dv.\end{aligned}$$

□

All of these alternative representations rely on the assumption of non-atomic marginals. Notice that Eq. (2.20) establishes the connection with the observations we made in Section 1.1 and justifies and generalises the population Spearman's ρ proposed in Definition 4. Eq. (2.21) is of interest as well; it suggests that Spearman's ρ can be interpreted as rescaled average distance between the gauged dependence structure C and independence. Finally, eq. (2.16) is of interest when compared with a similar representation of Kendall's τ, see eq. (2.28).

The sequence of equalities is sometimes used to define Spearman's ρ *analytically*; see, for instance [10, p. 208], [7, p. 480] or [12, p. 547]. Compare this distinction with the distinction in case of the concordance function (Rem. 3).

Definition 22 (Population Spearman's ρ, analytical definition)**.**
Let (X,Y) be a random vector with non-atomic marginals and let C be its unique associated copula. The *analytical definition* of Spearman's ρ is:

$$\rho(X,Y) = \rho(C) = 3\,Q_\Pi(C) = 12 \int_{[0;1]^2} C\ du\,dv - 3. \tag{2.22}$$

Kendall's τ

Another option that naturally comes to mind is to use an independent copy of (X, Y) as a reference. This leads to the probabilistic definition of Kendall's τ, see [7, p. 490], [12, p. 547] and [1, pp. 158–160]. This measure again gauges the whole copula C, yet the weights are defined by the copula itself.

Definition 23 (Population Kendall's τ, probabilistic definition)**.**
Consider an arbitrary random vector (X, Y). Let $(X^*, Y^*) \sim (X, Y)$ be a reference vector. We then define Kendall's τ as follows:

$$\tau(X, Y) := Q_{X^*,Y^*}(X, Y). \tag{2.23}$$

Let H be the joint cumulative distribution function of the random vector (X, Y). In line with Remark 11, we put:

$$\tau(H) := \rho(X, Y) \tag{2.24}$$

For Kendall's τ, we also present several alternative stochastic and analytical representations. Regarding the stochastic representation, see [7, p. 491, Prop. 9].

Theorem 20 (Stochastic representation of Kendall's τ)**.**
Assume (X, Y) has non-atomic marginals and let H be its joint distribution function. Then it holds that:

$$\tau(X, Y) = 4\,\mathsf{E}[H(X, Y)] - 1. \tag{2.25}$$

Theorem 21 (Alternative forms of Kendall's τ)**.**
Assume (X, Y) has non-atomic marginals. Let H be its joint CDF and let C be its unique associated copula. Let $(U, V) \sim C$ and put $W := C(U, V)$. Let F_W be the CDF of W. It then holds that:

$$\begin{aligned}
\tau(X, Y) &= Q_C(C) && (2.26)\\
&= 4 \int_{\mathbb{R}^2} C(u, v)\, dC(u, v) - 1 && (2.27)\\
&= 4\,\mathsf{E}\left[C\left(F(X), G(Y)\right)\right] - 1 && (2.28)\\
&= 4 \int_{\mathbb{R}} w\, dF_W(w) - 1. && (2.29)
\end{aligned}$$

Proof.

Eqs. (2.26) and (2.27) are immediate consequences of Theorem 17.

Eq. (2.28). Interpreting the previous equality as expectation under the distribution induced by the 2-CDF C, we get:

$$\tau(X, Y) = 4 \int_{\mathbb{R}^2} C(u, v)\, dC(u, v) - 1 = 4\,\mathsf{E}\left[C\left(F(X), G(Y)\right)\right] - 1.$$

Eq. (2.29). Since $(F(X), G(Y)) \sim C$, we then get:

$$\tau(X, Y) = 4\,\mathsf{E}\left[C\left(F(X), G(Y)\right)\right] - 1 = 4\,\mathsf{E}[W] - 1 = 4 \int_{\mathbb{R}} w\, dF_W(w) - 1.$$

□

The transformation $H(X, Y)$ is the two-variate analogue of probability integral transform (Th. 56); in our setting it is equivalent to $C\left(F(X), G(Y)\right)$. Because of eq. (2.29), the distribution of the random variable obtained by such transformation is called *Kendall's distribution* (see [4, p. 7]). The transform is, for instance, employed to develop diagnostic tools and goodness-of-fit tests used in copula modelling, see also [6, p. 352] and [1, p. 163].

The sequence of equalities is sometimes used to define Kendall's τ *analytically*; see, for instance [10, pp. 208–209, Th. 5], [7, p. 480] or [12, p. 547]. Compare this distinction with the distinction in the case of the concordance function (Rem. 3).

Definition 24 (Population Kendall's τ, analytical definition)**.**
Let (X, Y) be a random vector with non-atomic marginals and let C be its unique associated copula. The *analytical definition* of Kendall's τ is:

$$\tau(X, Y) = \tau(C) = Q_C(C) = 4 \int_{\mathbb{R}^2} C(u, v)\, dC(u, v) - 1. \tag{2.30}$$

Other examples

Choosing the FH-bounds as the reference multivariate structures we obtain Corrado Gini's correlation index. As we can see, it gauges the dependence structure's behaviour along the diagonals of the unit square $[0; 1]^2$, see also [1, pp.180–182].

Definition 25 (Gini's γ)**.**

$$\gamma(X, Y) := Q_M(C) + Q_W(C). \tag{2.31}$$

Denote the random variables median as m_X. If we use the point (m_X, m_Y) as a reference to define the the quadrants, and then calculate the difference between the probabilities of being concordant and discordant, we obtain *Blomqvist's medial correlation coefficient*. As we can see, it gauges the dependency structure at the single point $(\frac{1}{2}, \frac{1}{2})$, yet even this crude approach can be informative; see also [1, p. 182].

Definition 26 (Blomqvist's β)**.**

$$\beta(X, Y) := \mathsf{P}[(X - m_X)(Y - m_Y) > 0] - \mathsf{P}[(X - m_X)(Y - m_Y) < 0] \tag{2.32}$$

$$= 4\, C\left(\frac{1}{2}, \frac{1}{2}\right) - 1. \tag{2.33}$$

2.4.1 Measure of concordance

It turns out that the four aforementioned indices based on the concordance function do fulfil Scarsini's axioms and are measures of concordance in the sense of Definition 18.

Theorem 22. *Spearman's ρ, Kendall's τ, Gini's γ and Blomqvist's β are Scarsini's measures of concordance.*

Proof.

Spearman's ρ. See [10, pp. 207–208] or [1, p. 169, Th. 5.1.9].

Kendall's τ. See [10, pp. 208–210, Th. 5] or [1, p. 169, Th. 5.1.9].

Gini's γ. See [10, pp. 208–210, Th. 5].

Blomqvist's β. See [10, pp. 207–208].

□

Recall that initially we introduced measures of concordance as a way of extending the concordance partial ordering to a complete order. Needless to say that different Scarsini measures κ yield different extensions. The discrepancies in the values of two different measures of concordance can be informative; see, for instance, [1, pp.174–180, Sec. 5.1.3] where Nelsen explores the relationship between Spearman's ρ and Kendall's τ.

2.5 Further Dependency Concepts

Besides the positive and negative quadrant dependencies introduced in Definition 17, there are many more. We mention several of them for the sake of completeness. The following list is based on [7, p. 496, Def. 5].

Definition 27. Consider an arbitrary random vector (X, Y). Then

- Y is *left-tail decreasing* in X, denoted $DEP_2(X, Y)$, if and only if for each $y \in \mathbb{R}$ it holds that $\mathsf{P}[Y \leq y | X \leq x]$ is a non-increasing function of the variable x.
- Y is said to be *right-tail increasing* in x, denoted $DEP_3(X, Y)$, if and only if for each $y \in \mathbb{R}$ it holds that $\mathsf{P}[Y \leq y | X > x]$ is a non-decreasing function of the real variable x.
- Y is *stochastically increasing* in X, denoted $DEP_4(X, Y)$, if and only if for each $y \in \mathbb{R}$ it holds that $\mathsf{P}[Y \leq y | X = x]$ is a non-increasing function of x.
- X and Y are in *positive likelihood ratio dependence*, denoted $DEP_5(X, Y)$, if and only if the following holds:

  $$\forall x_1 < x_2 \in \mathbb{R},\ \forall y_1 < y_2 \in \mathbb{R}: \\ h(x_1, y_1)h(x_2, y_2) \geq h(x_1, y_2)h(x_2, y_1), \quad (2.34)$$

  where $h(x, y)$ is the probability or density function of (X, Y). The concept only makes sense if the density function exists.

In line with Remark 11, we may also write $DEP_i(H)$, where H is the joint CDF.

In case of the non-atomic margins, all of these concepts are captured by the unique associated copula.

Theorem 23. *Let (X, Y) be a random vector with non-atomic marginals. Let C be its unique associated copula. Then for $i \in \{1, \ldots, 5\}$ the following statements are equivalent:*

1. $DEP_i(X, Y)$ *holds,*

2. $DEP_i(C)$ *holds.*

Proof. Refer to [7, p. 497, Prop. 11]. □

Part II

Arbitrary Distributions

3 Copulae for Arbitrary Distributions

In Chapter 1, we established the foundations of the 2-copulae theory in the case of distributions with non-atomic marginals. The non-atomic marginals assumption is a popular setting, many classic books focus more or less exclusively on it (e.g., Nelsen [1] or Durante and Sempi [2]). It is safe to say that the classical copula theory is based on that assumption.

Although assuming the marginals are non-atomic is popular, it is rather limiting. In this chapter, we shall study the properties of copulae for arbitrary two-variate distributions. Apart from [7], there does not seem to be any review, nor summary texts dealing with the copula models for arbitrary distributions. One of the goals of this chapter is to collect in one place all the key facts about probabilistic copula models for arbitrary distributions.

The chapter is organised as follows. In Section 3.1, we demonstrate why the naive attempt to apply the methods we used previously to the case of arbitrary distribution fails. Section 3.2 reformulates Sklar's theorem for the case of atomic distributions and shows that the associated copula is no longer unique; the section demonstrates some consequences it has for the copula models and outlines different ways of dealing with the problems that arise from the atoms being present in the marginal distributions. Section 3.3 focuses on discrete distributions and showcases how detrimental the consequences of the fact that the associated copula is longer unique can be; it also presents several tools that allow us to assess the extent of the problems in each particular case. In Section 3.4 we explore the techniques that lead to deriving the so-called *standard extension copula*; we demonstrate that to a great extent it plays the role of the unique associated copula we worked with in the case of non-atomic marginals and hence can be considered as the embodiment of the dependence structure of an arbitrary random vector. Finally, in Section 3.5, we summarise the properties of synthetic copula models under the non-atomicity assumption, explore their behaviour in the case of arbitrary distributions and study how co-monotonicity and FH-bounds work when atoms are present in the marginals.

3.1 Differences with the Non-Atomic Case

Recall how we initially motivated the study of copulae in Section 1.1. We studied a random vector (X, Y), H was its joint cumulative distribution function, F and G were its marginal CDFs, which were assumed to be continuous. We then exploited the probability integral transform (Th. 56) to obtain a new transformed vector $(U, V) := (F(X), G(Y))$. Due to Theorem 56, it held that $U, V \sim Uni(0; 1)$, and the 2-CDF $C_{U,V}$ was the (unique) copula associated with the random vector (X, Y) and its 2-CDF H. Can we do the same trick without assuming the marginals to be non-atomic?

Alas, this approach does not work if there are atoms present in the marginal distributions. Specifically, the problem is that $F(X)$ is not distributed uniformly if X is arbitrary. Moreover, the closed-form expression for the associated copula

that we derived in Theorem 4 is not valid either. These facts are illustrated by the example below which loosely follows [7, pp. 478–479, Ex. 1].

Example 1. Consider a random vector (X, Y). Let the elements be independent Bernoulli random variables with probabilities of success p and q, respectively. Let H be its joint CDF, and let F and G be the marginal CDFs. It then holds that:

$$F(x) := \mathsf{P}[X \leq x] = \begin{cases} 0, & x \in (-\infty; 0); \\ 1-p, & x \in [0; 1); \\ p, & x \in [1; \infty). \end{cases} \tag{3.1}$$

It is easy to verify that $F(X)$ is a discrete random variable that is equal to $1-p$ with probability $1-p$, and with probability p it equals 1; clearly, it is not distributed uniformly. Hence the CDF of $(F(X), G(Y))$ by definition is not a copula.
Consider the quantile function (Def. 45) for X:

$$F_X^{-1}(u) := \inf\{x : F(x) \geq \check{u}\} = \begin{cases} 0, & u \in (-\infty; 1-p]; \\ 1, & u \in (1-p; \infty). \end{cases} \tag{3.2}$$

The expression $\check{x}$ in the equality above stands for the collar operator (Definition 7). Clearly, $H(F^{-1}(u), G^{-1}(v))$ is not a copula.

More details could be found, for instance, in [7, pp. 477–480].

3.2 Analytical Models. Extracting Associated Copulae from Joint CDF

As we have seen in the previous section, the first two out of three copula properties we outlined in Theorem 4 fail if atoms are present in the marginal distributions. The property we are left with is the coupling property, or Sklar's representation (Eq. (1.15)). Specifically, let H be an arbitrary two-variate CDF, and F and G be its marginal CDFs. Sklar's representation then goes as follows:

$$H(x, y) = C(F(x), G(y)), \tag{3.3}$$

where C is (usually) taken to be a copula. Given H arbitrary, is it possible to derive a copula that fulfils the Sklar's representation? Is the copula unique? If a copula is compatible with H, does it capture the dependence structure in a margin-free fashion as it was in the case of the non-atomic margins?

Recall that Sklar's theorem (Th. 9) as it was stated in Chapter 1 guarantees there exists at least one copula C compatible with H; if the marginals have atoms, it is however no longer guaranteed that the copula is unique. Consider the following example from [7, p. 488, Ex. 5] illustrating the issue.

Example 2. Let X and Y be independent identically distributed Bernoulli random variables with the probability of success equal to p. Clearly the product copula Π is associated with the random vector (X, Y). Yet the copula $C := \frac{M+W}{2}$ is also compatible with it.

This fact is explained by the ranges of F and G. Notice that Sklar's representation implies that H is defined by the behaviour of C on the set $\operatorname{Ran} F \times \operatorname{Ran} G$.

Indeed, if the marginals are non-atomic, then F and G are continuous functions. Hence $\operatorname{Ran} F, \operatorname{Ran} G \supseteq (0;1)$, H thus depends on the behaviour of the copula C at least on the interior of the unit square $(0;1)^2$. This fact allows us to restore the whole C uniquely.

The presence of atoms in the marginals corresponds to 'jumps' in the marginal CDFs. Hence the joint CDF H is determined by how an associated copula behaves on $\operatorname{Ran} F \times \operatorname{Ran} G \subset [0;1]$. We thus have no information about C outside of the marginals' ranges product, which may be a problem when we try to reconstruct the whole C.

This allows us to reformulate the Sklar's theorem in a more general way.

Theorem 24 (General Sklar's theorem)**.** *Consider an arbitrary random vector (X, Y) with the joint CDF $H_{X,Y}$. Let F_X and G_Y be the marginal distribution functions. Then the following holds:*

1. *There exists at least one copula C associated with (X, Y).*
2. *C is determined uniquely on the set $\overline{\operatorname{Ran} F \times \operatorname{Ran} G}$. Specifically, if X and Y are non-atomic, then the associated copula C is unique and thus can be denoted as $C_{X,Y}$ or C_H.*
3. *Conversely, if C is a copula and F and G are some arbitrary univariate CDFs, then $C(F(x), G(y))$ defines a valid 2-CDF.*

Proof. See, for instance, [1, p. 18, Theorem 2.3.3]. Various ways of proving the theorem are also available in [2, pp. 48–57]. The original arguments by Sklar and Schweizer are presented in [14] and [15]. □

Let us reiterate and summarise what these facts mean for probabilistic copula models. First, there may be many copulae associated with an arbitrary 2-CDF H. It means that the notation C_H is not valid anymore as it is not clear to which of the several compatible copulae we refer. We thus shall use it to denote the *set of all copulae associated* with the joint CDF H.

Second, even if we manage to find a copula that is compatible with H, generally it does not capture the dependency properties. This fact has long-lasting consequences that manifest in various ways. One such instance was given in Example 2: a copula $\frac{M+W}{2} \neq \Pi$ induced independence. Another example is that analytical definitions of Spearman's ρ and Kendall's τ (Defs. 22 and 24) do not coincide with the probabilistic ones (Defs. 21 and 23) anymore, the same goes for the concordance function. This seemingly strips copula models of the vast majority of their benefits.

Three ways to tackle this issue come to mind.

The first approach is to give up on insisting that we use copulae *per se* to describe the dependency structure. Theorem 24 ensures that we can uniquely reconstruct the restriction of an associated copula to the closure of $\operatorname{Ran} F \times \operatorname{Ran} G$ (such a restriction is often called a *subcopula*, see [1, p. 10, Def. 2.2.1] or [2, p. 49, Def. 2.3.2]); it thus seems natural to consider these restrictions of copulae as the embodiment of the dependency structure (at least for the given Fréchet class) and focus on them. This way of thought however seems to be unpopular.

The second approach is to put aside our mathematical rigour for a moment and 'call upon our internal engineer': we may hope that although the associated copula is not defined uniquely anymore, the set C_H of all compatible copulae would be reasonably 'small' with all its elements being 'close' to each other. Hence considering various associated copulae would lead to differences that are negligible, so taking some arbitrary compatible copula would make have no significant impact on the analysis. As we will see in Section 3.3, this hope may be quite misleading; it turns out that the copulae compatible with some given joint CDF may be quite 'far apart' from each other and represent rather different dependency structures, which renders this approach non-viable in the general case.

Finally, the third approach is to try to find the 'true' copula among all compatible copulae that would accurately represent the dependency structure. Section 3.4 focuses on that approach. As we will see, this technique, although not straight-forward, is rather productive.

3.3 Deeper Scrutiny of Associated Copulae Set

In this section, we are going to scrutinise the set C_H of copulae associated with a given joint cumulative distribution function H; we will however limit ourselves to studying only discrete random vectors.

Consider a *discrete* random vector (X, Y) with H being its joint CDF; let F and G be its marginal distribution functions and assume they are discrete. The first question that we may ask is how many copulae are compatible with H? There are two ways to approach this question.

3.3.1 Carley bounds

If both X and Y are discrete, the set of copulae compatible with H can be bounded pointwisely. The bounds were initially identified in 2002 by Carley [16] under the assumption that X and Y take finitely many values. The construction was generalised by Genest and Nešlehová [7, pp. 480–481] in 2007.

Theorem 25 (Carley bounds)**.**
Let (X, Y) be a discrete random vector with the joint CDF H. Without loss of generality, we may assume that X and Y a.s. take values in $\mathbb{N}_0$. Put

$$h_{ij} := \mathsf{P}[X = i, Y = j], \quad i, j \in \mathbb{N}_0. \tag{3.4}$$

Denote the marginal probabilities as

$$h_{i+} := \sum_{j=0}^{\infty} h_{ij} = \mathsf{P}[X = i], \qquad h_{+j} := \sum_{i=0}^{\infty} h_{ij} = \mathsf{P}[Y = j]. \tag{3.5}$$

Define the lower C_H^- and upper C_H^+ Carley bounds as follows:

$$C_H^-(u, v) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max\{0,\ -h_{ij} + \min(u - \gamma_{ij}, h_{ij}) + \min(v - \delta_{ij}, h_{ij})\}, \tag{3.6}$$

$$C_H^+(u, v) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max\{0,\ \min(u - \alpha_{ij},\ v - \beta_{ij}, h_{ij})\}, \tag{3.7}$$

where

$$\alpha_{ij} := \sum_{m=0}^{i-1} h_{m+} + \sum_{n=0}^{j-1} h_{in}, \qquad \beta_{ij} := \sum_{n=0}^{j-1} h_{+n} + \sum_{m=0}^{i-1} h_{mj},$$
$$\gamma_{ij} := \sum_{m=0}^{i-1} h_{m+} + \sum_{n=j+1}^{\infty} h_{in}, \qquad \delta_{ij} := \sum_{n=0}^{j-1} h_{+n} + \sum_{m=i+1}^{\infty} h_{mj}. \tag{3.8}$$

Let C_H be a set of copulae associated with H. For any $C \in C_H$, it then holds that:

$$\forall u, v \in \mathbb{R} : C_H^-(u, v) \leq C(u, v) \leq C_H^+(u, v). \tag{3.9}$$

Besides that, it holds that $C_H^-, C_H^+ \in C_H$.

Proof. See [16]. For more details, refer to [7, pp.480–486]; Example 2 in [7] is of special interest. □

Remark 4. Observe that if the joint distribution is discrete with values not in $\mathbb{N}_0^2$, we may enumerate the atoms. Let $x_i, y_j \in \mathbb{R}$ for $i, j \in \mathbb{N}_0$ denote the atoms of X and Y, respectively. Put $h_{ij} := \mathsf{P}[X = x_i, Y = y_j]$. The constructions then remains unchanged apart from the fact that instead of working with $\mathbb{N}_0^2$ we will work with the grid defined by the location of atoms x_i and y_j.

The theorem essentially says that for a given discrete H, we can derive the bounds for the set of compatible copulae C_H. These bounds C_H^- and C_H^+ are copulae themselves. Although we cannot directly apply the theory of concordance ordering developed in Chapter 2 to study discrete distributions, it is still valid for the elements of C_H. Together with Theorem 13, Equation 3.9 may be reinterpreted as that for any $C \in C_H$, the following holds:

$$C_H^- \preceq C \preceq C_H^+. \tag{3.10}$$

Carley bounds C_H^- and C_H^+ thus correspond, respectively, to the most discordant and concordant dependency structures compatible with H (recall Def. 15).

Carley bounds allow us to answer several questions.

The set C_H is uncountable

To begin with, recall Theorem 6: any convex combination of copulae is a copula again. Besides that, all elements of C_H coincide on $\operatorname{Ran} F \times \operatorname{Ran} G$. Hence $\forall \alpha \in (0; 1) : \alpha\, C_H^- + (1 - \alpha)\, C_H^+ \in C_H$. It means, that the set C_H is uncountable. So, at least when we are talking about discrete distributions, there exist *many* associated copulae.

Gap between minimum and maximum τ and ρ

As we have said, Carley bounds represent the most discordant and concordant dependency structures that are compatible with H. We thus may compare how far apart these structures span along the W–M concordance axis. Genest and Nešlehová [7, pp. 483–485] suggest the following.

Consider some abstract measure of concordance κ. From Scarsini axioms (Def. 18), specifically from the axiom of *Coherence*, we get that for any $C \in C_H$

it holds that $\kappa(C_H^-) \leq \kappa(C) \leq \kappa(C_H^+)$. Notice that these bounds are the best possible, since both Carley bounds belong to C_H. Moreover, considering various convex combinations of Carley bounds, we get that:

$$\forall \kappa \in \left[\kappa(C_H^-); \kappa(C_H^+)\right] \exists C \in C_H : \kappa(C) = \kappa. \tag{3.11}$$

Hence we may use the difference $0 \leq \kappa(C_H^+) - \kappa(C_H^-) \leq 2$ to quantify how 'large' the set C_H is.

Notice that all elements of C_H have non-atomic margins, hence $\kappa(C)$ is defined for all elements of C_H. Moreover, for instance, in case of Spearman's ρ or Kendall's τ, the probabilistic definitions (Defs. 21 and 23) coincide with the analytical ones (Defs. 22 and 24) and there is no ambiguity in writing $\tau(C)$. On the other hand, it is worth noting that, thus far, we have not defined measures of concordance for discrete distributions (see Def. 18) and the expression $\kappa(H)$ may not be well-defined. This is yet another illustration of how dependency properties of a joint CDF H are detached from those of the compatible copulae.

The following theorems give closed-form expressions for Kendall's τ and Spearman's ρ evaluated at the Carley bounds. The theorems are due to Genest and Nešlehová [7, p. 484, Props. 4 and 5].

Theorem 26 (Kendall's *tau* and Spearman's ρ for Carley bounds)**.** *Consider a discrete random vector (X, Y) with its values in $\mathbb{N}_0^2$ and let H be the corresponding joint CDF. Define h_{ij}, α_{ij}, β_{ij}, γ_{ij}, δ_{ij} as in Theorem 25.*
For Kendall's τ we then get:

$$\tau(C_H^-) = -1 + 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{m=0}^{i-1} \sum_{n=0}^{j-1} h_{ij}\, h_{mn}, \tag{3.12}$$

$$\tau(C_H^+) = 1 - 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{m=0}^{i-1} \sum_{n=j+1}^{\infty} h_{ij}\, h_{mn}. \tag{3.13}$$

For Spearman's ρ it holds that

$$\rho(C_H^-) = 1 - 6 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij}\, \left[\delta_{ij} + h_{ij} - (1 - \gamma_{ij})\right] \left[2(1 - \gamma_{ij}) - h_{ij}\right], \tag{3.14}$$

$$\rho(C_H^+) = 1 + 6 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij}\, (\beta_{ij} - \alpha_{ij})\, (2\alpha_{ij} + h_{ij}). \tag{3.15}$$

Proof. See [7, pp. 508–510, Appendices B and C]. □

The following example follows Example 3 in [7, pp. 484–485] and illustrates that the difference $\kappa(C_H^+) - \kappa(C_H^-)$ can be rather large.

Example 3. Let X and Y be Bernoulli random variables with the probabilities of success p and q, respectively. Put

$$r := \mathsf{P}[X = 0, Y = 0] \in [\max(0, 1 - p - q), \min(1 - p, 1 - q)]. \tag{3.16}$$

The joint distribution H is fully specified by p, q, r and a copula C is compatible with H if and only if $C(1 - p, 1 - q) = r$. From Theorem 26 we then get:

$$\tau(C_H^-) = 4r(p + q + r - 1) - 1, \tag{3.17}$$

$$\rho(C_H^+) = 1 - 6(1 - p - r)(1 - q - r)(2 - 2r - p - q), \tag{3.18}$$

$$\rho(C_H^-) = 6r(p + q + r - 1)(p + q + 2r - 1) - 1, \tag{3.19}$$

$$\tau(C_H^+) = 1 - 4(1 - p - r)(1 - q - r). \tag{3.20}$$

Consider the special case of $p = q$. Depending on p and r, we get:

- Kendall's τ for the lower Carley bound spans the interval $[-1; 0]$, for the upper bound it covers the interval $[0; 1]$. The difference $\tau(C_H^+) - \tau(C_H^-)$ attains values in the interval $[1; 2]$.
- Spearman's ρ for the lower bound varies in the interval $[-1; 0.5]$, the upper bound for ρ lies in the interval $[-0.5; 1]$. The difference $\rho(C_H^+) - \rho(C_H^-)$ covers the interval $[0.5; 2]$.

As we can see, in terms of concordance indices, C_H may potentially contain dependence structures that differ a lot; copulae compatible with a given joint cumulative distribution function H need not be 'close' to each other. Another point that is illustrated by the example is that copulae do not capture the dependence structure in a margin-free fashion: possible values of τ and ρ depend on the marginal distributions.

3.3.2 Flatness of margins

Carley bounds explore possible dependence structures compatible with a given element of some Fréchet class $\Gamma(F, G)$ represented. Another approach is to explore the whole set $\Gamma(F, G)$ and the constraints its elements put on the associated copulae.

As we know, for a non-atomic marginal CDF F it holds that $\operatorname{Ran} F \supseteq (0; 1)$. The presence of atoms in the distribution corresponds to discontinuities in F and leads to the parts of $\operatorname{Ran} F$ being 'clipped out', which subsequently limits our ability to reconstruct the copula associated with the two-variate CDF. One may hope though that if the atoms are not too large, a substantial part of $\operatorname{Ran} F$ will be intact and thus the constraint on the set of compatible copulae will be stringent enough to allow the reconstruction of the associated copulae with substantial precision.

Think of purely discrete random variables. If each atom is 'small', both $\operatorname{Ran} F$ and $\operatorname{Ran} G$ define 'fine' and 'dense' grids on the interval $[0; 1]$. Hence their product defines a 'dense' grid on the unit square $[0; 1]^2$, hopefully limiting the set of compatible copulae substantially. As it turns out, this idea does work; in 2007, Genest and Nešlehová proved the following theorem.

Theorem 27 (Diameter of C_H is given by the flatness of margins)**.**
Consider a discrete random vector (X, Y) with the joint CDF H. Without loss of generality, assume it takes values in $\mathbb{N}_0^2$. Let $C, D \in C_H$ be two arbitrary copulae compatible with H. Then it holds that:

$$\forall u, v \in (0, 1) : |C(u, v) - D(u, v)| \leq 2 \left(\max_{i \in \mathbb{N}} \mathsf{P}[X = i] + \max_{j \in \mathbb{N}} \mathsf{P}[Y = j] \right). \quad (3.21)$$

Proof. See [7, pp. 486–487, Prop. 6]. □

Genest and Nešlehová refer to the right-hand side as 'the flatness of margins'. Indeed, the lower is the right-hand side, the smaller are the atoms in the margins.

The theorem hence may be reformulated as follows. Introduce the supremum metric on the set of all copulae. Consider discrete CDFs F and G. The diameter of

the set of all copulae compatible with distributions from $\Gamma(F, G)$ is then bounded by the 'flatness' of F and G.

This theorem thus can be used to find the upper bounds for the differences between the properties of compatible copulae; if, for a given Fréchet class (Def. 14), the differences are small enough, one may use an arbitrary compatible copula and the error will be negligible.

3.4 Standard Extension Copula and Jittering

As we have discussed, atoms present in the marginal distributions correspond to jumps in the marginal CDFs and lead to the fact that the probability integral transform (Th. 56) does not yield a uniformly distributed random variable. Ultimately, it prevents us from reconstructing the unique associated copula.

On the sample level, atoms correspond to the presence of ties in a random sample. Among other consequences, it means that observations' ranks (Def. 49) are not defined uniquely anymore.

It seems natural to try breaking the ties by 'jittering' them; i.e., adding to them some (non-atomic) random noise. On the sample level, this technique is usually discouraged since the result of the analysis depends on an additional source of randomness which is undesirable for a variety of reasons. As it is shown below, on the population level, the technique is fairly productive and its application leads to interesting results.

In the following, we first outline the general jittering technique and the corresponding generalisation of probability integral transform. We then consider a special case of jittering which leads us to the so-called standard extension copula.

3.4.1 Jittered probability integral transform

The simplest instance of jittering can be illustrated as follows.

Example 4. Consider a discrete random variable X with values in $\mathbb{N}_0$. Consider a 'jittering' random variable $U \sim Uni(0; 1)$ independent of X. The 'jittered' version of X is defined as $X' := X + U - 1$. Notice that X' is a continuous random variable and thus it holds that $F_{X'}(X + U - 1) \sim Uni(0; 1)$. Nešlehová [12, p. 550, Note 1] refers to X' as 'continuous extension of X'. The jittered random variable X' *per se* is of secondary interest; its distribution function is much more important, since it could be perceived as a generalisation of the probability integral transform (Th. 56).

In 2007, Nešlehová generalised the technique for the case of arbitrary random variables, see [12, pp.549–553, Sec. 3].

Definition 28 (Jittered probability integral transform)**.** Consider an arbitrary random variable X and let F be the corresponding CDF. Let $U \sim Uni(0; 1)$ be independent from X. Put $\Delta F(x) := F(x) - F(x-) = \mathsf{P}[X = x]$.
Consider a mapping $\psi : \mathbb{R} \times [0; 1] \to [0; 1]$ defined as

$$\psi(x, u) := \mathsf{P}[X < x] + u\ \mathsf{P}[X = x] = F(x-) + u\,\Delta F(x). \tag{3.22}$$

The *jittered probability integral transform (JPIT)* is then defined as $\psi(X, U)$.

Observe, that $\psi(X, U) \sim Uni(0; 1)$. This technique is easily generalised to random vectors (see [12, p. 550]).

Definition 29 (Jittered probability integral transform for vectors)**.**
Let $\mathbf{Z} := (X, Y)$ be an arbitrary random vector. Consider a (jittering) random vector $\mathbf{U} := (U_1, U_2)$ such that $U_1, U_2 \sim Uni(0; 1)$. Assume $\mathbf{Z} \perp\!\!\!\perp \mathbf{U}$. The jittered probability integral transform is then defined element-wisely:

$$\Psi(\mathbf{Z}, \mathbf{U}) := (\psi(X, U_1), \psi(Y, U_2)) . \tag{3.23}$$

Notice that we do not specify the dependency structure of the random vector (U_1, U_2). Besides that, since its marginals are distributed uniformly, its joint distribution function is by definition a copula. Moreover, the transformed vector $\Psi(\mathbf{Z}, \mathbf{U})$ has uniformly distributed marginals; hence its 2-CDF is the unique copula associated with it.

An important property of the jittered probability integral transform is that it in certain sense preserves the dependency structure of the random vector (X, Y).

Theorem 28 (JPIT preserves the dependency structure)**.** *Consider the setting as in Definition 29. For any dependence structure of* $\mathbf{U}$*, the unique copula* C *associated with the vector* $\Psi(\mathbf{Z}, \mathbf{U})$ *is also compatible with the original vector* (X, Y)*.*

Proof. See [12, pp. 550–551, Prop. 4]. □

First of all, the theorem explains why jittering and JPIT are interesting for us in the context of copulae. As we have said, a naive attempt to use the probability integral transform to derive the associated copulae in the spirit of Section 1.1 fails when applied to arbitrary random vectors. Jittering allows us to generalise the probability integral transform to the case when atoms are present in the marginal distributions; Theorem 28 then states that JPIT indeed allows us to reconstruct associated copulae pretty much in the same fashion we did it in Section 1.1. Moreover, it allows us to describe the set C_H from another point of view: dependence structures of the jitter $\mathbf{U}$ correspond to the elements of C_H and provide a stochastic representation for them.

The copula corresponding to a jitter with independent elements is of particular interest. The following subsection is focused on exploring its properties.

3.4.2 Standard extension copula

Consider an arbitrary 2-CDF H and let F and G be the marginal distribution functions. Recall that Sklar's theorem (Th. 24) guarantees there exists at least one copula compatible with H and it is uniquely defined on the closure of $\operatorname{Ran} F \times \operatorname{Ran} G$; denote that uniquely defined restriction as C'_H. The elements of the set of all compatible copulae C_H may then be perceived as extensions of C'_H. Among other extensions, the so-called *standard extension copula*, also known as *bilinear interpolation copula*, plays a special role. It was initially used to prove Sklar's theorem, see [15, pp.46–48], yet, as we will see, there are further reasons, beside historical, why it is of interest.

Definition 30 (Standard extension copula)**.** Consider an arbitrary random vector (X, Y), let H be its joint CDF, and let F and G be the corresponding marginal

distribution functions. Let C' be the uniquely defined restriction of the associated copulae to the set $\overline{\operatorname{Ran} F} \times \overline{\operatorname{Ran} G}$. Consider a point $(u, v) \in [0; 1]^2$. Put

$$\begin{aligned} a_1 &:= \max\{x \in \overline{\operatorname{Ran} F} : x \leq u\}, & b_1 &:= \max\{y \in \overline{\operatorname{Ran} G} : y \leq v\}; \\ a_2 &:= \min\{x \in \overline{\operatorname{Ran} F} : x \geq u\}, & b_2 &:= \min\{y \in \overline{\operatorname{Ran} G} : y \geq v\}. \end{aligned} \tag{3.24}$$

Define

$$\lambda(a) := \begin{cases} \frac{u-a_1}{a_2-a_1}, & a_1 < a_2, \\ 1, & a_1 = a_2; \end{cases} \qquad \mu(b) := \begin{cases} \frac{v-b_1}{b_2-b_1}, & b_1 < b_2, \\ 1, & b_1 = b_2. \end{cases} \tag{3.25}$$

The *standard extension copula* is then defined as follows:

$$\begin{aligned} C_H^s(u, v) := (1 - \lambda(\check{u}))\,(1 - \mu(\check{v}))\,C'(a_1, b_1) + (1 - \lambda(\check{u}))\,\mu(\check{v})\,C'(a_1, b_2) + \\ \lambda(\check{u})\,(1 - \mu(\check{v}))\,C'(a_2, b_1) + \lambda(\check{u})\,\mu(\check{v})\,C'(a_2, b_2), \end{aligned} \tag{3.26}$$

where $\check{x}$ denotes the collar operator (Def. 7).

The standard extension copula C_H^s is defined uniquely and is indeed a copula compatible with H, for proof see [1, pp.19–21, Lemma 2.3.5]. For a deeper understanding of the mechanics inside C_H^s, see, for instance, [10, pp. 213–215]. See also example in Appendix B.1 which presents a step-by-step calculation of the standard extension copula for a random vector whose marginals have Bernoulli distributions.

The following theorem establishes the stochastic representation for the standard extension copula using JPIT.

Theorem 29 (Standard extension copula corresponds to independent jitters)**.** *Let $\mathbf{Z} := (X, Y)$ be an arbitrary random vector. Consider a (jittering) random vector $\mathbf{U} := (U_1, U_2)$ such that $U_1, U_2 \sim Uni(0; 1)$ and $U_1 \perp\!\!\!\perp U_2$. Assume $\mathbf{Z} \perp\!\!\!\perp \mathbf{U}$. Let $C_{X,Y}^s$ be the standard extension copula corresponding to the random vector (X, Y). Let C_Ψ be the unique copula associated with the transformed random vector $\Psi(\mathbf{Z}, \mathbf{U})$.*
Then $C_{X,Y}^s = C_\Psi$.

Proof. See [12, pp. 550–551, Prop. 4]. □

In terms of capturing the dependence structure, in [12] Nešlehová establishes several arguments in favour of considering the standard extension copula C_H^s as a reasonable equivalent for the unique associated copula C_H that was available in the case of non-atomic marginals. For instance, C_H^s is invariant under *continuous* strictly increasing transformations of the marginals, compare with Theorem 10.

Theorem 30 (Invariance of C^s under continuous increasing transformations)**.** *Consider an arbitrary random vector (X, Y) and its standard extension copula C^s. Let f and g be continuous strictly increasing transformations defined on $Ran X$ and $Ran Y$, respectively.*
Then C^s is also the standard extension copula of the vector $(f(X), g(Y))$.

Proof. See [12, pp. 551, Cor. 6] □

The standard extension copula captures independence.

Theorem 31 (C^s captures independence)**.** *Consider an arbitrary random vector* $\mathbf{Z} = (X, Y)$ *and let* C^s *be the corresponding standard extension copula. Then it holds that* $X \perp\!\!\!\perp Y \iff C^s = \Pi$.

Proof.
$\Longrightarrow$ Assume $X \perp\!\!\!\perp Y$. Let $\mathbf{U} \sim Uni(0;1)^2$. Then from Theorem 29 we obtain that C^s is the unique copula associated with the random vector $\mathbf{V} := \Psi(\mathbf{Z}, \mathbf{U})$. Yet the elements of $\mathbf{V}$ are independent, hence $C^s = \Pi$.

$\Longleftarrow$ Assume $C^s = \Pi$. Let H be the joint CDF of (X, Y), let F and G be the marginal distribution functions. From Theorems 28 and 29 we know that C^s is associated with H. From Sklar's theorem (Th. 24) we then get:

$$H(x, y) = C^s(F(x), G(y)) = \Pi(F(x), G(y)) = F(x) \cdot G(y). \tag{3.27}$$

It then follows that X and Y are independent. □

In case of discrete distributions, the standard extension copula also preserves the more sophisticated dependency concepts introduced in Definitions 27 and 17. Compare the following theorem with Theorem 23.

Theorem 32. *Consider a random vector* (X, Y) *with values in* $\mathbb{N}^2$. *Let* C^s *be its standard extension copula. Then for* $i \in \{1, \ldots, 5\}$ *the following holds:*

$$DEP_i(X, Y) \iff DEP_i(C^s). \tag{3.28}$$

Proof. See [7, p. 497, Prop. 11]. □

As we will see in Chapter 4, the standard extension copula preserves the concordance partial order (Th. 39), concordance function (Th. 41), and Spearman's ρ and Kendall's τ (Th. 43).

In light of Theorem 29, it is worth observing that all properties preserved or captured by the standard extension copula C^s are also preserved by the JPIT with independent jitters.

It is important to stress that the standard extension copula may not capture all dependency concepts or indices that are reflected entirely by the unique associated copula in case of the non-atomic marginals. For instance, if there are atoms present in the marginals, the standard extension copula C^s cannot be equal to either of the FH-bounds. This is what Nešlehová [12, p. 554] writes: *'Another issue is that though the standard extension copula coincides with the independence copula if the marginals are independent, it is always different from the Fréchet–Hoeffding bounds, even if the marginals are countermonotonic and comonotonic, respectively (i.e. when the upper, respectively lower, Fréchet bound is a possible copula of* $\mathbf{X}$*). This is due to the fact that as soon as the closure of the product of the ranges of the marginal distribution functions does not fill out the entire unit square, the standard extension copula cannot be singular. Although the standard extension copula is bounded from below and above by standard extension copulas corresponding to the perfect monotonic case, these bounds are not simply related'*

Another example is the so-called *Joe's tail dependency coefficients*. As Genest and Nešlehová [7, p. 497] point out, it is not entirely clear whether these indices are preserved by C^s. Same goes for the so-called *monotone regression dependence*, DEP_2 and DEP_3 (Def. 27), see [7, p. 498].

3.5 Synthetic Models and Copula Adaptations

Up till now we focused our attention on the *analytical* copula models. In other words, we considered a pre-specified arbitrary joint cumulative distribution H and explored ways of analysing the dependence structure of H using its associated copulae. In this section, we shall briefly explore *synthetic* models. Specifically, we shall explore how copulae can be used to find $H \in \Gamma(F, G)$ fulfilling some required dependency properties (e.g., specified ρ, τ or having one of DEP_i properties from Def. 27) when F and G are arbitrary CDFs.

Non-atomic marginals

Assume both F and G are non-atomic. Then in light of Sklar's theorem (Th. 9) and Chapter 3 the task is almost trivial. As we know, in this setting, Sklar's theorem ensures there exists a bijection between the set of all copulae and the elements of $\Gamma(F, G)$. Besides that, Theorem 10 guarantees that any property invariant under strictly increasing transformations is captured by copulae. Let us formulate this as a theorem (bear Remark 11 in mind).

Theorem 33. *Let F and G be non-atomic distribution functions. Consider a copula C and put:*

$$H(x, y) := C(F(x), G(y)), \quad x, y \in \mathbb{R}. \tag{3.29}$$

Consider a property A such that $A(X, Y)$ holds if and only if $A(f(X), g(Y))$ holds, where f and g are strictly increasing. Then we get:

$$A(C) \iff A(H). \tag{3.30}$$

Proof. Follows from Theorems 10, 56 and 57. □

Constructing $H \in \Gamma(F, G)$ such that $A(H)$ holds is then reduced to finding copulae for which A holds. Consider the following example. Suppose that one seeks to construct $H \in \Gamma(F, G)$ such that $\tau(H) = 0.42$. We then get:

$$\{H \in \Gamma(F, G) : \tau(H) = 0.42\} = \{C(F, G) : \tau(C) = 0.42,\ C \textit{ is a copula}\}; \tag{3.31}$$

a particular choice of C may be governed by further constraints and boundary conditions.

Arbitrary marginals

Now let F and G be arbitrary. First of all, general Sklar's theorem (Th. 24) ensures that $H(x, y) := C(F(x), G(y))$ is a valid 2-CDF whenever C is a copula. Moreover, it still holds that $H \in \Gamma(F, G)$. The problem is that the correspondence between $\Gamma(F, G)$ and the set of all copulae is no longer one-to-one; several copulae can yield the same joint distribution function H. This is the reason why generally $\tau(C) \neq \tau(H)$, same goes for ρ. Some properties though are still transferred from the copula to the constructed H as they should.

Theorem 34. *Let C be a copula. Assume F and G are arbitrary distribution functions. For $x, y \in \mathbb{R}$ put $H(x, y) := C(F(x), G(y))$. Then for $i \in \{1, \ldots, 5\}$ the following implication holds:*

$$DEP_i(C) \implies DEP_i(H), \tag{3.32}$$

where DEP_i are from Definitions 17 and 27.

Proof. See [7, p. 497, Prop. 11]. □

The theorem may be interpreted as follows; although the DEP_i properties of copula C are maintained, the 'synthesised' joint CDF H may have some additional properties as a result of the associated copula no longer being unique.

Copula adaptations

In light of the facts that we explored in Section 3.4 and will study in depth in Sections 4.1–4.4, it appears that it is not the copula C that is of primary interest, but rather the standard extension copula corresponding to it.

Definition 31 (Copula adaptation to $\Gamma(F, G)$)**.**
Consider arbitrary distribution functions F and G. Let C be a copula. Let C' be a restriction of C to the set $\overline{\operatorname{Ran} F} \times \overline{\operatorname{Ran} G}$. Denote the standard extension copula (Def. 30) corresponding to C' as $C^s_{F,G}$. We then call $C^s_{F,G}$ the *adaptation of C to the Fréchet class* $\Gamma(F, G)$.

Observe that Theorem 29 ensures that any copula adaptation is a copula in the sense of Definition 5.

When working with synthetic models for the Fréchet class $\Gamma(F, G)$, it is in fact the adaptations to $\Gamma(F, G)$ that are of interest. Moreover, as a result of adaptation to $\Gamma(F, G)$ copula C may obtain some further properties (see Th. 34). The new definition also allows us to reformulate Nešlehová's words we quoted in the previous section. See also example in Appendix B.1 demonstrating what W^s_{FG} may look like.

Theorem 35. *Consider arbitrary distribution functions F and G. The following statements are equivalent:*

1. *F and G are non-atomic.*
2. *$W^s_{F,G} = W$.*
3. *$M^s_{F,G} = M$.*

Proof. Follows from Theorem 29 and Definition 28. □

Theorem 36 (FH-bounds for copula adaptations)**.** *Let F and G be arbitrary distribution functions. Then for any copula C it holds that:*

$$W^s_{F,G} \leq C^s_{F,G} \leq M^s_{F,G}. \tag{3.33}$$

Theorem 37. *Let C be a copula. Let M' and C' be the restrictions to the set $\overline{Ran\,F} \times \overline{Ran\,G}$. Recall the standard extension copula definition (Def.30):*

$$C^s_{F,G}(u,v) := (1-\lambda(\check{u}))\,(1-\mu(\check{v}))\,C'(a_1,b_1) + (1-\lambda(\check{u}))\,\mu(\check{v})\,C'(a_1,b_2) + \lambda(\check{u})\,(1-\mu(\check{v}))\,C'(a_2,b_1) + \lambda(\check{u})\,\mu(\check{v})\,C'(a_2,b_2), \quad (3.34)$$

where the coefficients λ and μ are determined by the Fréchet class $\Gamma(F,G)$.

Compare the sums for $C^s_{F,G}$ and $M^s_{F,G}$. The coefficients in each summand are non-negative and same in both sums. At the same time, Theorem 7 guarantees that $C' \leq M'$. It then follows that $C^s_{F,G} \leq M^s_{F,G}$. Proof for $W^s_{F,G}$ is similar.

Co-monotonicities

Recall that under the assumption of non-atomic marginals, the notions of the strong (Definition 13) and weak (Def. 12) co-monotonicities coincided and both of them were induced by the upper FH-bound M (Thm. 11). The situation changes when we consider arbitrary distributions.

When arbitrary distributions are considered, the notions of weak and strong co-monotonicity do not coincide anymore, see examples in App. B.2 and B.1 . As we can see, the upper FH-bound M now generally corresponds to the weak co-monotonicity, rather than strong. The following theorem formalises that result, see also [7, p. 489, Ex. 6 and Prop. 7].

Theorem 38 (Upper FH-bound and weak co-monotonicity)**.**
Consider an arbitrary random vector (X,Y). Let F and G be the corresponding marginal distributions. Then M is compatible with (X,Y) in the sense of Definition 10 if and only if there exists a random variable $U \sim Uni(0;1)$ such that $X = F^{-1}(U)$ a.s. as well as $Y = G^{-1}(U)$ a.s., where F^{-1} and G^{-1} are the respective quantile functions (Def. 45).

Proof. See [17, pp. 192–193, Th. 2]. □

4 Measures of Concordance for Arbitrary Distributions

4.1 Partial Concordance Order

In this section, we generalise the idea of partial concordance order we introduced in Chapter 2 for the case of arbitrary distributions. Consider an arbitrary random vector (X, Y). Notice that Definition 15 introducing concordance partial order on the Fréchet class $\Gamma(X, Y)$ is valid for arbitrary distributions. Moreover, Theorem 13 establishing the relation between the concordance partial order and the corresponding joint CDFs remains valid as well. As we have stated in Section 2.1, this setting is rich enough for all of the constructions we are interested in.

Extending the partial concordance order so that we can compare elements of different Fréchet classes poses more challenges. Recall that in the case of non-atomic marginals we required that the order remains unchanged under strictly increasing transformations of the marginals. This led us to relying on the unique associated copulae to generalise the concordance partial order to all random vectors with non-atomic marginals, see Definition 16. Since in the case of arbitrary distributions the associated copula is not defined uniquely anymore, we cannot apply the same approach directly.

One approach to the generalisation was suggested by Scarsini in his original paper [10, pp. 212–213, Def. 6], yet it is rather cumbersome. Following Nešlehová [12] and Genest [7], in this chapter we limit ourselves to always working with some given Fréchet class $\Gamma(F, G)$. Luckily, in this setting the concordance partial order can be again reduced to studying copulae. Compare the following theorem with Definition 16.

Theorem 39 (C^s and concordance partial order)**.**
Consider an arbitrary random vector $\mathbf{Z} = (X, Y)$. Let $\mathbf{Z}^* = (X^*, Y^*) \in \Gamma(X, Y)$. Let $C^s_{\mathbf{Z}}$ and $C^s_{\mathbf{Z}^*}$ be the corresponding standard extension copulae. Then the following statements are equivalent:

1. $(X, Y) \succeq (X^*, Y^*)$;
2. $\forall u, v \in \mathbb{R} : C^s_{\mathbf{Z}}(u, v) \geq C^s_{\mathbf{Z}^*}(u, v)$.

Proof. See [12, p. 553, Cor. 6.1]. □

Notice that in light of Theorem 13, the theorem can be reformulated as

$$\mathbf{Z} \succeq \mathbf{Z}^* \iff C^s_{\mathbf{Z}} \succeq C^s_{\mathbf{Z}^*}. \tag{4.1}$$

To sum up, for the arbitrary distributions, the concordance partial order works more or less in the same way as for the distributions with non-atomic marginals; the role of the unique copula is taken by the unique standard extension copula. The only major difference is that we have to limit ourselves to studying the elements of the same Fréchet class, which is not a significant limitation.

4.2 Concordance Function

In this section, we explore how the ideas we introduced in Section 2.3 generalise to the arbitrary two-variate distributions. Recall that both the definition of a reference random vector (Def. 19) as well as the probabilistic definition of the concordance function (Def. 20) are general and hold for arbitrary random vectors. It is worth observing though that the concordance function does not reflect the possible ties in any way, it only accounts for the cases of strictly concordant and strictly discordant pairs; this will have some consequences later on.

As in the previous section, the problem arises when we try to apply the analytical expressions relying on the associated copula being unique, specifically Theorem 15. In the case of arbitrary distributions, the associated copula generally is not unique and evaluating Q at different elements of C_H may yield different values, see [12, pp. 547–548, Ex. 2]. Fortunately, the standard extension copula solves that issue.

Theorem 40 (Some JPIT preserve concordance function)**.**
Consider $\mathbf{Z}$ *and* $\mathbf{U}$ *as in Theorem 29. Now let* $(\mathbf{Z}^*, \mathbf{U}^*)$ *be an independent copy of the four-element random vector* $(\mathbf{Z}, \mathbf{U})$*. The following holds:*

$$Q_{\mathbf{Z}^*}(\mathbf{Z}) = Q_{\Psi(\mathbf{Z}^*, \mathbf{U}^*)}\left(\Psi(\mathbf{Z}, \mathbf{U})\right). \tag{4.2}$$

Proof. See [12, p. 552, Th. 5]. □

The theorem formulation seems convoluted, so it is worth elaborating on what is going on. As in Definition 20, we have an arbitrary gauged random vector $\mathbf{Z} = (X, Y)$ and a reference $\mathbf{Z}^* = (X^*, Y^*) \in \Gamma(X, Y)$ for which we want to evaluate the concordance function. We now want to apply the 'same' jittered probability integral transform (Def. 29) to both gauged and reference vector; the jitters that we want to use for the JPIT are the 'independent jitters' from Theorem 29. So, for the gauged vector we prepare the jitter $\mathbf{U}$ whose elements are independent with $Uni(0; 1)$ distribution, and the jitter itself is independent from both $\mathbf{Z}$ and $\mathbf{Z}^*$. For the reference vector, we prepare a jitter $\mathbf{U}^*$, which is a copy of $\mathbf{U}$ independent from all of the aforementioned objects.

Theorem 40 is formulated in terms of random variables and their transformations. We are however much more interested in its analytical consequences. The following two theorems present *analytical* as well as *stochastic* representations for the concordance function generalising Theorems 15 and 16. Proofs for both theorems are presented in [12, p. 552, Th. 5].

Theorem 41 (Standard extension copula captures concordance function)**.**
Consider an arbitrary gauged random vector $\mathbf{Z} = (X, Y)$ *and a reference vector* $\mathbf{Z}^* = (X^*, Y^*)$*. Let* H *and* H^* *be the corresponding joint CDFs. Let* C^s *and* C^{s*} *be the corresponding standard extension copulae. Then the following chain of equalities holds.*

$$Q_{X^*,Y^*}(X, Y) = Q_{C^{s*}}(C^s) = 4 \int_{\mathbb{R}^2} C^s(u, v)\, dC^{s*}(u, v) - 1. \tag{4.3}$$

Theorem 42 (General stochastic representation of concordance function)**.** *Assume the setting of Theorem 41 holds. Then we get:*

$$Q_{X^*,Y^*}(X,Y) = \mathsf{E}[H(X^*,Y^*) + H(X^*-,Y^*)] + \mathsf{E}[H(X^*,Y^*-) + H(X^*-,Y^*-)] - 1. \quad (4.4)$$

Notice that Theorems 40–42 indeed do generalise Theorems 15 and 16. If we assume that H has non-atomic marginals, the expressions will be reduced to those in Th. 15 and 16.

4.3 Generalised Scarsini Measures

It is time now to reconsider Scarsini axioms. Recall that the abstract measures of concordance were initially introduced as an auxiliary tool for extending the concordance partial ordering to a complete order. The following 'updated' list of axioms for Scarsini measures of concordance defined for arbitrary distributions is suggested by Nešlehová [12, p. 554]. The differences with Def. 18 are typeset in bold font.

Definition 32 (Generalised Scarsini measure of concordance)**.** Consider a mapping $\kappa : \Gamma \to \mathbb{R}$ where Γ stands for the set of **arbitrary** two-variate random vectors. We say κ is a *(general) measure of concordance* if and only if the following holds:

1. *(Symmetry)* $\kappa(X,Y) = \kappa(Y,X)$.
2. *(Bounds)* $\kappa(X,Y) \in [-1;1]$.
3. *(Normalisation)* If $X = f(Y)$ a.s. with f strictly increasing and **continuous** on the range of Y, then $\kappa(X,Y) = 1$. If $X = f(Y)$ a.s. with f strictly decreasing and continuous on the range of Y, then $\kappa(X,Y) = -1$.
4. *(Independence)* If X and Y are independent, then $\kappa(X,Y) = 0$.
5. *(Change of sign)* If f is strictly monotone and **continuous** on the range of X, then
$$\kappa(f(X),Y) = \begin{cases} \kappa(X,Y), & \text{for } f \text{ strictly increasing,} \\ -\kappa(X,Y), & \text{for } f \text{ strictly decreasing.} \end{cases}$$
6. *(Continuity)* Consider a sequence of random vectors $(X_n,Y_n) \in \Gamma$ such that $(X_n,Y_n) \xrightarrow{D} (X,Y) \in \Gamma$. Then $\kappa(X_n,Y_n) \underset{n\to\infty}{\longrightarrow} \kappa(X,Y)$.
7. *(Coherence)* For $(X,Y) \succeq (X^*,Y^*)$ it holds that $\kappa(X,Y) \geq \kappa(X^*,Y^*)$.

The reason for this update is the following. Theory behind Scarsini measures of concordance introduced in Definition 18 relies heavily on the classic copula theory. Recall that under the non-atomic marginals assumption, Theorem 10 ensured that the unique associated copula is invariant. Hence the *Normalisation* and *Change of sign* axioms were formulated under the assumption of strictly

monotone transformations. The situation changes when we consider arbitrary distributions. In that setting, the standard extension copula C^s is the counterpart of the unique associated copula we worked with under the non-atomic marginals assumption. Theorem 30 however guarantees that C^s is invariant under strictly increasing *continuous* transformations. Hence the change in axioms.

As we will see, the axiom of *Continuity* is somewhat problematic and questionable, it is rare that a concordance index fulfils it.

It is also in place to update the definition of strong co-monotonicity (Def. 13).

Definition 33 (Continuous strong co-monotonicity)**.**
Consider an arbitrary random vector (X, Y). Let f be a transformation continuous and strictly increasing on $\operatorname{Ran} Y$. We say that X and Y are *continuously strong co-monotonic* if and only if $X = f(Y)$ almost surely.

The notion of continuous strong counter-monotonicity is then introduced in the similar way.

The updated axiom of *Normalisation* thus requires that κ ascribes its boundary values ± 1 to the random vectors whose elements are continuously strongly co-/counter-monotonic.

We are now going to explore whether there exist any mappings that fulfil the general Scarsini axioms. In particular, we shall study the behaviour of Spearman's ρ and Kendall's τ on the class of arbitrary two-variate distributions.

4.4 Spearman's ρ and Kendall's τ

In this section, we use the facts established above to explore the behaviour of Spearman's ρ and Kendall's τ in the case of arbitrary distributions. First of all, notice that the probabilistic versions of these two indices (Defs. 21 and 23) are well-defined for arbitrary random vectors. As we have already stated, the issue arises when it comes to the analytical expressions (Th. 19 and 21) since they were derived under the assumption of marginals being non-atomic. These theorems however are special instances of Theorem 15, which was generalised in the previous section by Theorem 41.

We thus obtain that the standard extension copula properly captures Spearman's ρ and Kendall's τ. In other words, the probabilistic definitions (Defs. 21 and 23) coincide with the analytical expressions evaluated for C^s.

Theorem 43 (C^s captures Spearman's ρ and Kendall's τ)**.** *Consider an arbitrary random vector (X, Y) and its standard extension copula C^s. The following holds:*

$$\rho(X, Y) = \rho(C^s) = 12 \int_{[0;1]^2} C^s(u, v)\, du\, dv - 3, \tag{4.5}$$

$$\tau(X, Y) = \tau(C^s) = 4 \int_{\mathbb{R}^2} C^s(u, v)\, dC^s(u, v) - 1. \tag{4.6}$$

Proof. A direct implication of Theorem 41 as well as Definitions 21 and 23. □

Observe that in case of τ we may also integrate over the unit square $[0; 1]^2$ since C^s assign no mass to the points outside of it. On the other hand, integration over $\mathbb{R}^2$ highlights the fact that the integral is in fact expectation E with respect to the distribution induced by C^s.

Using Theorem 42, we also get the generalised stochastic representations for Spearman's ρ and Kendall's τ, see [7, p. 491, Prop. 9]; compare with Ths. 18 and 20.

Theorem 44 (General stochastic representation for ρ and τ)**.**
Consider an arbitrary random vector (X, Y). Let H be the corresponding joint distribution function. Let (X^*, Y^*) be a reference vector such that $X^* \perp\!\!\!\perp Y^*$. Then the following holds:

$$\rho(X,Y) = 3\ \mathsf{E}[H(X^*,Y^*) + H(X^*-,Y^*)] + \mathsf{E}[H(X^*,Y^*-) + H(X^*-,Y^*-)] - 3, \quad (4.7)$$

$$\tau(X,Y) = \mathsf{E}[H(X,Y) + H(X-,Y)] + \mathsf{E}[H(X,Y-) + H(X-,Y-)] - 1. \quad (4.8)$$

See example in Appendix B.2 demonstrating the usage of these formulae.

4.4.1 Normalisation issues

Seemingly, the story unfolds positively; up till now, we were able to generalise the ideas presented in Chapter 2 to the case of arbitrary distributions. Recall that the constructions we scrutinised in Chapter 2 reached its peak in Theorem 22 which demonstrated that, for the class of distributions with non-atomic marginals, Spearman's ρ and Kendall's τ indeed are measures of concordance in the sense of Definition 18. Do Scarsini axioms hold for these indices if we consider arbitrary distributions?

Alas, the answer is no, they do not. In particular, the *Normalisation* axiom does not hold anymore. As Nešlehová [12, pp. 551–551, Ex. 2] shows, in case of atoms being present in the marginals, it may happen that $\tau(X,Y) \neq 1 \neq \rho(X,Y)$ even when X and Y are strongly co-monotonic (Def. 13); same goes for -1 and strong counter-monotonicity. Consider the following example by Genest and Nešlehová [7].

Example 5. Consider two Bernoulli random variables X and Y with the same probability of success equal to $p \in (0;1)$. Assume $\mathsf{P}[X = 0, Y = 0] = 1 - p$. Then $X = Y$ a.s., hence they are continuously strongly co-monotonic (Def. 33). Applying the stochastic representations (Th. 44), we get

$$\tau(X,Y) = \rho(X,Y) = p(1-p) < 1. \quad (4.9)$$

Assume now $\mathsf{P}[X = 0] = 1 - p = \mathsf{P}[Y = 1]$ and let $\mathsf{P}[X = 0, Y = 0] = 0$. Then $X = 1 - Y$ a.s., hence they are continuously strongly counter-monotonic. It also holds that $\tau(X,Y) = \rho(X,Y) = -p(1-p) > -1$.

See also example in Appendix B.2 demonstrating the extent of the issue in the case of weak co-monotonicity.

As we know, both Spearman's ρ and Kendall's τ are based on the concordance function (Defs. 21 and 23). Recall the concordance function's (Def. 20) structure:

$$Q_{X^*,Y^*}(X,Y) := \mathsf{P}\left[(X - X^*)(Y - Y^*) > 0\right] - \mathsf{P}[(X - X^*)(Y - Y^*) < 0]. \quad (4.10)$$

It considers only the cases of sharply concordant pairs (the minuend) and sharply discordant pairs (the subtrahend); the ties—cases when $X = X^*$ or $Y = Y^*$—are not accounted for. In case of the non-atomic marginals, it bears no consequences since they occur with probability zero. The presence of atoms in the marginals changes the situation.

This is also closely related to the fact that adapted copulae typically do not reach FH-bounds. Indeed, let F and G be distribution functions such that at least one of them is not non-atomic. Consider the adapted (Definition 31) upper FH-bound $M^s_{F,G}$. Recall that a copula adaptation is a copula (Th. 29), hence Theorem 7 ensures that $M^s_{F,G} \leq M$ and Th. 35 then ensures that $M^s_{F,G} < M$, by which we mean $\exists u, v \in [0;1]^2 : M^s_{F,G}(u,v) < M(u,v)$. Observe that both M and its adaptation are copulae, hence their marginals are non-atomic and thus Theorem 22 holds; i.e., for these two objects, both τ and ρ fulfil Scarsini axioms (Def. 18). It then could be shown that $\rho(M^s_{F,G}) < \rho(M)$.

To sum up, we have the following chain of inequalities going along the W–M axis (Sec. 2.1):

$$W < W^s_{F,G} \leq C^s_{F,G} \leq M^s_{F,G} < M, \tag{4.11}$$

where $C^s_{F,G}$ is an arbitrary adapted copula. Any Scarsini measure of concordance κ is normalised so that it ascribes the boundary values ± 1 to the poles W and M. Yet for the Fréchet class $\Gamma(F,G)$ we are working with, the actual attainable poles are $W^s_{F,G}$ and $M^s_{F,G}$, and they differ from the FH-bounds. Thus the actual values that κ can reach when working with $\Gamma(F,G)$ are narrower than required by Scarsini axioms. In order to fulfil the *Normalisation* Scarsini axiom, it is natural then to consider some re-normalisations.

4.4.2 Sharp re-normalisations

In the rest of this section, we shall use κ as a placeholder for both Spearman's ρ or Kendall's τ.

The first approach is to tailor τ or ρ to the Fréchet class at hand. A minor nuisance is that generally $|\kappa(W^s_{F,G})| \neq |\kappa(M^s_{F,G})|$, see [12, p. 555, Fig. 2]. Considering these facts, Genest and Nešlehová [7, p. 494–495, Def. 4] suggest the following generalisation. Notice that it can be interpreted as a probabilistic (i.e., based on Defs. 21 and 23), analytical (Thm. 43) or stochastic (Thm. 44) expression.

Definition 34 (Sharp re-normalisation of ρ and τ)**.** Consider arbitrary distribution functions F and G. For $H \in \Gamma(F,G)$, put:

$$\kappa_{FG}(H) := \frac{\kappa(H)}{d_{FG}(H)}, \tag{4.12}$$

where $d_{FG}(H)$ is a normalising function defined as:

$$d_{FG}(H) := \begin{cases} |\kappa(M^s_{F,G})|, & \kappa(H) \geq 0; \\ |\kappa(W^s_{F,G})|, & \kappa(H) < 0. \end{cases} \tag{4.13}$$

The following fact about the mechanics of κ_{FG} is worth noticing. Observe that Fréchet class fully specifies d_{FG}; hence, for a given $\Gamma(F,G)$, both d_{FG} and $\kappa_{FG}(H)$ are functions of $\kappa(H)$.

Observe that this re-normalisation indeed generalises κ, since under the non-atomicity assumption Theorem 35 guarantees that $\kappa_{FG}(H) = \kappa(H)$. Another nice property is that $\kappa_{FG}(H) = 1$ whenever H is weakly co-monotonic (Def. 12), which is ensured by Theorem 38.

There are, however, certain disadvantages of choosing in favour of such re-normalisation. One of them is that the sharp bounds $\kappa(W^s_{FG})$ and $\kappa(M^s_{FG})$ may not be easy to calculate analytically as Nešlehová shows [12, p. 555, Fig. 2].

Another issue stems from the fact that the normalising function jumps at zero. Consider a fixed Fréchet class $\Gamma(F, G)$. As we have already mentioned, for $H \in \Gamma(F, G)$ it holds that $\kappa_{FG}(H)$ and $d_{FG}(H)$ are functions of $k := \kappa(H)$. The problem is that generally $d_{FG}(k)$ is discontinuous at $k = 0$. Hence, although $\kappa_{FG}(k)$ is continuous, it is not differentiable at zero. It poses no problem on the population level, yet when it comes to estimating κ_{FG}, this fact may lead to instabilities, especially when $\kappa(H)$ is close to zero. Hence it may be reasonable to consider a smooth version of the normaliser d_{FG}.

Another problem is that, as Genest and Nešlehová [7, p. 495] admit, the properties of such generalisations κ_{FG} are not clear. Specifically, it is not clear whether the mappings κ_{FG} are generalised Scarsini measures of concordance in the sense of Definition 32.

4.4.3 Alternative re-normalisations

Another approach suggested by Nešlehová [12] is based on a boundary for the concordance function.

Theorem 45 (Concordance function boundary)**.** *Consider an arbitrary random vector (X, Y) and let (X^*, Y^*) be a reference vector. Assume F and G are the corresponding marginal CDFs. Put $\Delta F(x) := F(x) - F(x-) = \mathsf{P}[X = x]$. Then the following holds:*

$$|Q_{X^*,Y^*}(X, Y)| \leq \sqrt{(1 - \mathsf{E}[\Delta F(X)])\ (1 - \mathsf{E}[\Delta G(Y)])}. \tag{4.14}$$

Proof. See [12, p. 555–556, Cor. 7]. □

Kendall's τ_b

The following theorem states that if the reference vector has the same dependence structure as the gauged one (or, in other words, if we are calculating τ), the bounds are attained exactly when the random variables at hand are in continuous strong co-/counter-monotonicity.

Theorem 46. *Consider an arbitrary random vector (X, Y) and let F and G be the corresponding marginals. Assume $X = f(Y)$ for a transformation f on $\mathrm{Ran}Y$ continuous strictly monotone. Then it holds that*

$$\tau(X, Y) = \begin{cases} 1 - \mathsf{E}[\Delta F(X)] & \text{for } f \text{ increasing,} \\ -1 + \mathsf{E}[\Delta F(X)] & \text{for } f \text{ decreasing.} \end{cases} \tag{4.15}$$

Proof. See [12, p. 556, Prop. 8]. □

Relying on the boundary we have just introduced, the following generalisation of τ is suggested, see [12, p. 557, Def. 9].

Definition 35 (Kendall's τ_b)**.** Consider an arbitrary random vector (X, Y) with joint distribution function H. Let F and G be the corresponding marginals. Define Kendall's τ_b as follows:

$$\tau_b(X, Y) = \frac{\tau(X, Y)}{\sqrt{(1 - \mathsf{E}\left[\Delta F(X)\right])\ (1 - \mathsf{E}\left[\Delta G(Y)\right])}}. \tag{4.16}$$

There is a particularly nice probabilistic reformulation for τ_b which is easy to obtain, see also [7, p. 492, Def. 3].

Theorem 47 (Probabilistic formulation for τ_b)**.**
Consider an arbitrary random vector (X, Y). Let $(X^*, Y^*) \sim (X, Y)$ be a reference vector in the sense of Definition 19. It then holds that

$$\tau_b(X, Y) := \frac{\tau(X, Y)}{\sqrt{\mathsf{P}[X \neq X^*]\ \mathsf{P}[Y \neq Y^*]}}. \tag{4.17}$$

Proof. Recall that $\Delta F(x) := F(x) - F(x-) = \mathsf{P}[X = x]$. Let x_i be the atoms (i.e., the points where $\Delta F(x_i) \neq 0$). Denote $p_i := \mathsf{P}[X = x_i]$. $\Delta F(X)$ is then a discrete random variable with values in $\{p_i : i \in \mathbb{N}\} \cup \{0\}$. We then get that

$$\mathsf{E}[\Delta F(X)] = \sum_{i \in \mathbb{N}} p_i \cdot p_i = \mathsf{P}[X = X^*]. \tag{4.18}$$

It then follows that $1 - \mathsf{E}[\Delta F(X)] = \mathsf{P}[X \neq X^*]$. Repeating the similar constructions for Y, we complete the proof. □

Observe that τ_b indeed generalises Kendall's τ introduced in Def. 23: under the non-atomicity assumption, the denominator is equal to 1. Nešlehová [12, p. 557] points out that τ_b is not a Scarsini measure of concordance in the sense of Definition 32: it fulfils all axioms apart from the *Continuity* axiom.

The disadvantage of the approach compared to κ_{FG} renormalisation is that, although compensating for possible ties, τ_b is not sensitive enough to detect weak co-monotonicity, see example in Appendix B.2 and refer to [7, p. 493, Ex. 9].

Spearman's ρ

We begin by observing the following facts about Spearman's ρ.

Theorem 48. *Consider an arbitrary random vector (X, Y) with marginal distribution functions F and G. Let $(U, V) \sim Uni(0; 1)^2$. Then it holds that*

$$\rho(X, Y) = r(\psi(X, U), \psi(Y, V)), \tag{4.19}$$

where r stands for Pearson's r (Def. 1), and ψ is the JPIT (Def. 28).

Proof. From Theorem 43, we get $\rho(X, Y) = \rho(C^s_{X,Y})$. Theorem 29 then ensures that $\rho(C^s_{X,Y}) = \rho(C_\Psi)$, where C_Ψ is the unique copula (as well as the joint CDF) associated with the vector transformed with JPIT. Finally, from Equation (2.20) we get $\rho(C_\Psi) = r(C_\Psi)$. By Definition 1, $r(C_\Psi) = r(\psi(X, U), \psi(Y, V))$, which completes the proof. □

Let us explore the expression $\mathsf{E}[\psi(X,U)]$ a bit closer. From the definition of JPIT (Def. 28), we get:

$$\begin{aligned}\mathsf{E}[\psi(X,U)] &= \mathsf{E}[F(X-) + U\Delta F(X)] \\ &= \mathsf{E}[F(X-)] + E[U]\ \mathsf{E}[\Delta F(X)] \\ &= \mathsf{E}[F(X-)] + 0.5\,\mathsf{E}[F(X) - F(X-)] \\ &= \mathsf{E}\left(\frac{F(X)+F(X-)}{2}\right), \end{aligned} \tag{4.20}$$

where we used that $U \sim Uni(0;1)$ is independent of X. This leads to the following theorem.

Theorem 49. *Let X be an arbitrary random variable. Let F be the corresponding distribution function. Then it holds that:*

$$\mathsf{E}\left[\frac{F(X)+F(X-)}{2}\right] = \frac{1}{2}, \tag{4.21}$$

$$var\left[\frac{F(X)+F(X-)}{2}\right] = \frac{1}{12}\left[1 - \mathsf{E}[\Delta F(X)]^2\right]. \tag{4.22}$$

Proof. See [12, pp. 550 and 558] □

These results lead to the following conclusion:

Theorem 50. *Consider an arbitrary random vector (X,Y) with marginal distribution functions F and G. Put*

$$\overline{X} := \frac{F(X)+F(X-)}{2}, \qquad \overline{Y} := \frac{G(Y)+G(Y-)}{2}. \tag{4.23}$$

It then holds that

$$\rho(X,Y) = 12\,\mathsf{cov}(\overline{X},\overline{Y}). \tag{4.24}$$

Proof. See [12, p. 558]. □

The idea behind the re-normalisation of ρ is to put it equal to $r(\overline{X},\overline{Y})$, see [7, pp. 492–493, Def. 3] and [12, p. 558, Def. 11].

Definition 36 (Spearman's ρ_s)**.**
Consider an arbitrary random vector (X,Y) and let F and G be the corresponding marginal distribution functions. We then put

$$\rho_s(X,Y) := \frac{\rho(X,Y)}{\sqrt{\mathrm{var}[\overline{X}]\,\mathrm{var}[\overline{Y}]}} = \frac{\rho(X,Y)}{\sqrt{\left(1 - \mathsf{E}\left[\Delta F(X)\right]^2\right)\left(1 - \mathsf{E}\left[\Delta G(Y)\right]^2\right)}}, \tag{4.25}$$

where $\overline{X}$ and $\overline{Y}$ are defined as in Theorem 50.

Again, observe that ρ_s indeed generalises Spearman's ρ as introduced in Definition 21; specifically, under the non-atomic marginals assumption, the denominator turns to 1. Similarly to Kendall's τ_b, it turns out that, strictly speaking, ρ_s is not

a general Scarsini measure of concordance (Def. 32); it fulfils all of the axioms apart from the *Continuity* axiom, see [12, p. 559].

Another disadvantage is that, as in the case of Kendall's τ_b, the coefficient generally fails to reveal weak co-monotonicities, see example in Appendix B.2 and refer to [7, p. 493].

On the other hand, it may hint on the fact that the *Continuity* axiom is ill-formulated since, as Nešlehová [12, p. 554] admits, it is rare that a concordance index fulfils it.

Conclusion

In this text, we examined probabilistic copula models and their applications to measures of concordance and modelling dependence between random variables. We began by studying the bivariate distributions with non-atomic marginals. We summarised the basic facts of copula theory and demonstrated that it provides a solid framework for studying dependencies between random variables. The cornerstone of classic copula theory is the fact that the unique copula associated with a given random vector preserves all 'scale-invariant' characteristics of the vector and fully describes its dependency structure. On the one hand, this fact could be used to analyse the dependencies between the elements of a given a random vector, which corresponds to the so-called *analytical* models. In addition, it allows us to construct joint distributions having pre-specified margins and dependency properties, which is referred to as *synthetic* models. We also introduced the notions of counter- and co-monotonicity.

We subsequently applied copula theory to develop the notion of concordance ordering, which yielded the W–Π–M concordance axis. We introduced Scarsini measures of concordance, which reflect the position of a dependence structure on this axis. In particular, it allowed us to define the population versions for Kendall's τ and Spearman's ρ. Additionally, copula theory provided deeper insights into the inner mechanics and behaviour of these correlation coefficients.

We then proceeded by broadening our scope to arbitrary bivariate distributions. We highlighted the differences from classic copula models and outlined the problems that arise when considering arbitrary distributions. The major problem is that the associated copula is no longer unique. We introduced the standard extension copula, which resolves many issues arising from the presence of atoms in the marginals, effectively generalising the unique associated copula available under the non-atomicity assumption. In capturing important dependency properties, specifically τ and ρ, the standard extension copula validates the use of copulae for analytical models. Although synthetic copula models remain valid even when atoms are present in the marginal distributions, the so-called *copula adaptations* provide a more refined perspective, albeit with potential computational challenges. We also demonstrated the need to differentiate between weak and strong counter-/co-monotonicity in the case of arbitrary distributions.

Using the standard extension copula, we generalised the concordance ordering and Scarsini measures of concordance. The presence of atoms in the marginals means that while τ and ρ are well-defined for arbitrary distributions, they are no longer properly normalised and fail to capture even strong co-monotonicity adequately, not to mention the weak. Various re-normalisations seeking to adjust for this were suggested and studied. Taking into account the fact that none of them fulfilled Scarsini's axiom of Continuity, a possible re-consideration of Scarsini's axioms may be necessary.

All theoretical concepts were illustrated using simple and understandable examples.

Statistical inference for copulae and measures of concordance is out of the scope of this text. It is thus pertinent to suggest texts that cover this topic. A good introduction to copula model building, diagnostics and statistical inference

under the non-atomicity assumption is available in [6]. Texts [7] and [3] focus specifically on statistical copula models for arbitrary distributions; see also [18] and [19]. Publication [20] is dedicated to the non-parametric estimator of the associated copula known as *empirical copula*. Paper [21] illustrates the use of copulae for tests of independence for arbitrary distributions.

Overall, this thesis provides a comprehensive overview of probabilistic copula models and population measures of concordance, both under the common assumption of non-atomic marginals and for arbitrary distributions. This text could serve as a primer or introductory text for those interested in the behaviour of Kendall's τ or Spearman's ρ at the population level. While being understandable for any practitioner with the basic knowledge of statistics and probability, the text provides an important mathematical background and theory standing behind the inner mechanics of these indices. To the best of the author's knowledge, there is no publication that summarising these facts all in one place.

Bibliography

[1] Roger B. Nelsen. *An Introduction to Copulas*. 2nd. Springer series in statistics. New York: Springer, 2006. 269 pp. ISBN: 978-0-387-28659-4. URL: https://link.springer.com/book/10.1007/0-387-28678-0.

[2] Fabrizio Durante and Carlo Sempi. *Principles of copula theory*. Boca Raton, FL [u.a.]: CRC Press, 2016. ISBN: 9781439884423.

[3] John Ery. "Semiparametric Inference for Copulas with Discrete Margins". MA thesis. Ecole Polytechnique Fédérale deLausanne, 2016.

[4] Christian Genest and Johanna Nešlehová. "Copulas and Copula Models". In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014. ISBN: 9781118445112. DOI: https://doi.org/10.1002/9781118445112.stat07523. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat07523. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07523.

[5] Harry Joe. *Multivariate models and dependence concepts*. OCLC: 883391243. Boca Raton, FL: Chapman & Hall/CRC, 2001. ISBN: 978-1-4899-3105-4. DOI: 10.1007/978-1-4899-3105-4.

[6] Christian Genest and Anne-Catherine Favre. "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask". In: *Journal of Hydrologic Engineering* 12.4 (1st July 2007), pp. 347–368. ISSN: 1084-0699. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(347). URL: https://ascelibrary.org/doi/10.1061/%28ASCE%291084-0699%282007%2912%3A4%28347%29 (visited on 26/10/2022).

[7] Christian Genest and Johanna Nešlehová. "A Primer on Copulas for Count Data". In: *ASTIN Bulletin: The Journal of the IAA* 37.2 (2007). Publisher: Cambridge University Press, pp. 475–515. ISSN: 0515-0361. DOI: 10.1017/s0515036100014963.

[8] Menahem E. Yaari. "The Dual Theory of Choice under Risk". In: *Econometrica* 55.1 (1987), pp. 95–115. ISSN: 0012-9682. DOI: 10.2307/1911158.

[9] E. L. Lehmann. "Some Concepts of Dependence". In: *The Annals of Mathematical Statistics* 37.5 (Oct. 1966), pp. 1137–1153. ISSN: 0003-4851. DOI: 10.1214/aoms/1177699260.

[10] Marco Scarsini. "On measures of concordance". In: *Stochastica* 7.1-2 (1984), pp. 39–44. ISSN: 1593-8883. DOI: 10.1007/bf02106373.

[11] Fabrizio Durante, Erich Peter Klement and José Juan Quesada-Molina. "Bounds for Trivariate Copulas with Given Bivariate Marginals". In: *Journal of Inequalities and Applications* 2008.1 (2008), p. 161537. ISSN: 1029-242X. DOI: 10.1155/2008/161537.

[12] Johanna Nešlehová. "On rank correlation measures for non-continuous random variables". In: *Journal of Multivariate Analysis* 98.3 (Mar. 2007), pp. 544–567. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2005.11.007.

[13] Jan Dhaene and Marc J. Goovaerts. “Dependency of Risks and Stop-Loss Order”. In: *ASTIN Bulletin* 26.2 (1996), pp. 201–212. ISSN: 0515-0361. DOI: 10.2143/ast.26.2.563219.

[14] M. Sklar. “Fonctions de répartition à N dimensions et leurs marges”. In: *Annales de l’ISUP* VIII.3 (1959), pp. 229–231. URL: https://hal.science/hal-04094463.

[15] B. Schweizer and A. Sklar. “Operations on distribution functions not derivable from operations on random variables”. eng. In: *Studia Mathematica* 52.1 (1974), pp. 43–52. URL: http://eudml.org/doc/217948.

[16] Holly Carley. “Maximum and Minimum Extensions of Finite Subcopulas”. In: *Communications in Statistics - Theory and Methods* 31.12 (Dec. 2002), pp. 2151–2166. ISSN: 1532-415X. DOI: 10.1081/sta-120017218.

[17] Paul Embrechts, Alexander J. McNeil and Daniel Straumann. “Correlation and Dependence in Risk Management: Properties and Pitfalls”. In: *Risk Management*. Cambridge University Press, Jan. 2002, pp. 176–223. DOI: 10.1017/cbo9780511615337.008.

[18] Ivan Kojadinovic and Tommaso Martini. “Copula-like inference for discrete bivariate distributions with rectangular supports”. In: *Electronic Journal of Statistics* 18.1 (Jan. 2024). ISSN: 1935-7524. DOI: 10.1214/24-ejs2261.

[19] Bouchra R. Nasri and Bruno N. Remillard. *Identifiability and inference for copula-based semiparametric models for random vectors with arbitrary marginal distributions*. 2023. DOI: 10.48550/ARXIV.2301.13408.

[20] Christian Genest, Johanna G. Nešlehová and Bruno Rémillard. “Asymptotic behavior of the empirical multilinear copula process under broad conditions”. In: *Journal of Multivariate Analysis* 159 (July 2017), pp. 82–110. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2017.04.002.

[21] C. Genest et al. “Testing for independence in arbitrary distributions”. In: *Biometrika* 106.1 (Jan. 2019), pp. 47–68. DOI: 10.1093/biomet/asy059.

[22] Jun Shao. *Mathematical statistics*. Springer, 2003, p. 591. ISBN: 0387953825.

A Recapitulating Basics of Probability Theory and Mathematical Statistics

In this section, we briefly summarise the basic definitions, facts, and conventions from probability theory and mathematical statistics that are used throughout in this text.

Definition 37 (Random element)**.** Consider a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space $(A, \mathcal{A})$. A measurable mapping $\mathbf{X} : (\Omega, \mathcal{F}) \to (A, \mathcal{A})$ is a *random element*.

Definition 38 (Distribution of a random element)**.** Let $\mathbf{X}$ be a random element. Consider the measure $P_{\mathbf{X}}$ defined on the target measurable space of $\mathbf{X}$ as follows:

$$P_{\mathbf{X}}(A) := P\left(\mathbf{X}^{-1}[A]\right) := P\left(\{\omega \in \Omega : \mathbf{X}(\omega) \in A\}\right), \quad A \in \mathcal{A}.$$

We call the measure $P_{\mathbf{X}}$ the *distribution of the random element* $\mathbf{X}$, and we say that $P_{\mathbf{X}}$ is induced on $(A, \mathcal{A})$ by the random element (or mapping) $\mathbf{X}$.

Definition 39 (Discrete distributions and random elements)**.** Let $B \in \mathcal{A}$ be countable. Let $P_{\mathbf{X}}(B) = 1$. Then we say that $P_{\mathbf{X}}$ is a *discrete* distribution, and $\mathbf{X}$ is a discrete random element.

Definition 40 (Non-atomic distributions and random elements)**.** We say that distribution $P_{\mathbf{X}}$ is *non-atomic* if and only if $\forall a \in A : P_{\mathbf{X}}(\{a\}) = 0$.
We say that random element $\mathbf{X}$ is non-atomic if its distribution is non-atomic.

Remark 5. In literature, non-atomic distributions are often referred to as *continuous*, e.g., [12, p. 545]. However, the term is ambiguous. On the one hand, it is in the opposition to the *discrete* distributions, implying that continuous distributions are those without atoms, which is in line with the notion of non-atomicity and with how it is used by, e.g., Nešlehová [12, p. 545]. On the other hand, it is often used to refer to the distributions (absolutely) continuous with respect to Lebesgue measure. The term *non-atomic* is thus preferred.

Definition 41 (Random variable)**.** Let $\mathcal{B}$ be the Borel σ-algebra of real numbers $\mathbb{R}$. Consider a random element X whose target space is $(\mathbb{R}, \mathcal{B})$. Such element is an element *real random variable*.

Remark 6. To denote random variables, we will mainly use capital letters R, S, U, V, X and Y, possibly with indices, typeset in plain font.

Definition 42 (Random vector)**.** Consider a random element $\mathbf{X}$, whose target space is $(\mathbb{R}^n, \mathcal{B}^n)$. Such an element is a *real random vector*.

Remark 7. We are mainly interested in two-element random vectors and we will usually define them by elements: simply listing the random variables constituting the elements of the vector. Sometimes we will also use bold letters to refer to a random vector. For instance, $\mathbf{U}$, $\mathbf{Z} := (X, Y)$ or (X_1, X_2) are random vectors.

Remark 8. We call real random variables and vectors simply *random variables* or *random vectors*. We also assume that all random variables and vectors are defined on the same probability space.

Definition 43 (Cumulative distribution function of a random variable)**.** Consider a random variable X. Function $F_X : \mathbb{R} \to [0; 1]$ defined as $F_X(x) := \mathsf{P}[X \leq x]$ is called the *cumulative distribution function (CDF) of* X.

Definition 44 (Univariate cumulative distribution function)**.**
We call function $F : \mathbb{R} \to [0; 1]$ a *cumulative distribution function* if and only if the following holds:

1. F is non-decreasing;
2. F is right-continuous;
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

Theorem 51 (Characterization of univariate CDFs)**.**

1. *Consider a univariate cumulative distribution function F (as in Definition 44). Then there exists a random variable X such that $F_X = F$.*
2. *Consider a random variable X. Its CDF F_X is a univariate cumulative distribution function, in other words it meets the requirements set it Definition 44.*

Proof. See [22, pp. 8–9]. □

Definition 45 (Quantile function)**.** Let F be a univariate CDF. The *generalised inverse* or *quantile function* of F is a mapping $F^{-1} : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ defined as

$$F^{-1}(u) := \inf\{x : F(x) \geq \breve{u}\},$$

where $\breve{u}$ stands for the collar operator (Def. 7).

Definition 46 (Joint cumulative distribution function of a random vector)**.** Consider a random n-vector $\mathbf{X}$ consisting of the elements $(X_1, \ldots, X_n)$. *Joint cumulative distribution function of random vector* $\mathbf{X}$ is a function $H_{\mathbf{X}} : \mathbb{R}^n \to [0; 1]$ defined as follows:

$$H_{\mathbf{X}}(x_1, \ldots, x_n) := \mathsf{P}[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

Remark 9. For joint CDFs, we will typically use capital letters C or H typeset in plain font, while letters F and G are reserved for CDFs of random variables.
In this work, we are mainly interested in random vectors consisting of two elements, so the random variables constituting the vector at hand will be used as the index. For instance, the joint CDF of random vector (X, Y) will be typically referred to as $H_{X,Y}$, while its elements' CDFs are F_X and G_Y, respectively.

Definition 47 (Two-variate cumulative distribution function)**.**
Let $H : \mathbb{R}^2 \to [0; 1]$ be a function. We say that H is a *two-variate cumulative distribution function* if and only if the following holds:

1. $\lim_{x,y\to+\infty} H(x,y) = 1$;
2. $\forall x, y \in \mathbb{R} : \lim_{x\to-\infty} H(x,\cdot) = 0 = \lim_{y\to-\infty} H(\cdot,y)$;
3. $\forall x, y \in \mathbb{R} : H(x,\cdot)$ and $H(\cdot,y)$ are right-continuous;
4. $\forall\ x_1 < x_2,\ y_1 < y_2 : H(x_2,y_2) - H(x_1,y_2) - H(x_2,y_1) + H(x_1,y_1) \geq 0$.

Remark 10. Property 4 from Definition 47 is sometimes called *rectangle inequality* (see [5, p. 11]). Functions fulfilling property 4 are sometimes called *two-increasing*, see [1, Section 2.1] for more information on two-increasing functions. Rectangle inequality can be generalised to n-variate functions; see, for instance, [5, p. 12] and [1, pp. 43–44, Definition 2.10.2].
Essentially, the property of being two-increasing stems from the fact, that we want to construct a measure using the function, and it is inclusion-exclusion principle combined with measure's non-negativity.

Theorem 52 (Characterization of two-variate CDFs)**.**

1. *Consider random vector (X,Y). Then its joint CDF $H_{X,Y}$ is a two-variate cumulative distribution function.*
2. *Consider a two-variate cumulative distribution function $H(x,y)$. Then there exists random vector (X,Y), such that $H_{X,Y}(x,y) = H(x,y)$.*

Proof. See [22, pp. 8–9]. □

Definition 48 (Marginals of a two-variate random vector)**.** Consider a random vector (X,Y). Distributions P_X and P_Y are its *marginal distributions*, while CDFs F_X and G_Y are its *marginal distribution functions*.

Theorem 53 (Univariate CDF and distribution define each other)**.**

1. *Let F be a univariate CDF. Then there exists a unique distribution P defined on $(\mathbb{R},\mathcal{B})$, such that $\forall x \in \mathbb{R} : F(x) = P((-\infty;x])$.*
2. *Let P be a distribution defined on $(\mathbb{R},\mathcal{B})$. Then there exists a unique univariate CDF such that $\forall x \in \mathbb{R} : P((-\infty;x)]) = F(x)$.*

Theorem 54. *See [22, p. 4, Prop. 1.2].*

Theorem 55 (Two-variate CDF and distribution define each other)**.**

1. *Let H be a two-variate CDF. Then there exists a unique distribution P defined on $(\mathbb{R}^2,\mathcal{B}^2)$, such that $\forall x,y \in \mathbb{R} : H(x,y) = P((-\infty;x] \times (-\infty;y)])$.*
2. *Let P be a distribution defined on $(\mathbb{R}^2,\mathcal{B}^2)$. Then there exists a unique univariate CDF such that $\forall x,y \in \mathbb{R} : P((-\infty;x] \times (-\infty;y)]) = H(x,y)$.*

Proof. See [22, p. 6]. □

Remark 11. The previous two theorems establish a direct connection between CDFs and distribution: these object uniquely define each other. To simplify the notation, we will often use them interchangeably without introducing the notation explicitly. For instance, Pearson's r for a random vector (X, Y) with the 2-CDF H could be equivalently written as $r(X, Y)$ and $r(H)$ or even $r(P_{X,Y})$, where $P_{X,Y}$ is the corresponding distribution.

Theorem 56 (Probability integral transform of a non-atomic variable)**.** *Let X be a random variable with non-atomic distribution. Let $F_X(x)$ be its CDF. Then the following holds:*

$$F(X) \sim Uni(0; 1).$$

Proof. Let $u \in (0; 1)$. We then get

$$\mathsf{P}[F(X) \leq u] = \mathsf{P}[X \leq F^{-1}(u)] = F(F^{-1}(u)) = u. \tag{A.1}$$

□

Theorem 57 (Inverse probability integral transform)**.**
Consider a random variable $U \sim Uni(0; 1)$. Let F be a distribution function. Put $X := F^{-1}(U)$, where F^{-1} is a quantile function from Definition 45.Then it holds that $X \sim F$.

Proof. Let $x \in \mathbb{R}$. Then

$$\mathsf{P}[F^{-1}(U) \leq x] = \mathsf{P}[U \leq F(X)] = F(X). \tag{A.2}$$

□

Definition 49 (Rank of an observation)**.** Let $X_1, \ldots, X_n$ be a random sample. We define the rank of the i-th observation as follows:

$$R_i := \sum_{j=1}^{n} \mathbb{I}[X_j \leq X_i].$$

Definition 50 (Empirical CDF)**.** Let $X_1, \ldots, X_n$ be a random sample. We define the empirical cumulative distribution function (ECDF) as follows:

$$\widehat{F}_n(x) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}[X_j \leq x].$$

Immediately from Definitions 49 and 50 follows the connection between the empirical CDF and observations' ranks.

Theorem 58 (Empirical CDF and ranks)**.** *Let $X_1, \ldots, X_n$ be a random sample. Notice that for the i-th observation, the following holds from the definitions of empirical CDF and observation's rank:*

$$R_i := \sum_{j=1}^{n} \mathbb{I}[X_j \leq X_i] = n \cdot \widehat{F}_n(X_i).$$

Theorem 59 (Glivenko-Cantelli)**.** *Let $\{X_i\}_{i=1}^n$ be a random sample from a distribution with CDF F_X. Then the following holds:*

$$\|\widehat{F}_n - F_X\|_\infty = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F_X(x)| \xrightarrow[n \to +\infty]{a.s.} 0. \tag{A.3}$$

Proof. See [22, p. 320]. □

B Examples

B.1 Computing Standard Extension Copula

Consider a random variable $U \sim Uni(0;1)$. Define X and Y as follows:

$$X := \begin{cases} 0, & U \in (0;\frac{1}{3}); \\ 1, & U \in [\frac{1}{3};1); \end{cases} \qquad Y := \begin{cases} 0, & U \in (0;\frac{2}{3}); \\ 1, & U \in [\frac{2}{3};1); \end{cases} \tag{B.1}$$

Let H be the joint CDF and F and G the marginal distribution functions of the random vector (X,Y). Clearly both X and Y are Bernoulli random variables with the probabilities of success equal to 2/3 and 1/3, respectively. Tables B.1 and B.2 present the joint probability cumulative distribution functions. Notice that the upper FH-bound M (Def. 8) is compatible with the random vector (X,Y) in the sense of Definition 10. For instance, $H(0,0) = 1/3$. On the other hand, we have:

$$\min\left(F(0), G(0)\right) = \min\left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{3} = H(0,0). \tag{B.2}$$

Observe also that X and Y are weakly co-monotonic (Def. 12) by construction, but they are not strongly co-monotonic (see Definition 13).

$X \setminus Y$	0	1	Σ
0	1/3	0	1/3
1	1/3	1/3	2/3
Σ	2/3	1/3	1

Table B.1 Joint probability function $h(x,y) = \mathsf{P}[X = x, Y = y]$ of a weakly co-monotonic vector.

$X \setminus Y$	0	1
0	1/3	1/3
1	2/3	1

Table B.2 Joint cumulative distribution function $H(x,y) = \mathsf{P}[X \leq x, Y \leq y]$ of a weakly co-monotonic vector.

$v \setminus v$	0	2/3	1
0	0	0	0
1/3	0	1/3	1/3
1	0	2/3	1

Table B.3 Values of the associated subcopula $C'(u,v)$.

Observe that $\operatorname{Ran} F = \{0, 1/3, 1\}$ and $\operatorname{Ran} G = \{0, 2/3, 1\}$. The unique subcopula C' (see Thm. 24) is discrete and its values are given in Table B.3. We are now going to reconstruct the unique standard extension copula (Def. 30)

associated with (X, Y). We begin with determining the pairs (a_1, a_2) and b_1, b_2 for different values of $(a, b) \in [0; 1]^2$. Following Definition 30, we obtain:

$$(a_1, a_2) := \begin{cases} (0,0), & a = 0; \\ (0, \frac{1}{3}), & a \in (0; \frac{1}{3}); \\ (\frac{1}{3}, \frac{1}{3}), & a = \frac{1}{3}; \\ (\frac{1}{3}, 1), & a \in (\frac{1}{3}; 1); \\ (1,1), & a = 1. \end{cases} \qquad (b_1, b_2) := \begin{cases} (0,0), & a = 0; \\ (0, \frac{2}{3}), & b \in (0; \frac{2}{3}); \\ (\frac{2}{3}, \frac{2}{3}), & b = \frac{2}{3}; \\ (\frac{2}{3}, 1), & b \in (\frac{2}{3}; 1); \\ (1,1), & b = 1. \end{cases} \tag{B.3}$$

We are now in position to calculate the coefficients λ and μ for different regions of $[0; 1]^2$.

$$\lambda(u) := \begin{cases} 1, & u \in \{0, \frac{1}{3}, 1\}; \\ \frac{u-0}{1/3-0} = 3u, & u \in \left(0, \frac{1}{3}\right); \\ \frac{u-1/3}{1-1/3} = \frac{3}{2}u - \frac{1}{2}, & u \in \left(\frac{1}{3}, 1\right), \end{cases} \tag{B.4}$$

and

$$\mu(v) := \begin{cases} 1, & v \in \{0, \frac{2}{3}, 1\}; \\ \frac{v-0}{2/3-0} = \frac{3}{2}v, & v \in \left(0, \frac{2}{3}\right); \\ \frac{v-2/3}{1-2/3} = 3v - 2, & v \in \left(\frac{2}{3}, 1\right). \end{cases} \tag{B.5}$$

Using the coefficients, we may now reconstruct the values of the standard extension copula C^s. We will do it for the four regions of $[0; 1]^2$ defined by the point $(1/3, 2/3)$.
For $u \in [1/3; 1]$ and $v \in [2/3; 1]$, we get:

$$\begin{aligned} C^s(u, v) := & \frac{1}{3}\left(1 - \frac{3}{2}u + \frac{1}{2}\right)(1 - 3v + 2) + \\ & \frac{1}{3}\left(1 - \frac{3}{2}u + \frac{1}{2}\right)(3v - 2) + \\ & \frac{2}{3}\left(\frac{3}{2}u - \frac{1}{2}\right)(1 - 3v + 2) + \\ & 1\left(\frac{3}{2}u - \frac{1}{2}\right)(3v - 2) \\ = & \frac{1}{2}(3uv - u - v + 1), \qquad u \in \left[\frac{1}{3}; 1\right],\ v \in \left[\frac{2}{3}; 1\right]. \end{aligned} \tag{B.6}$$

For $u \in [0; 1/3)$ and $v \in [2/3; 1]$, we get:

$$\begin{aligned} C^s(u, v) := & 0\,(1 - 3u)(1 - 3v + 2) + \\ & 0\,(1 - 3u)(3v - 2) + \\ & \frac{1}{3}(3u)(1 - 3v + 2) + \\ & \frac{1}{3}(3u)(3v - 2) \\ = & u, \qquad u \in \left[0; \frac{1}{3}\right),\ v \in \left[\frac{2}{3}; 1\right]. \end{aligned} \tag{B.7}$$

For $u \in [0; 1/3)$ and $v \in [0; 2/3)$, we get:

$$\begin{aligned} C^s(u,v) := & \; 0\,(1-3u)\left(1-3\frac{v}{2}\right) + \\ & 0\,(1-3u)\left(3\frac{v}{2}\right) + \\ & 0\,(3u)\left(1-3\frac{v}{2}\right) + \\ & \frac{1}{3}\,(3u)\left(3\frac{v}{2}\right) \\ = & \; \frac{3}{2}uv, \qquad\qquad u \in \left[0; \frac{1}{3}\right),\ v \in \left[0; \frac{2}{3}\right). \end{aligned} \tag{B.8}$$

Finally, for $u \in [1/3; 1]$ and $v \in [0; 2/3)$, we get:

$$\begin{aligned} C^s(u,v) := & \; 0\left(1-\frac{3}{2}u+\frac{1}{2}\right)\left(1-3\frac{v}{2}\right) + \\ & \frac{1}{3}\left(1-\frac{3}{2}u+\frac{1}{2}\right)\left(3\frac{v}{2}\right) + \\ & 0\left(\frac{3}{2}u-\frac{1}{2}\right)\left(1-3\frac{v}{2}\right) + \\ & \frac{2}{3}\left(\frac{3}{2}u-\frac{1}{2}\right)\left(3\frac{v}{2}\right) \\ = & \; \frac{3}{4}uv, \qquad\qquad u \in \left[\frac{1}{3}; 1\right],\ v \in \left[0; \frac{2}{3}\right). \end{aligned} \tag{B.9}$$

Observe that the expressions above coincide with the values specified by C' (see Table B.3).

B.2 Weak Co-monotonicity

Construction

Consider a random variable $U \sim Uni(0;1)$. Define X and Y as follows:

$$X := \begin{cases} 0, & U \in (0;0.5); \\ 1, & U \in [0.5;1); \end{cases} \qquad Y := \begin{cases} 0, & U \in (0;1/3); \\ 1, & U \in [1/3;2/3); \\ 2, & U \in [2/3;1). \end{cases} \tag{B.10}$$

Let H be the joint CDF of the random vector (X,Y) and denote the marginal distribution functions as F and G.

Clearly $X \sim Uni\{0,1\}$ and $Y \sim Uni\{0,1,2\}$, where by Uni we mean discrete uniform distributions. The joint probability function and joint CDF for vector (X,Y) are given in Tables B.4 and B.5. By construction, it is weakly co-monotonic (Def. 12); it is also easy to verify that the upper FH-bound M (Def. 8) is compatible with (X,Y). Further, obviously this vector cannot be strongly co-monotonic (Def. 13) since by pigeonhole principle there is no bijection between the sets consisting of three and two elements.

$X \setminus Y$	0	1	2	Σ
0	1/3	1/6	0	1/2
1	0	1/6	1/3	1/2
Σ	1/3	1/3	1/3	1

Table B.4 Joint probability function $h(x,y) = \mathsf{P}[X = x, Y = y]$ of a discrete weakly co-monotonic vector.

Kendall's τ and τ_b

To evaluate $\tau(X,Y)$, we are going to use its stochastic representation from Theorem 44. Tables B.5–B.8 present the values the calculations are based on.

We thus get

$$\begin{aligned} \tau(X,Y) &= \mathsf{E}\left[H(X,Y) + H(X-,Y)\right] + \mathsf{E}\left[H(X,Y-) + H(X-,Y-)\right] - 1 \\ &= \frac{23}{36} + \frac{1}{4} + \frac{1}{3} + \frac{2}{9} - 1 = \frac{4}{9}. \end{aligned} \tag{B.11}$$

Observe that although (X,Y) is weakly co-monotonic and the upper Fréchet-Hoeffding bound M is compatible with it, $\tau(X,Y) < 0.5$. This illustrates how devastating ties are for the default measure of concordance.

Now let (X^*,Y^*) be an independent copy of (X,Y). For the probabilities of not observing ties it then holds that $\mathsf{P}[X \neq X^*] = 1/2$ and $\mathsf{P}[Y \neq Y^*] \neq 2/3$. For τ_b from Definition 35 this yields:

$$\tau_b(X,Y) := \frac{\tau(X,Y)}{\sqrt{\mathsf{P}[X \neq X^*]\ \mathsf{P}[Y \neq Y^*]}} = \frac{4}{9}\sqrt{3} \approx 0.77. \tag{B.12}$$

$X \setminus Y$	0	1	2
0	1/3	1/2	1/2
1	1/3	2/3	1

Table B.5 Joint cumulative density function $H(x, y) = \mathsf{P}[X \leq x, Y \leq y]$ of a discrete weakly co-monotonic vector.

$X \setminus Y$	0	1	2
0	0	0	0
1	1/3	1/2	1/2

Table B.6 Values of $H(x-, y) = \mathsf{P}[X < x, Y \leq y]$.

$X \setminus Y$	0	1	2
0	0	1/3	1/2
1	0	1/3	2/3

Table B.7 Values of $H(x, y-) = \mathsf{P}[X \leq x, Y < y]$.

$X \setminus Y$	0	1	2
0	0	0	0
1	0	1/3	1/2

Table B.8 Values of $H(x-, y-) = \mathsf{P}[X < x, Y < y]$.

Spearman's ρ and ρ_s

Consider a reference vector (X^*, Y^*) from Definition 23. Specifically, let $X^* \sim X$ and $Y^* \sim Y$ while $X^* \perp\!\!\!\perp Y^*$. Let h^* and H^* denote the joint probability and cumulative distribution functions, respectively; their values are presented in Tables B.9 and B.10. We now compute $\rho(X, Y)$ using the stochastic representation from Theorem 44:

$$\rho(X,Y) = 3\,\mathsf{E}\left[H(X^*,Y^*) + H(X^*-,Y^*)\right] + \mathsf{E}\left[H(X^*,Y^*-) + H(X^*-,Y^*-)\right] - 3, \quad \text{(B.13)}$$

where the expectation is taken with respect to the distribution H^*. We again use Tables B.5–B.8 to calculate the summands and obtain

$$\rho(X,Y) = 3\left(\frac{1}{2} + \frac{2}{9} + \frac{11}{36} + \frac{5}{36}\right) - 3 = 3 \cdot \frac{7}{6} - 3 = \frac{1}{2}. \quad \text{(B.14)}$$

Again (X, Y) is weakly co-monotonic and the upper FH-bound M is compatible with it, yet $\rho(X, Y) = 0.5 < 1$.

Let us now calculate the normalised version of Spearman's ρ_s from Definition 36. Recall that $\Delta F(x) := F(x) - F(x-) = \mathsf{P}[X = x]$. Since both X and Y are discrete uniform, we get that $\Delta F(X) = 1/2$ a.s. and $\Delta F(X) = 1/3$ a.s. We thus obtain:

$$\rho_s(X,Y) := \frac{\rho(X,Y)}{\sqrt{\left(1 - \mathsf{E}\left[\Delta F(X)\right]^2\right)\left(1 - \mathsf{E}\left[\Delta G(Y)\right]^2\right)}} = \frac{1}{2} : \frac{1}{\sqrt{\frac{3}{4} \cdot \frac{8}{9}}} = \frac{\sqrt{6}}{4} \approx 0.61 < 1. \quad \text{(B.15)}$$

$X^* \setminus Y^*$	0	1	2
0	1/6	1/6	1/6
1	1/6	1/6	1/6

Table B.9 Values of the joint probability function $h^*(x, y) = \mathsf{P}[X^* = x, Y^* = y]$ under elements' independence.

$X^* \setminus Y^*$	0	1	2
0	1/6	2/6	3/6
1	2/6	4/6	1

Table B.10 Joint distribution function $H^*(x, y) = \mathsf{P}[X^* \leq x, Y^* \leq y]$ under elements' independence.