FACULTY OF MATHEMATICS AND PHYSICS
Charles University

BACHELOR THESIS

Matej Mifkovič

Copulae for non-continuous distributions

Department of Probability and Mathematical Statistics

Supervisor of the bachelor thesis: doc. RNDr. Michal Pešta, Ph.D.
Study programme: Mathematics
Study branch: Financial Mathematics

Prague 2022

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

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 Prague, 10th May 2022 ....

Author's signature

I thank my supervisor doc. RNDr. Michal Pešta, Ph.D. for his valuable insights, remarks and words of encouragement.

Title: Copulae for non-continuous distributions

Author: Matej Mifkovič

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Copulas are a popular choice when assessing the dependence structure between continuous random variables. However, major difficulties arise as soon as one of the random variables is non-continuous. This thesis introduces the basics of copula theory based on the cited literature. The main focus of this thesis is to introduce the reader to the field of non-continuous copula modelling and highlight all major issues. At the same time, empirical evidence with discussion is presented to suggest that copula modelling and inference may be a viable option when additional care and caution are applied. Afterwards, accumulated theoretical knowledge is demonstrated on real-world data concerning bike-sharing.

Keywords: copula, non-continuous distribution, discrete distribution

Contents

1. Introduction

Studying dependence between random variables belongs among the most important disciplines in inferential statistics. Apart from its omnipresent theoretical relevance, where dependence (resp. independence) in premises of a theorem or a proposition yields vastly different outcomes, the concept of dependence plays a pivotal role in many real-world applications where the interaction of two or more entities is encountered. Whenever we observe an interaction of 2 or more random variables, the concept of joint distribution is natural, yet rather complicated, to work with. Independence is the most convenient form of dependence given that it enables us to work rather with the product of marginals than with the joint distribution itself, provided the simpler nature of a marginal distribution. Therefore, it was natural to search for a way how to “couple” a joint distribution with respective marginals through a functional relationship. By the virtue of Abe Sklar and his eponymous theorem, we are, indeed, able to capture the relationship between a joint distribution and respective marginals through a special two-argument real function named copula, no matter the dependence nature of random variables. Moreover, having only continuous random variables the copula is unique, which proves essential in copula modelling and subsequent inference. Although it took a while, copulas rose to the limelight and became an integral part of dealing with dependence in fields ranging from quantitative finance to biomedical statistics provided that the dependence structure is captured much simpler than in dealing with joint distribution alone. However, as soon as we allow for at least one of the random variables to be non-continuous, there is a catch. The much-desired uniqueness property no longer holds for non-continuous random variables, which leads to a plethora of both theoretical and practical issues in copula modelling.

In this work, we will provide sufficient preliminary theoretical knowledge in order to properly define a copula. Afterwards, we turn our attention to several useful properties of a copula and state the paramount Sklar’s theorem. Next, we will restrict ourselves to a specific subcategory of copulas named Archimedean copulas which are especially convenient to work with. Then, we introduce a rank-based dependence measure called Kendall’s τ and highlight its connection to copula modelling and inference. Concluding the chapter on theory, we elaborate on several issues that occur as soon as we leave the continuous setting, and we provide partial remedies to these issues arguing in favour of copula modelling in a non-continuous setting, despite the toll it takes due to the lack of uniqueness. Even though we are concerned with non-continuous distributions, several pieces of theory will be stated only for discrete distributions regardless of the existence of possible generalizations to non-continuous settings.

For the application part, we will utilize the accumulated knowledge from the previous chapter which will be applied to real-world data, more specifically, data set regarding bike-sharing will be scrutinized. We will evaluate dependence structure between a hourly number of registered users and casual (non-registered) users, and subsequently construct copula models from observed data where we will use some of most popular Archimedean copula families, namely Clayton, Gumbel, Frank copula families. In a similar fashion as in the chapter about theory, we will examine copula modeling and inference in case when both marginal distributions are discrete. Although generalization to higher dimension is fairly straightforward for the majority of introduced theoretical and practical concepts, we will restrict to bivariate settings for the purposes of this work.

2. Theory

As stated in the introduction, we will consider only bivariate case even though the extension of theory to higher dimensional cases is fairly straightforward for the vast majority of theorems and definitions stated below. For details see *Chapter 2 (Section 2.10)* in Nelsen (2006) or refer to *Chapter 6* of McNeil et al. (2015). First, we will do the groundwork for a copula definition, then we define copula and its properties, and subsequently we link copulas to probability theory through Sklar's Theorem. Since the focus of this work is on copulas for non-continuous distributions, we introduce the non-uniqueness issue. Then, we will restrict ourselves to parametric families of copulas, namely Archimedean copulas, with related theoretical background. Subsequently, we move to establishing so-called measures of concordance (namely Kendall's τ) and their relationship to copulas. Finally, having covered all of the necessary theory, we delve into the discussion of all major issues, related to copula modelling and inference, arising as soon as we leave continuous settings. Possible remedies to such issues will be provided and we will state our verdict on copula modelling and inference for non-continuous random variables.

2.1 Preliminaries

Prior to delving into defining copula, its basic properties and the fundamental theorem of copula modelling, i.e. Sklar's theorem, we need to make familiar some mathematical concepts and its properties in order to make the conversation about copulas more fluent. First, we define a 2-increasing function and state its properties that are used in theory related to copulas. 2-increasing function is sort of a generalisation of a non-decreasing function in 1 dimension. In order to define 2-increasing function, we will consider extended real-number set $\mathbb{R}_* = (-\infty, \infty) \cup \{-\infty, +\infty\}$, 2-place real function H, whose domain is nonempty subset of $\mathbb{R}_*^2$ and that for every rectangle $[x_1, x_2] \times [y_1, y_2] \subseteq dom(H)$ we have so-called *H-volume of rectangle B*, denoted $V_H(B)$ given by

$$V_H(B) = H(x_2, y_2) + H(x_1, y_1) - H(x_2, y_1) - H(x_1, y_2).$$

Quite naturally, we establish first order differences with respect to the first and second coordinate which we define $\Delta_{x_1}^{x_2} = H(x_2, y) - H(x_1, y)$, $\Delta_{y_1}^{y_2} = H(x, y_2) - H(x, y_1)$, respectively (Nelsen, 2006). Finally, we have managed to define all concepts required for a neat definition of a 2-increasing function.

Definition 1. *(Nelsen, 2006) Let, $A_1 \neq \emptyset \neq A_2$, H is a real 2-placed real function such that $dom(H) = A_1 \times A_2$. Next, consider a rectangle $B = [x_1, x_2] \times [y_1, y_2] \subseteq dom(H)$. Then, we say that H is 2-increasing if $V_H(B) \geq 0$, $\forall B \subseteq dom(H)$.*

Note. H-volume of 2-place real function is equal to the second degree difference of H, i.e. $V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y)$ since: $\Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) = \Delta_{y_1}^{y_2}(H(x_2, y) - H(x_1, y)) = \Delta_{y_1}^{y_2} H(x_2, y) - \Delta_{y_1}^{y_2}(H(x_1, y) = H(x_2, y_2) - H(x_2, y_1) - (H(x_1, y_2) - H(x_1, y_1)) = V_H(B)$.

Whereas a 2-increasing function does not imply that it is non-decreasing in each of its arguments, nor the opposite is true (for details and examples see Nelsen (2006)), the following lemma, however, provides groundwork for proving non-decreasing property in arguments of 2-increasing function under an additional assumption. The author presents his own proof given that there was no proof provided in the original text and this proof helps familiarise with definitions provided above.

Lemma 1. *(Nelsen, 2006) Let $A_1, A_2 \subseteq \mathbb{R}_*$, $A_1 \neq \emptyset \neq A_2$, H is a 2-increasing function such that $dom(H) = A_1 \times A_2$. Then, functions defined as $t \mapsto H(t, y2) - H(t, y_1)$, $t \mapsto H(x_2, t) - H(x_1, t)$ are non-decreasing on A_1, A_2, $\forall x_1, x_2 \in A_1 : x_1 \leq x_2$, $\forall y_1, y_2 \in A_2 : y_1 \leq y_2$, respectively.*

Proof. We will only prove that $t \mapsto H(t, y2) - H(t, y_1)$ is non-decreasing, the other would be proved analogously. Let us take $t_1, t_2 \in A_1 : t_1 \leq t_2$. Given that H is 2-increasing, we have from definition $V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq 0$. Fix $x_2 := t_2$, $x_1 := t_1$. Straightaway, we obtain that $H(t_2, y_2) - H(t_2, y_1) \geq H(t_1, y_2) - H(t_1, y_1)$, which is by definition a non-decreasing function with respect to $t \in A_1$.

⊠

This lemma along with additional assumption that H is a *grounded* function, i.e. given that a_1, a_2 are minimal elements of sets A_1, A_2, respectively, we have that $H(x, a_2) = 0 = H(a_1, y)$, $\forall (x, y) \in A_1 \times A_2$ (Nelsen, 2006). This additional assumption on H gives us finally all tools required for showing that this 2-increasing function is non-increasing in each of its arguments. Again, the author provided his own version of the proof, which is more detailed and precise than "the proof by reference" in the original text.

Lemma 2. *(Nelsen, 2006) H is grounded 2-increasing function with domain $A_1 \times A_2$, $A_1 \neq \emptyset \neq A_2, A_1, A_2 \in \mathbb{R}_*$. Then H is non-decreasing in each argument.*

Proof. Let $a_1 = min(A_1), a_2 = min(A_2)$, set $x_1 := a_1, x_2 := a_2$, given that $t \mapsto H(x_2, t) - H(x_1, t)$ is non-decreasing, and that $H(x_1, t) = H(a_1, t) = 0$, we immediately have that, $t \mapsto H(x_2, t) - H(x_1, t) = H(x_2, t) - H(a_1, t) = H(x_2, t)$ is a non-decreasing function $\forall t \in A_2$, $\forall x_2 \in A_1$. Proof for the second argument of H is analogous.

⊠

Now, let us consider greatest elements of sets A_1, A_2, i.e. set $b_1 := max(A_1)$, $b_2 = max(A_2)$. We define *margins* of 2-increasing function $H : A_1 \times A_2 \to \mathbb{R}$ as functions F, G, whose respective domains are $dom(F) = A_1$, $dom(G) = A_2$, and which are given by $F(x) = H(x, b_2)$, $\forall x \in A_1$, $G(y) = H(b_1, y)$, $\forall y \in A_2$, respectively (Nelsen, 2006).

To conclude this section, we finish with with the lemma about inequality between grounded 2-increasing function and its margins. We present our own, more detailed version of the proof from Nelsen (2006).

Lemma 3. *H is grounded 2-increasing function with margins with $dom(H) = A_1 \times A_2$, $A_1, A_2 \subseteq \mathbb{R}_*$, $A_1 \neq \emptyset \neq A_2$. Let $(x_1, y_1), (x_2, y_2)$ be any points in $A_1 \times A_2$. Then*

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Proof. First, we consider case $x_2 \geq x_1$ (case $x_2 \leq x_1$ is analogous). Fix $\alpha(t) = H(x_2, t) - H(x_1, t) \geq 0$, $\forall t \in A_2$, $\alpha(t)$ is non-negative because H is non-decreasing in each argument. Next, set $b_2 := max(A_2)$, then $\max_{t \in A_2} \alpha(t) = \alpha(b_2)$. From that follows $\alpha(b_2) = H(x_2, b_2) - H(x_1, b_2) = F(x_2) - F(x_1)$, form which we directly get $\forall y_2 \in A_2 : H(x_2, y_2) - H(x_1, y_2) \leq F(x_2) - F(x_1)$. Along with the analogous case, we get that $\forall y_2 \in A_2 : |H(x_2, y_2) - H(x_1, y_2)| \leq |F(x_2) - F(x_1)|$.

The same holds for $y_2 \leq y_1, y_1 \leq y_2$, respectively, i.e. $\forall x_1 \in A_1 : |H(x_1, y_2) - H(x_1, y_1)| \leq |G(y_2) - G(y_1)|$.

Finally, from triangle inequality and addition of the obtained inequalities we get

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)| \leq$$
$$\leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

⊠

2.2 Copula definition and Sklar's Theorem

2.2.1 Copula definition and its properties

We start this subsecition with the informal definition of a *2-subcopula* (or just a *subcopula*, due to dealing only with 2-dimensions). A *subcopula* is a grounded 2-increasing function with margins. This leads us to the *copula* definition. Informally, a copula C is a subcopula, whose domain is a unit square, i.e. $dom(C) = [0, 1] \times [0, 1] = \mathbb{I}^2$, where interval $[0, 1]$ is denoted $\mathbb{I}$ for brevity. The more formal definition of a copula, loosely based on Nelsen (2006), is given below.

Definition 2. *a copula is a function $C : \mathbb{I}^2 \to \mathbb{I}$ such that:*

(a) $\forall u, v \in \mathbb{I} : C(0, v) = C(u, 0) = 0$

(b) $\forall u, v \in \mathbb{I} : C(1, v) = v, C(u, 1) = u$

(c) $\forall u, v \in \mathbb{I} : 0 \leq C(u, v) \leq 1$

(d) $\forall 0 \leq u_1 \leq u_2 \leq 1, 0 \leq v_1 \leq v_2 \leq 1$*:*

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$$

Note that $C(u, v) = V_C([0, u] \times [0, v]), \forall u, v \in \mathbb{I}$.

Note. For the remainder of this work we will remain rather loose in terminology regarding copulas, especially when non-continuous random variables will be of concern. We will refer to object related to non-continuous random variables as copulas even though Genest and Nešlehová (2007) refrain from doing so. Yet, terminology is not unified throughout the literature.

After defining copula, the following theorem specifies the bounds that any copula C can attain. We provide our version of the proof based on the proof from the original text, however, our theorem concern copulas, whereas theorem from Nelsen (2006) deals with subcopulas.

Theorem 4. *(Nelsen, 2006) Let C be a copula, then $\forall u, v \in \mathbb{I} : \max(u+v-1, 0) \leq C(u, v) \leq \min(u, v)$.*

Proof. Let $u, v \in \mathbb{I}$, then from the copula definition we have $C(u, v) \leq C(u, 1) = u$, $C(u, v) \leq C(1, v) = v \implies C(u, v) \leq \min(u, v)$. Furthermore, it holds that $C(u, v) \geq 0$ and that $V_C([u, 1] \times [v, 1]) \geq 0 \implies C(1, 1) - C(u, 1) - C(1, v) - C(u, v) \geq 0 \implies C(u, v) \geq u + v - 1$. Gluing it all together, we obtain $\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$.

⊠

It is apparent that margins of a copula give an identity function, i.e. $C(u, 1) = F(u) = u, C(1, v) = G(v) = v$, which implies that when a copula is used in the context of random variables, the copula margins have uniform distribution (however, the considered random variables must be continuous). The upper and lower copula bounds are called Fréchet-Hoeffding lower and upper bounds and are commonly denoted $M(u, v) = \min(u, v)$, and $W(u, v) = \max(u + v - 1, 0)$. These bounds themselves are copulas, and together with the product copula defined as $\Pi(u, v) = uv$ they constitute 3 most important copulas, referred to as *fundamental* copulas by McNeil et al. (2015). Their importance will be highlighted after stating Sklar's Theorem. Furthermore, the following theorem shows that copulas are uniformly continuous on its domain. Another property of copulas that can be directly observed from the inequality below is that copulas are Lipschitz continuous.

Theorem 5. *(Nelsen, 2006) For any copula C and $\forall (u_1, v_1), (u_2, v_2) \in \mathbb{I}^2$ holds the inequality:*

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

Proof. Direct consequence of Lemma 3.

⊠

2.2.2 Sklar's Theorem

So far we considered copulas solely as 2-increasing grounded real valued functions with their domain and range being $\mathbb{I}^2$, $\mathbb{I}$, respectively. Now, we apply the concept of copulas to probabilistic theory, since the main use of copulas lies in a convenient way of capturing and modelling the dependence of 2 (or more) random variables. The theorem around which the copula modelling revolves was devised by Abe Sklar in 1959 in his paper *Fonctions de répartition à n dimensions et leurs marges.*

Theorem 6 (**Sklar's Theorem**). *Let H be a joint distribution function with margins F, G, respectively. Then, there exist a copula C such that $\forall x, y \in \mathbb{R}_*$ we have*

$$H(x, y) = C(F(x), G(y)). \tag{2.1}$$

If F and G are continuous, then the copula C is unique, otherwise C is only unique on $Ran(F) \times Ran(G)$. Alternatively, the above may be restated as: having a copula C and marginal distribution functions F and G, the the function H defined by by the above equation is a joint distribution function with margins F and G.

Proof. Nelsen (2006).

⊠

Before exploring the non-continuous case, we illustrate the result of combining Sklar's Theorem with Theorem 4. The result it yields is that the Fréchet-Hoeffding bounds for a unique copula C for joint distribution H with respective continuous margins F and G are:

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) = C(F(x), G(y)) \leq min(F(x), G(y)) \tag{2.2}$$

The copula C equals to the lower bound, i.e. $W(F(x), G(y)) = \max(F(x) + G(y) - 1, 0)$ when random variables with distribution functions F and G are in a perfectly counter-monotonic relationship, i.e. given random variables X_1, X_2, we have $(X_1, X_2) \stackrel{D}{=} (f_1(X_3), f_2(X_3))$, where f_1, f_2 are increasing and decreasing functions, respectively, and X_3 is some random variable. Analogously C equals to the upper bound when random variables are in a perfectly co-monotonic relationship, i.e. $(X_1, X_2) \stackrel{D}{=} (f_1(X_3), f_2(X_3))$, where f_1, f_2 are increasing functions and X_3 is some random variable (McNeil et al., 2015). As for the product copula Π, the equality $H(x, y) = C(F(x), G(y)) = \Pi(F(x), G(y)) = F(x)G(y)$ implies that the product copula captures the independence between random variables with distribution functions F and G.

As argued by Genest and Nešlehová (2007) the uniqueness property serves as "the corner stone for of inference in copula modelling." However, the uniqueness property no longer holds true in the non-continuous case as the copula for 2 provided non-continuous distributions is guaranteed to be unique only on the mesh of ranges of respective marginal distributions. This leads to the class of copulas that satisfy the equality in Sklar's Theorem to be infinite. We will demonstrate this somehow abstract notion on the specific example adopted from McNeil et al. (2015).

Example. Consider bivariate Bernoulli distribution of (X_1, X_2) given by:

$$P[X_1 = 0, X_2 = 0] = \frac{1}{8},\ P[X_1 = 1, X_2 = 0] = \frac{2}{8},$$

$$P[X_1 = 0, X_2 = 1] = \frac{2}{8},\ P[X_1 = 1, X_2 = 1] = \frac{3}{8},$$

from which immediately follows that $P[X_1 = 0] = \frac{3}{8} = P[X_2 = 0]$, i.e. $X_1, X_2 \sim Bernoulli(\frac{5}{8})$. Adopting the notation used in Sklar's Theorem, we denote cumulative distribution functions (cdf) of X_1, X_2 as F and G, respectively, and their respective joint cdf as H. It is apparent that

$$F(x) = \begin{cases} 0, \ x < 0 \\ \frac{3}{8}, \ 0 \leq x < 1 \\ 1, \ x \geq 1 \end{cases}$$

from which $Ran(F) = \{0, \frac{3}{8}, 1\}$ and the same holds for G, i.e. $Ran(F) = Ran(G)$. By the definition of copula, we have by default determined values

$$C(0, G(y)) = C(F(x), 0) = 0,$$

$$C(F(x), 1) = F(x)\ , C(1, G(y)) = G(y),\ C(1, 1) = 1.$$

This means that the only one point on the grid $Ran(F) \times Ran(G)$, i.e. the point $(\frac{3}{8}, \frac{3}{8})$, puts restriction on the copula C and this condition is that $C(\frac{3}{8}, \frac{3}{8}) = H(0, 0) = \frac{1}{8}$, and the class of copulas fulfilling this condition is infinitely large, thus no unique copula exist for this bivariate Bernoulli distribution.

Note. It feels appropriate to state that in fact none of the "copulas" out the aforementioned class is a copula, but it is a subcopula because domain-wise $\mathbb{I}^2 \neq Ran(F) \times Ran(G)$. In a nutshell, subcopula is the copula, with relaxed condition on its domain; to be specific, the domain of subcopula must be the subset of $\mathbb{I}$ containing 0 and 1. For details see the definition in Nelsen (2006). In defence of this, Nelsen (2006) shows that "any subcopula can be extended to a copula," yet this extension is not unique in general. For proof, see *Lemma 2.3.5* in Nelsen's copula monograph (2006). Hence, if class of subcopulas is infinite, so is the class of copulas that are extensions of the former.

Note. The same is shown in a similar fashion in *Example 1* in Genest and Nešlehová (2007). However, they are rather explicit that given copula representations are not copulas.

Another noticeable thing is that marginal distribution functions F and G are not uniformly distributed on the interval $(0, 1)$ as they both are piecewise constant functions with jumps, which is essentially true for any univariate non-continuous distribution.

The lack of uniqueness of copula for non-continuous random variables invalidates the use of the inversion method for copula construction, which is rather straightforward in case of continuous distributions. This stems from the fact that the generalised inverse (or "quasi-inverse") of a distribution function F, i.e. $F^{(-1)}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$ does not coincide with the inverse of F for non-continuous distributions since F is not strictly continuous, which means that the equality $F^{(-1)}(F(x)) = x,\ \forall x \in \mathbb{R}$ no longer holds. This obstacle compromises the validity of constructing copulas from the joint distribution functions. For details on quasi-inverses, its relationship to inverses and (sub)copulas refer to Nelsen (2006).

Even though the lack of uniqueness represents an issue in copula modelling, it may not be an insuperable one given that the class of copulas, for which the equation 2.1 holds, is reasonably small. Genest and Nešlehová (2007) argue that such class of copulas would mimic fairly closely the dependence characteristic of the joint distribution H. In order to analyze the size of the copula class, we need to measure the smallest and largest elements of the class as well as their associated measure of dependence. Let us restrict ourselves only to discrete distributions and denote such copula class by $\mathcal{C}$. To compute, the size of this class, we will make use of Carley's bounds (Carley, 2002). While Carley (2002) in her paper defined Carley's bounds for a subcopula with a finite support, Genest and Nešlehová (2007) proposed the generalized version of those bounds for a copula that is specified at the grid of points $(F(i), G(j)),\ i, j \in \mathbb{N}_0$. We will consider, without loss of generality, that random variables X_1 and X_2 take only integer values. The following construction proceeds from Genest and Nešlehová (2007).

Note. For the rest of this work, we will denote the set of natural numbers including 0 by $\mathbb{N}_0$, and the set of natural numbers starting from 1 by $\mathbb{N}$.

Consider joint distribution H of (X_1, X_2) with the respective margins F and G for X_1 and X_2. Denote

$$h_{ij} = P[X_1 = i, X_2 = j],$$

$$H(i, j) = P[X_1 \leq i, X_2 \leq j],$$

where $h_i j$ represents the probability mass assigned by C to a rectangle

$$(F(i-1), F(i)] \times (G(j-1), G(j)],$$

and $F(-1) = G(-1) = 0$ by convention. Respective marginal probability mass functions (pmf) of X_1 and X_2 are given by

$$h_{i+} = \sum_{j=0}^{\infty} h_{ij} = P[X_1 = i],$$

$$h_{+j} = \sum_{i=0}^{\infty} h_{ij} = P[X_2 = j].$$

Then we define the upper Carley's bound as

$$C_H^+(u, v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \max\{0, \min\{u - \alpha_{ij}, v - \beta_{ij}, h_{ij}\}\}, \tag{2.3}$$

and the lower Carley's bound as

$$C_H^-(u, v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \max\{0, -h_{ij} + \min\{u - \gamma_{ij}, h_{ij}\} + \min\{v - \delta_{ij}, h_{ij}\}\} \tag{2.4}$$

where

$$\alpha_{ij} = \sum_{k=0}^{i-1} h_{k+} + \sum_{l=0}^{j-1} h_{il},\ \ \beta_{ij} = \sum_{l=0}^{j-1} h_{+l} + \sum_{k=0}^{i-1} h_{kj},$$

$$\gamma_{ij} = \sum_{k=0}^{j-1} h_{k+} + \sum_{l=j+1}^{\infty} h_{il},\ \ \delta_{ij} = \sum_{l=0}^{j-1} h_{+l} + \sum_{k=i+1}^{\infty} h_{kj}.$$

Note. By convention, an empty sum equals to zero.

The definition above allows us to formulate the statement adopted from Genest and Nešlehová (2007), specifically *Proposition 2.*

Theorem 7. *Let H be a bivariate discrete distribution on $\mathbb{N}_0^2$, $\mathcal{C}$ the class of all copulas that fulfills the equality 2.1 for H. Then the upper and lower Carley's bounds for $\mathcal{C}$ are given by equations 2.3 and 2.4, respectively.*

Proof. See Carley (2002).

⊠

Note. Even though the proof from Carley (2002) takes into account only finite support of H, Genest and Nešlehová (2007) state that the proof of their proposed generalised theorem is "essentially the same."

Apart from previous method, size of the class $\mathcal{C}$ can be approximated from above by the following theorem stated as *Proposition 6* in Genest and Nešlehová (2007). We provide a slightly expanded version of the proof in order to demonstrate usefulness of theorems stated above that enable us to prove the following inequality.

Theorem 8. *Let (X_1, X_2) be the pair of discrete integer-valued non-negative random variables with the joint distribution function H and their their respective margins F and G. Denote the rectangle $(F(i-1), F(i)] \times (G(j), G(j-1)]$ as R_{ij}, $\forall i, j \in \mathbb{N}_0$. Then $\forall C_1, C_2 \in \mathcal{C}$:*

$$|C_1(u,v) - C_2(u,v)| \leq 2\sum_{i=0}^{\infty}\sum_{j=o}^{\infty}(P[X_1 = i] + P[X_2 = j])\mathbb{1}\{(u,v) \in R_{ij}\} \leq$$

$$\leq 2(\max_{i\in\mathbb{N}_0}\{P[X_1 = i]\} + \max_{j\in\mathbb{N}_0}\{P[X_2 = j]\}) \tag{2.5}$$

Proof. Based on the proof from Genest and Nešlehová (2007). From Lemma 2 we have that

$$|C_1(u,v) - C_1(F(i), G(j))| \leq |u - F(i)| + |v - G(j)|$$

and the same is true for C_2. Since $F(i) = P[X_1 \leq i]$, $G(j) = P[X_2 \leq j]$, $\forall i, j \in \mathbb{N}_0$, we have that

$$F(i) - u = P[X_1 \leq i] - u \leq P[X_1 \leq i] - P[X_1 \leq i-1] = P[X_1 = i],$$

since $F(i-1) < u \leq F(i)$, and analogous holds for X_2. Next, we obtain from the triangle inequality that $\forall (u,v) \in R_{ij}$:

$$|C_1(u,v) - C_2(u,v)| \leq |C_1(u,v) - C_1(F(i), G(j))| + |C_2(u,v) - C_2(F(i), G(j))|.$$

Now, we need to notice that thanks to Sklar's Theorem, we have that

$$C_1(F(i), G(j)) = C_2(F(i), G(j)),\ \forall i, j \in \mathbb{N}_0$$

as all copulas from $\mathcal{C}$ coincide on $Ran(F) \times Ran(G)$. Finally, combining all of the above together we obtain:

$$|C_1(u,v) - C_2(u,v)| \leq 2\{|u - F(i)| + |v - G(j)|\} \leq 2\{P[X_1 = i] + P[X_2 = j]\}.$$

Since (u, v) can lie only in one rectangle R_{ij} at a time, we have

$$\sum_{i=0}^{\infty}\sum_{j=o}^{\infty}(P[X_1 = i] + P[X_2 = j])\mathbb{1}\{(u,v) \in R_{ij}\} = P[X_1 = i] + P[X_2 = j],$$

which proves the first inequality in 2.5. Given the previous equality, the second inequality in 2.5 is fulfilled trivially.

⊠

This result suggests that the bigger the range of discrete random variables X_1, X_2, the less serious the uniqueness issue may be as the size of $\mathcal{C}$ get smaller. However, as pointed out by Genest and Nešlehová (2007), the formal proof for this conjecture needs to be presented.

Indeed, the lack of uniqueness, although a problem, does not necessarily mean that copula modelling with discrete (or even non-continuous) random variables is invalid. This claim is supported by Genest and Nešlehová (2007), who despite additional (and even more severe) issues, which we will approach later, argue that copula modelling remains a valid option for constructing multivariate distributions with discrete margins, yet one must proceed with caution. Others advocating for copula modelling despite the uniqueness issue are Nikoloulopoulos (2013) and Trivedi and Zimmer (2017). The former states that non-identifiability is a separate theoretical issue and does not have any bearing on copula dependence modelling for discrete data, whereas the latter argue that the uniqueness problem diminishes when the count outcomes of discrete random variable cover the outcome domain in a more complete way, which simply puts in other words the above conjecture made by Genest and Nešlehová (2007). The latter claim will be discussed in more depth at the end of this chapter. All in all, all non-continuous issues that hinder copula modelling and inference for non-continuous data stem from the non-uniqueness issue.

2.3 Archimedean copulas

Adopting classification used in McNeil et al. (2015), we classify copulas into 3 categories: fundamental, explicit and implicit copula. Fundamental copulas, i.e. M, W and Π copulas, have been mentioned earlier. Further, we will restrict ourselves to explicit copulas, i.e. copulas that can be expressed in simple-closed forms and the dependence modelled by them is governed by copula parameter(s). Namely, we restrict to one-parameter Archimedean copulas since they are pretty convenient to work with as they can be constructed with ease, we have a great variety of them to choose from, and they have many convenient properties (Nelsen, 2006). In general, we denote the family of all single-parameter copulas by $\mathcal{C}_\theta$ and a copula governed by the parameter θ belonging to this family as C_θ. The theory from this section is adopted from Nelsen (2006).

First, we need to define an appropriate "inverse" of a function that will be fundamental in Archimedean copula construction.

Definition 3. *Consider a continuous strictly decreasing function* $\varphi : \mathbb{I} \mapsto [0, \infty]$ *fulfilling* $\varphi(1) = 0$. *Then, the pseudo-inverse of* φ *is defined as the function* $\varphi^{[-1]} : [0, \infty] \mapsto \mathbb{I}$ *is given by:*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), \ 0 \leq t \leq \varphi(0) \\ 0, \ \varphi(0) \leq t \leq \infty \end{cases} \tag{2.6}$$

where the case for $t = \infty$ *we understand as* $\lim_{t \to \infty} \varphi^{[-1]}(t) = 0$.

Note. Definitions of inverse and pseudo-inverse blend together when

$$\lim_{t \to 0^+} \varphi(t) = \infty.$$

The following theorem provides the form of Archimedean copulas, meticulously shows that they are indeed copulas, and highlights the key property of the function that generates an Archimedean copula. The theorem was adapted as the combination of 2 distinct lemmas and 1 theorem from Nelsen (2006), and proofs for *(a)* and *(b)* parts of this theorem are provided in order to provide them in bigger detail to the reader than they are in the original text.

Theorem 9. *Let* $\varphi : \mathbb{I} \mapsto [0, \infty]$ *be a continuous, strictly decreasing function such that* $\varphi(1) = 0$ *and* $\varphi^{[-1]}(t)$ *is its pseudo-inverse. Define a function* $C : \mathbb{I}^2 \mapsto \mathbb{I} :$ $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$. *Then:*

(a) C satisfies copula boundary conditions of Theorem 4.

(b) Fulfilling (a), C is 2-increasing if and only if

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1, \ \forall u_1, u_2, v \in \mathbb{I} : u_1 \leq u_2.$$

(c) C is a copula if and only if φ *is convex.*

Copulas of the form $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ *are called Archimedean copulas and* φ *their generator.*

Proof. Adapted from Nelsen (2006).
(a) Directly from the Definition 3 we have:

$$C(u, 0) = \varphi^{[-1]}(\varphi(u) + \varphi(0)) = 0,$$

$$C(u, 1) = \varphi^{[-1]}(\varphi(u) + \varphi(1)) = \varphi^{[-1]}(\varphi(u)) = u,$$

thanks to the fact that $\varphi(1) = 0$, $\varphi(u) \geq 0$ and $\varphi(u) \leq \varphi(0)$, $\forall u \in \mathbb{I}$. The analogous holds for the other argument of C.

(b) Given the Definition 1, C is 2-increasing if $V_C([u_1, u_2] \times [v, 1]) \geq 0$, and $0 \leq V_C([u_1, u_2] \times [v, 1]) = C(u_2, 1) - C(u_2, v) - C(u_1, 1) + C(u_1, v) = u_2 - C(u_2, v) - u_1 + C(u_1, v)$, from which we get

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1.$$

To prove the opposite implication, we assume that $C(u_2, v) - C(u_1, v) \leq u_2 - u_1$, and choose $v_1, v_2 \in \mathbb{I} : v_1 \leq v_2$. Given the Definition 3, it holds that

$$C(0, v_2) = 0 \leq v1 \leq v_2 = C(1, v_2).$$

$\varphi, \varphi^{[-1]}$ being continuous cause C to be continuous, and therefore there exist t such that $C(t, v_2) = v_1$, i.e. $\varphi(v_2) + \varphi(t) = \varphi(v_1)$. To complete the proof, it suffices to show that $V_C([u_1, u_2] \times [v_1, v_2]) \geq 0$, which can be seen from

$$\begin{aligned} C(u_2, v_1) - C(u_1, v_1) = \varphi^{[-1]}(\varphi(u_2) + \varphi(v_1)) - \varphi^{[-1]}(\varphi(u_1) + \varphi(v_1)) = \\ = \varphi^{[-1]}(\varphi(u_2) + \varphi(v_2) + \varphi(t)) - \varphi^{[-1]}(\varphi(u_1) + \varphi(v_2) + \varphi(t)) \leq \\ \leq \varphi^{[-1]}(\varphi(u_2) + \varphi(v_2)) - \varphi^{[-1]}(\varphi(u_1) + \varphi(v_2)) = C(u_2, v_2) - C(u_1, v_2). \end{aligned}$$

(c) see *Theorem 4.1.4* of Nelsen (2006).

⊠

Note. It is straightforward to show that Archimedean copulas are symmetric in arguments, i.e. $\forall u, v \in \mathbb{I} : C(u, v) = C(v, u)$ as $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \varphi^{[-1]}(\varphi(v) + \varphi(u)) = C(v, u)$

The convenience of Archimedean copulas lies in the fact that we only need to know (or find) the generator in order to construct them. The condition for a generator to fulfill is that it needs to be a decreasing convex continuous function $\varphi : \mathbb{I} \mapsto [0, \infty]$ fulfilling $\varphi(1) = 0$. This causes that there are many Archimedean copula families which provide the variety of dependence structures, which is one of the main reason for their usefulness in statistical modelling as argued by Nelsen (2006).

For the purposes of this work, we will work with Clayton, Frank and Gumbel copulas that all belong to Archimedean copulas. The following table, whose design is inspired by the table found in Nelsen (2006), lists the basic properties of these selected copulas. Similar table for more Archimedean families can be found in Nelsen (2006).

Note. Various authors refer to Gumbel copula mentioned above as Gumbel-Hougaard copula. The reason behind this that there is another copula, given by

$$C(u, v) = uv \exp\left\{-\theta \log(u) \log(v)\right\},$$

with the generator $\varphi(t) = \log(1 - \theta \log(t))$ bears Gumbel's name as well. However, various authors tend to refer to the latter Gumbel copula as Gumbel-Barnett copula (Nelsen, 2006).

	$C_\theta(u, v)$	$\varphi_\theta(t)$	$Dom(\theta)$
Clayton	$(\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}}$	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1, \infty) \setminus 0$
Frank	$-\frac{1}{\theta} \log(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$	$-\log(\frac{e^{-\theta t}-1}{e^{-\theta}-1})$	$(-\infty, \infty) \setminus 0$
Gumbel	$exp\{-((-\log(u))^\theta + (-\log(v))^\theta)^{\frac{1}{\theta}}\}$	$(-\log(t))^\theta$	$[1, \infty)$

Table 2.1: Overview of selected copula families.

The final shape of copula depends on the copula parameter θ, so it is appropriate to examine how the shape of a given copula changes as θ tends to its limiting values. Given that for any value of θ from the respective domain of a considered copula we have a univariate continuous function of θ, we can examine limit cases by simply taking limits and use the the usual tools such as L'Hopital rule if need be. In our case, we have the following table for limiting cases.

Clayton	$\theta \to -1: \ C_\theta = W, \ \theta \to 0: \ C_\theta = \Pi, \ \theta \to \infty: \ C_\theta = M$
Frank	$\theta \to -\infty: \ C_\theta = W, \ \theta \to 0: \ C_\theta = \Pi, \ \theta \to \infty: \ C_\theta = M$
Gumbel	$\theta \to 1: \ C_\theta = \Pi, \ \theta \to \infty: \ C_\theta = M$

Table 2.2: Limiting cases of selected copula families.

For step-by-step solution how limiting cases of Clayton copula were determined refer to Appendix A. The limiting cases for the other 2 copulas would be computed in a similar manner; their results are obtained from Nelsen (2006).

Table 2.2 provides a valuable insight in terms of the selection of the proper copula prior to the estimation of parameter θ. One can easily notice that choosing Gumbel copula is not an appropriate choice when there is observed a negative dependence structure between observed random variables. Hence, the knowledge of limiting cases of Archimedean families, along with exploratory analysis of data, enables us to narrow the group of possible copulas suitable for the modelling of the dependence in a given data even before the estimation of θ.

2.4 Dependence and measures of concordance

The usefulness of copulas in terms of application lies in their ability to capture the dependence (or association) between random variables. While there are several ways how to measure dependence, some of them are relevant only in specific settings and can be misleading when applied inappropriately without accounting for the settings of a given model. Below, we will put linear correlation under scrutiny and demonstrate why it is not an appropriate measure of dependence when it comes to capturing dependence structure between random variables in general settings. Later, we establish better dependence measures for our purposes that link well with copulas.

The most basic concept of dependence, familiar to everybody introduced to elementary statistics, is Pearson correlation coefficient, or shortly correlation. Although it is widely used in measuring the dependence between random variable, it should be used with caution as it has its limitations since correlation is only natural to multivariate normal distribution (or more generally speaking to elliptical models) as a dependence concept, where linearity plays a central role (McNeil et al., 2015). It is important to realise that correlation is the measure of *linear* dependence, and thus it can be misleading when other than linear dependence among random variables is encountered. Furthermore, other pitfalls of correlation are that it is not invariant under nonlinear strictly increasing transformations applied to random variables, and that correlation is defined only when we have finite second moments of random variables. For more thorough discussion about pitfalls of linear correlation refer to the section *7.2.2 Linear Correlation* of McNeil et al. (2015). The takeaway message from McNeil et al. (2015) is that "the concept of correlation is meaningless unless applied in the concept of a well-defined joint model. Any interpretation of correlation values in the absence of such a model should be avoided." As a consequence, more general measures of dependence were proposed.

Much better measures of dependence are rank correlations, i.e. a scalar measure of dependence observing the relationship between orderings of observed random variables which are ordinal in nature. Apparent advantages over linear correlation are that considering ranks of random variables, we are able to observe more general dependence structures than the linear dependence. Moreover, from the practical standpoint, we only need to know the order of the observed realizations of given random variables, not their nominal values. This makes it oftentimes simpler in terms of data processing and numerical processes where difficulty may arise when differences between observed values are very small and floating-point arithmetic limitations may be encountered. Nelsen (2006) argues that another advantage of rank correlations is that they are "scale-invariant measures of dependence", as the dependence measure remains the same after we apply strictly increasing transformations to given random variables, whereas in case of linear correlation, we only have invariance when linear increasing transformations are being considered, which may be rather restricting. Furthermore, rank correlations are very useful in the parametrization of a copula, i.e. estimating the dependence parameter θ, as there is a functional relationship between a given rank correlation measure and particular copula family. This is due to the fact that rank correlation does not depend on the marginal distributions of given random variables, but it depends only on the copula of their bivariate distribution (McNeil et al., 2015).

Note. Nelsen's argument stems from the premise that we have *continuous random variables*. We will show later in the next section that number of difficulties arise when non-continuous random variables are considered. The same is true for the rank correlations and the copula dependence parameter θ.

Two most common rank correlations are Kendall's τ and Spearman's ρ denoted $\tau^{(K)}$, $\rho^{(S)}$, respectively. Both are (among many others) so-called measures of concordance in the sense of Scarsini (1984). However, we will restrict ourselves only to informal and formal definitions of concordance and discordance. For axioms, which must be respected by any concordance measure, and theory related to copulas refer to the work of Scarsini (1984). Considering bivariate context, concordance informally means that large values of one random variable tend to be associated with large values of the other random variable. Conversely, discordance means that large values of one random variable are associated with small values of the other. To elaborate, we provide the formal definition of concordance in the setting where random variables are continuous. We will define both probabilistic and sample versions of concordance and discordance.

Note. There exist several other measures of concordance such as Blomqvist's β and Gini's γ. For details refer to the section *5.1.4 Other Concordance Measures* of Nelsen (2006).

Definition 4. *(Nelsen, 2006) Let $X_1,\ X_2$ be continuous random variables, consider $(x_1^{(1)}, \dots, x_n^{(1)})$ to be a finite random sample from X_1 and $(x_1^{(2)}, \dots, x_n^{(2)})$ to be the random sample from X_2. Then pairs $(x_i^{(1)}, x_i^{(2)})$ and $(x_j^{(1)}, x_j^{(2)})$ for any $i, j \in \{1, \dots, n\}$ are concordant if $(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) > 0$ and discordant if $(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) < 0$.*

From probabilistic point of view, let $X_1^{(c)}, X_2^{(c)}$ be independent copies of X_1, X_2, respectively. Then the probability of concordance, denoted $P(C)$, is defined as

$$P(C) = P[(X_1 - X_1^{(c)})(X_2 - X_2^{(c)}) > 0].$$

Conversely, the probability of discordance, denoted $P(D)$, is defined as

$$P(D) = P[(X_1 - X_1^{(c)})(X_2 - X_2^{(c)}) < 0].$$

Note. Given that X_1, X_2 are continuous, we have $P(C) + P(D) = 1$. Also note that when there is a perfect monotone dependence between X_1 and X_2 we have that $P(C) = 1$, and analogously in case of perfect negative dependence we have $P(D) = 1$.

2.4.1 Kendall's tau

We will narrow our attention to Kendall's τ as our measure of concordance of interest in this work. It is proper to start with its formal definition. Once again probabilistic and sample versions are provided. We will denote the probabilistic version of Kendall's τ by $\tau^{(K)}$, and its sample version $t^{(K)}$. For convenience, we will consider only continuous random variables in this subsection as it facilitates establishing basic properties and their links to copula theory.

Definition 5. *(Nelsen, 2006) Given X_1, X_2 are continuous random variables and using the same notation as in Definition 4, the probabilistic version of Kendall's tau is defined as*

$$\tau^{(K)} = P[(X_1 - X_1^{(c)})(X_2 - X_2^{(c)}) > 0] - P[(X_1 - X_1^{(c)})(X_2 - X_2^{(c)}) < 0].$$

The sample version of Kendall's τ, denoted $t^{(K)}$ for distinction reasons, is defined as

$$t^{(K)} = \frac{N_c - N_d}{N_c + N_d} = \frac{N_c - N_d}{\binom{n}{2}},$$

where

$$N_c = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \mathbb{1}\{(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) > 0\},$$

$$N_d = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \mathbb{1}\{(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) < 0\},$$

i.e. N_c is the number of concordant pairs of the sample and N_d is the numbers of discordant pairs of the sample.

Note. Although the definition above is adapted from Nelsen (2006), the author of this rank correlation measure is Kendall, who first published this measure in 1938 in his paper *A new measure of rank correlation*.

Alternatively, we may write $\tau^{(K)} = P(C) - P(D)$. In the absence of ties, it is also straightforward to realise that $N_c + N_d = \sum_{i=1}^{n} \sum_{j \neq i}^{n} 1 = \frac{n(n-1)}{2} = \binom{n}{2}$. Furthermore, the range of $\tau^{(K)}$ is $[-1, 1]$, when the lower bound is reached when there is a perfect negative dependence among random variables and the converse is true for the upper bound. The following theorem establishes the way how $\tau^{(K)}$ is calculated from the unique copula of continuous random variables.

Theorem 10. *(Nelsen, 2006) Let X_1, X_2 be continuous random variables and C their unique copula representation. Then, the probabilistic version of Kendall's τ is given by*

$$\tau^{(K)} = 4 \iint_{\mathbb{I}^2} C(u, v)\, dC(u, v) - 1$$

Proof. It suffices to link together the Definition 5 with *Theorem 5.1.1* from Nelsen (2006), where we set $C_1 := C, C_2 := C$, which completes the proof.

⊠

Note. As observed by Nelsen (2006), we can interpret the integral in Theorem 10 as the expected value of the function $C(U, V)$, where $C(U, V)$ is the joint distribution function of $U, V \sim Unif(0, 1)$, i.e. we have

$$\tau^{(K)} = 4 \iint_{\mathbb{I}^2} C(u, v)\, dC(u, v) - 1 = 4E[C(U, V)] - 1$$

.

Sometimes, we will need to specify whether we refer to $\tau^{(K)}$ in the copula context or in the context of random variables. When computing $\tau^{(K)}$ from a copula C, we will use the notation $\tau^{(K)}(C)$ since the value of Kendall's τ depends on the copula C (viz. Theorem 10). In a similar fashion, when random variables X_1, X_2 are concerned, we will stick to the notation $\tau^{(K)}(X_1, X_2)$. Note that when X_1, X_2 are continuous random variables, then $\tau^{(K)}(C) = \tau^{(K)}(X_1, X_2)$ by the virtue of Sklar's Theorem and the unique copula representation for the joint distribution of X_1 and X_2.

Note. Having mentioned Spearman's ρ, it feels appropriate to (at least) informally define the measure and state how it can be obtained from copula. Both are adapted from Nelsen (2006). Denoting it by $\rho^{(S)}$, we define

$$\rho^{(S)} = 3\{P[(X_1 - X_1^{(c_1)})(X_2 - X_2^{(c_2)}) > 0] - P[(X_1 - X_1^{(c_1)})(X_2 - X_2^{(c_2)}) < 0]\},$$

where $(X_1^{(c_1)}, X_2^{(c_1)})$, $(X_1^{(c_2)}, X_2^{(c_2)})$ are independent copies of (X_1, X_2). Having continuous random variables X_1, X_2 with their unique copula C, Spearman's ρ is given by

$$\rho^{(S)} = 12 \iint_{\mathbb{I}^2} uv\, dC(u, v) - 3 = 12 \iint_{\mathbb{I}^2} C(u, v)\, du\, dv - 3.$$

For more specific information about $\rho^{(S)}$ and its relationship with $\tau^{(K)}$ refer to Nelsen (2006).

2.4.2 Archimedean copulas and Kendall's tau

Returning to Archimedean copulas, their another convenient feature is that it is easy to calculate $\tau^{(K)}$ given the copula is Archimedean since we only need to evaluate the integral of the generator, i.e. a univariate function, which leads to the evaluation of a simple integral instead of a double integral as the following theorem from Nelsen (2006) states. We will also demonstrate the proof which was adapted from Nelsen (2006) but there were various references made to other parts of the monograph so the decision was made to demonstrate somehow more compact and slightly more detailed proof.

Theorem 11. *Let X_1, X_2 be continuous random variables with an Archimedean copula C generated by a generator function φ. Then*

$$\tau^{(K)} = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt.$$

Proof. Prior to the proof itself, we need to lean on an additional theoretical background. For details refer to *Theorem 4.3.4* and *Corollary 4.3.6* of Nelsen (2006). We will state the following as facts.

Fact 1: Given that C is an Archimedean copula generated by φ, then $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}$, $t \in \mathbb{I}$ denotes the C-measure of the set $\{(u,v) \in \mathbb{I}^2 : \varphi(u) + \varphi(v) \geq \varphi(t)\}$. C-measure is given by $V_C([0,u] \times [0,v]) = C(u,v)$; for more details see pages 26-27 of Nelsen (2006). Note that φ is convex so that it is differentiable almost everywhere so that we can exchange $\varphi'(t)$ for $\varphi'(t^+)$.

Fact 2: K_C from the Fact 1 is the distribution function of the random variable $C(U,V)$, where C (an Archimedean copula) is the joint distribution function of $U, V \sim Unif(0,1)$.

Finally, we can proceed with the proof itself. Utilising the note below Theorem 10 we get that $\tau^{(K)} = 4E[C(U,V)] - 1$. Moreover, making use of Fact 2 that K_C is the distribution function of the random variable $C(U,V)$, we get

$$\tau^{(K)} = 4E[C(U,V)] - 1 = 4\int_0^1 t\,dK_C(t) - 1.$$

Applying per partes on the integral gives

$$\int_0^1 t\,dK_C(t) = [t\;K_C(t)]_0^1 - \int_0^1 K_C(t)\,dt,$$

where

$$[t\;K_C(t)]_0^1 = \left[t^2 - t\frac{\varphi(t)}{\varphi'(t)}\right]_0^1 = 1$$

since $\varphi(1) = 0$ by definition. In addition, we have that

$$\int_0^1 K_C(t)\,dt = \int_0^1 t - \frac{\varphi(t)}{\varphi'(t)}\,dt = \frac{1}{2} - \int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt.$$

Combining it all together, we obtain

$$\tau^{(K)} = 4\left(\frac{1}{2} + \int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt\right) - 1 = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt.$$

⊠

This theorem provides the toolbox for capturing the functional relationship between the copula parameter θ and $\tau^{(K)}$, i.e. one becomes the function of the other and vice versa. It is apparent that different copula families will have different functional relationship with respect to $\tau^{(K)}$. The following table captures such a relationship for our selected families of copulas (Nelsen, 2006). The detailed step-by-step calculation for Clayton and Gumbel copulas is included in the Appendix B.

Clayton	$\tau^{(K)} = \frac{\theta}{\theta+2}$
Frank	$\tau^{(K)} = 1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}\,dt\right)$
Gumbel	$\tau^{(K)} = \frac{\theta-1}{\theta}$

Table 2.3: Functional relationship between $\tau^{(K)}$ and θ for selected copula families.

2.5 Non-continuous issues

Difficulties arise as soon as we leave the continuous settings since the interpretability of $\tau^{(K)}$ is hindered, rendering the inference based on copula models problematic. All of the issues in this section originate from the non-uniqueness issue mentioned earlier.

2.5.1 Equality of Kendall's taus no longer holds

Having joint distribution of non-continuous random variables X_1, X_2, the class $\mathcal{C}$ of copulas fulfilling the equation 2.1 consists of infinitely many copulas. As a consequence, the relationship $\tau^{(K)}(C) = \tau^{(K)}(X_1, X_2)$ no longer holds for every copula in $\mathcal{C}$. In case this class is reasonably small, the copula models could capture the dependence structure of X_1 and X_2 fairly closely (Genest and Nešlehová, 2007). As discussed above, the size of $\mathcal{C}$ is measured by Carley's bounds. First, we compute C_H^-, C_H^+, then we calculate $\tau^{(K)}(C_H^-)$, $\tau^{(K)}(C_H^+)$, respectively. From there, we will get that

$$\forall C \in \mathcal{C} : \tau^{(K)}(C_H^-) \leq \tau^{(K)}(C) \leq \tau^{(K)}(C_H^+),$$

which was shown by Tchen (1980). Given that C_H^-, C_H^+ are respective maximum and minimum extensions of a (sub)copula C fulfilling the equation 2.1, we observe that concordance ordering, referred by Scarsini (1984) as "coherence property," is preserved under non-continuous settings, i.e. $C_H^- \prec_c C \prec_c C_H^+ \implies \tau^{(K)}(C_H^-) \leq \tau^{(K)}(C) \leq \tau^{(K)}(C_H^+)$ in the context of $\tau^{(K)}$ (Tchen, 1980). We notice that now $\tau^{(K)}$ is not a single value but the range of possible values capturing dependence structure between X_1 and X_2, which hinders the intepretability of $\tau^{(K)}$ as the measure of dependence, should the range be too wide. To put it differently, Carley's bounds enable us to measure the breadth of degrees of dependence that possibly depicts the dependence structure between X_1 and X_2 (Genest and Nešlehová, 2007).

Restricting ourselves to discrete margins, we are able to state formulas for calculating $\tau^{(K)}(C_H^-)$, $\tau^{(K)}(C_H^+)$ stated in Genest and Nešlehová (2007), viz. *Proposition 4*.

Theorem 12. *Consider discrete random variables X_1, X_2 with their supports on $\mathbb{N}_0$ whose joint distribution is H and their margins are F and G, respectively. Let $\mathcal{C}$ be the class of copulas such that the equation 2.1 holds for any $C \in \mathcal{C}$ and let $h_{ij} = P[X_1 = i, X_2 = j]$. Then:*

$$\tau^{(K)}(C_H^-) = -1 + 4\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{i-1}\sum_{l=0}^{j-1} h_{kl}h_{ij},$$

$$\tau^{(K)}(C_H^+) = 1 - 4\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{i-1}\sum_{l=j+1}^{\infty} h_{kl}h_{ij},$$

where $C_H^- = \min\{C : C \in \mathcal{C}\}$ and $C_H^+ = \max\{C : C \in \mathcal{C}\}$.

Proof. See *Appendix B* of Genest and Nešlehová (2007). ⊠

These formulas could be then applied to obtain the range of values describing the dependence structure between X_1 and X_2. In case the range in reasonably small, we can determine the dependence structure of X_1 and X_2 based on the range bounds with no significant interpretability problems. On the other hand, when bounds are farther away from each other, the interpretability of dependence structure becomes ambiguous and even impossible if the bounds imply different dependence structure (e.g. a lower bound suggests independence, whereas the upper one suggests positive dependence). Thus, computing the value of $\tau^{(K)}$ (or any other concordance measure) for C_H^- and C_H^+ gives us a valuable insight whether doing inference on a dependence structure in copula modelling is even reasonable given specific data.

2.5.2 Bounds for Kendall's tau are narrower

Another issue we encounter in the non-continuous settings is that the range of values of $\tau^{(K)}$ is narrower that $[-1, 1]$. Having the non-continuous random variables, the probability of a tie is non-zero as opposed to the continuous case. For further purposes, let us denote the probability of a tie of 2 random variables X_1 and X_2 by $P(T)$. Formally, we define the probability of a tie as

$$P(T) = P\left[X_1 = X_1^{(c)} \vee X_2 = X_2^{(c)}\right],$$

where $(X_1^{(c)}, X_2^{(c)})$ is the independent copy of (X_1, X_2) and $\vee$ is the "or" logical operator. We can immediately see from applying the principle of inclusion and exclusion that

$$P(T) = P\left[X_1 = X_1^{(c)}\right] + P\left[X_2 = X_2^{(c)}\right] - P\left[X_1 = X_1^{(c)},\ X_2 = X_2^{(c)}\right]. \tag{2.7}$$

Realising that the probabilistic version of $\tau^{(K)}$ does not account for ties (viz. Definition 5), we derive the $\tau^{(K)}$ range inequality for non-continuous random variables stated by Denuit and Lambert (2005). Recalling that $\tau^{(K)} = P(C) - P(D)$ and that now $P(C) + P(D) + P(T) = 1$ since $P(T) > 0$, we can get that $\tau^{(K)} = 2P(C) + P(T) - 1$. From there it is straightforward that

$$-1 + P(T) \leq \tau^{(K)} \leq 1 - P(T)$$

given that $\tau^{(K)} = P(C) - P(D) \leq P(C) + P(D) = 1 - P(T)$ since $0 < P(C),\ P(D) < 1$. Now, without loss of generality let us assume that $P[X_1 = X_1^{(c)}] \geq P[X_2 = X_2^{(c)}]$. Subsequently, we obtain from (2.7) that

$$P(T) \geq \max\left\{P\left[X_1 = X_1^{(c)}\right], P\left[X_2 = X_2^{(c)}\right]\right\} = P\left[X_1 = X_1^{(c)}\right]$$

since $P[X_2 = X_2^{(c)}] \geq P[X_1 = X_1^{(c)},\ X_2 = X_2^{(c)}]$, which trivially holds. Combining it all together we arrive at the string of inequalities

$$-1 < -1 + \max\left\{P\left[X_1 = X_1^{(c)}\right], P\left[X_2 = X_2^{(c)}\right]\right\} \leq -1 + P(T) \leq \tau^{(K)} \leq$$

$$\leq 1 - P(T) \leq 1 - \max\left\{P\left[X_1 = X_1^{(c)}\right], P\left[X_2 = X_2^{(c)}\right]\right\} < 1 \tag{2.8}$$

where the strict inequalities arise from $P[X_1 = X_1^{(c)}],\ P[X_2 = X_2^{(c)}] > 0$ provided that $X_1,\ X_2$ are non-continuous random variables. As a consequence of this inequality, we can notice that not matter the joint distribution of X_1 and X_2 we always get that $-1 < \tau^{(K)} < 1$ even if there is a perfect monotone dependence between random variables. Moreover, Genest and Nešlehová (2007) argue that $\tau^{(K)}$ and its range both depend on marginal distributions of discrete random variables as opposed to the continuous case when $\tau^{(K)}$ is a margin-free dependence measure and its range is $[-1, 1]$.

Considering the implication of the above result for practical applications, Denuit and Lambert (2005) argue that since "the population version of $\tau^{(K)}$ is restricted to a narrower range than $[-1, 1]$," we must take this into the account when examining the dependence structure and its strength of non-continuous random variables. Examining the string of inequalities 2.8, we can immediately notice that the bigger the probability of a tie, the narrower the range for $\tau^{(K)}$. Narrowing our focus to discrete margins whose support lies on $\mathbb{N}$ (resp. $\mathbb{N}_0$), we can notice that $P(T)$ decreases as the probability mass is spread between higher number of atoms and when the probability masses assigned to a single atoms decreases. Hence, it is apparent that the issue of narrower bounds for $\tau^{(K)}$ range is a far more serious problem for Bernoulli margins than it is for margins governed by Poisson or negative binomial distributions, especially when their respective parameters are such that no single atom is assigned significant probability mass. Moreover, an illustrative example is shown in Denuit and Lambert (2005) that highlights that $\tau^{(K)}$ upper bound of margins $X_1,\ X_2 \sim Poisson(\lambda)$ tends to 1 as $\lambda \to \infty$, which is expected as the probability of tie tends to zero in this case. This example also illustrates the dependence of $\tau^{(K)}$ bounds on marginal distributions. From the practical viewpoint, the seriousness of the issue of the narrower $\tau^{(K)}$ range diminishes when we deal random variables whose probability mass is distributed on many atoms with no single atom being assigned too much probability mass, and when we do not expect extremely strong monotone dependence between random variables.

Apart from allowing for narrower bounds of $\tau^{(K)}$, we may resort to adapting a rescaled version of $\tau^{(K)}$ that account for ties. There are many modified versions of $\tau^{(K)}$ that account for ties; for reference refer to Genest and Nešlehová (2007). It is worth noting that the presence of ties makes the range of $\tau^{(K)}$ variable depending on marginal distribution, therefore, rescaling of $\tau^{(K)}$ that accounts for ties tries to tackle this issue. We will stick to Kendall's τ_b as our rescaled version whose the probabilistic and the sample version are adapted from Nešlehová (2007). Given the definition from Nešlehová (2007), we will restrict only to discrete random variables.

Definition 6. *Consider a pair of discrete random variables X_1, X_2 with finite random samples $(x_1^{(1)}, \dots, x_n^{(1)})$, $(x_1^{(2)}, \dots, x_n^{(2)})$ with r and s distinct values $\xi_1 < \dots < \xi_r$ and $\eta_1 < \dots < \eta_s$, respectively, where $r,\ s \le n$. The probabilistic version of Kendall's τ_b, denoted $\tau_b^{(K)}$, is given as*

$$\tau_b^{(K)} = \frac{\tau^{(K)}}{\sqrt{P[X_1 \neq X_1^{(c)}]P[X_2 \neq X_2^{(c)}]}},$$

where $X_1^{(c)}$ and $X_2^{(c)}$ are independent copies of X_1 and X_2, respectively. The sample version of Kendall's τ_b, denoted $t_b^{(K)}$, is defined as

$$t_b^{(K)} = \frac{N_c - N_d}{\sqrt{\binom{n}{2} - t^{(1)}}\sqrt{\binom{n}{2} - t^{(2)}}},$$

where

$$N_c = \sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbb{1}\{(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) > 0\},$$

$$N_d = \sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbb{1}\{(x_i^{(1)} - x_j^{(1)})(x_i^{(2)} - x_j^{(2)}) < 0\},$$

$$t^{(1)} = \sum_{j=1}^{r} \binom{\sum_{i=1}^{n} \mathbb{1}\{x_i^{(1)} = \xi_j\}}{2},$$

$$t^{(2)} = \sum_{j=1}^{s} \binom{\sum_{i=1}^{n} \mathbb{1}\{x_i^{(2)} = \eta_j\}}{2},$$

i.e. N_c, N_d are numbers of concordant and discordant pairs, respectively, and $t^{(1)}$, $t^{(2)}$ are respective sums of the number of all possible pairs of the number of tied sample values per given distinct sample value for finite samples of X_1 and X_2. By convention, $\binom{a}{b} = 0,\ a < b,\ a, b \in \mathbb{N}_0$.

Note. When X_1, X_2 are continuous random variables we have that $\tau^{(K)} = \tau_b^{(K)}$ and $t^{(K)} = t_b^{(K)}$.

Another way of dealing with ties in random discrete variables has been proposed by Denuit and Lambert (2005) who approached this problem by continuing discrete variables by adding perturbation continuous random variables U_1, U_2, with the domain $(0, 1)$ and strictly increasing cumulative distribution functions (uniform distribution seems as the most obvious choice here). The reason we opted for the approach chosen by Nešlehová (2007) is that sample and probabilistic version of Kendall's τ_b proposed by her coincide, which was also formally stated and proven; for proof refer to *Theorem 14* of Nešlehová (2007). Even though correcting for ties broadens the range of $\tau^{(K)}$, the bounds do not always reach ± 1 when we have perfect monotone dependence as argued by Genest and Nešlehová (2007). Nonetheless, Nešlehová (2007) shows that under more strict conditions bounds ± 1 are guaranteed to be reached when perfect monotonic dependence is encountered, however, the condition $X_2 \overset{a.s.}{=} f(X_1)$, where f is strictly monotonous continuous function defined on the range of X_1, is rather restrictive.

2.5.3 Biased estimation of θ

Furthermore, another problem arises when estimation of copula parameter θ of a given parametric copula family (e.g. any of Archimedean copulas) is concerned. To construct a copula model capturing the dependence structure of random variables X_1, X_2, the estimation of θ is the essential step. When X_1 and X_2 are continuous, the estimation process is straightforward. There are numerous techniques of estimation, but by far, the most convenient technique is so-called inversion method as argued by Genest and Nešlehová (2007). This method makes use of the functional relationship between θ and any concordance measure, i.e. $\tau^{(K)}$ in our case (viz. Table 2.3). This is possible thanks to the fact that the mapping $\tau^{(K)} : \theta \mapsto \tau^{(K)}(C_\theta)$ for a given C_θ from provided parametric copula family (e.g. Clayton) is 1-to-1, which enables inversion, and the fact that the estimation of θ is known to be consistent and asymptotically normal (Genest and Nešlehová, 2007). Provided that X_1, X_2 are continuous with margins F and G, the generalised inverses of margins coincide with ordinary inverses and they are strictly increasing, so $U = F^{(-1)}(X_1)$, $V = G^{(-1)}(X_2)$ are strictly increasing transformations of X_1 and X_2 and we have that $\tau^{(K)}(X_1, X_2) = \tau^{(K)}(U, V)$ given that any measure of concordance, in the sense of Scarsini (1984), is invariant under strictly monotone transformations. Refer to *Theorem 1* of Scarsini (1984) or *Theorem 5.1.9* of Nelsen (2006) for formal statements with proofs, respectively.

However, in the non-continuous case it no longer holds because the probability of ties is non-zero, and therefore, the mapping ceases to be 1-to-1, which makes inversion irreversible. Adapting sample versions that correct for ties such as $t_b^{(K)}$ in the Definition 6 and other proposed in Genest and Nešlehová (2007) does not bear any improvement. Genest and Nešlehová (2007) show that when margins of X_1 and X_2 are discrete, the estimation of θ through inversion method is biased and also inconsistent, even if continuing X_1 and X_2, as proposed by Denuit and Lambert (2005), is concerned. However, this does not necessarily mean that θ is a priori doomed to failure rather than that inference for the dependence parameter θ should not be based on the inversion of $\tau^{(K)}$ or any other concordance measure due to bias (Genest and Nešlehová, 2007).

On the other hand, using standard maximum likelihood estimation could work properly as shown in *Example 13* of Genest and Nešlehová (2007). However, as argued by Genest and Nešlehová (2007), the conditions under which this can be supported by theory are yet to be discovered. Trivedi and Zimmer (2017), restricting to count outcomes only, show in their Monte Carlo simulation study that estimation of θ tends to be unbiased and consistent when the sample is heterogeneous and large enough, and also when regression structure is applied to the sample. The argument behind both is that bias in estimation of θ diminishes as the outcome domain is more covered by count outcomes.

Considering discrete random variables with their support on $\mathbb{N}$ (resp. $\mathbb{N}_0$), the higher mean of outcome variable implies more heterogeneous sample, and it is also apparent that lower probability of ties is encountered with more heterogeneous sample. In a similar fashion, Trivedi and Zimmer (2017) argue that given that means of outcome variables increase "the partition of the unit interval induced by the quantile functions becomes finer as μ [mean] increases, the lack of identification of θ likewise should diminish as μ increases."

As for applying regression structure by adding covariates (be it continuous or discrete), it changes marginal distribution to conditional distribution since "with covariate, the argument to the copula functions are expected means, rather than outcome variables themselves" Trivedi and Zimmer (2017). Trivedi and Zimmer (2017) show that marginals conditioned upon covariates tend to perform better in the estimation of θ even if means are smaller. Yet, asymmetric copulas (e.g. Clayton and Gumbel copulas) needs to be treated with caution as they require larger samples in order for the estimation to be unbiased and consistent, provided we have the benefit of larger means or/and covariates, as has been shown in the simulation study of Trivedi and Zimmer (2017).

Even though modelling dependence structure of non-continuous random variables is fraught with difficulties that have been highlighted above, it is still a valid option. However, modelling and subsequent inference need to be treated with a substantial caution. Estimation of the dependence parameter θ remains a possible approach, yet conditions guaranteeing its theoretical validity are yet to be discovered. Therefore, prior to thinking of constructing a copula model, data must be scrutinized, and appropriateness of applying copula model to this data must be accessed.

3. Application on Real-World Data

After a thorough presentation of theoretical background and discussion about the appropriateness of copula model usage in non-continuous settings, we will provide an application of the aforementioned on the real-world data. We will construct 3 copula models from the same data set. Copula families of our choice are Clayton, Gumbel and Frank copulas, which are all single-parameter Archimedean copulas whose basic properties were introduced in the previous chapter. Our demonstration will concern the case when both random variable are discrete given that the bulk of the theoretical background presented above was mainly concerned with discrete, not non-continuous, random variables, although the generalisation to non-continuous setting should be straightforward in general.

The data set is available at `http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset`, and it originates from Fanaee-T and Gama (2013). The data set concerns bicycle-sharing data from Washington, D.C. from years 2011 and 2012 obtained from Capital Bike Sharing (CBS). The package of data set provides both hourly and daily data, however, we restrict ourselves to the exploration of hourly data only. This data set contains 17 variables, namely the instant (number of index of the entry), date, weather season, year, month, hour, indicator whether it is a holiday or not, weekday number (1–Monday, 7–Sunday), indicator of working day (negation of holiday indicator), categorical indicator of weather situation (4 states), temperature (normalised), feeling temperature (normalised), humidity (normalised), wind speed (normalised), and finally number of registered users per hour, number of casual users per hour (without subscription) and total number of users per hour. We will use copula models to model and infer the dependence structure between number of registered users and number of casual users that are our random variables of interest.

3.1 Data preparation

The original data set concerning hourly data has 17379 entries. However, quite naturally, the behaviour of users is strongly influenced by season, time of the day, whether it is working day or not and weather. As a consequence, it is not possible for such data to follow any standard discrete distributions, which eventually leads to rather wild mixture distribution constructions if one wishes to proceed with this data. Alternative approach would be to make use of the regression structure of the data and proceed with incorporating regression analysis in copula modelling, however, this is beyond the scope of this work. Therefore, in order not to overcomplicate, we decided to impose several conditions on data that provides less dependent data on variable that are not of interest in this work.

We restricted ourselves to hourly data of number of users that were recorder on working days, from April to November (including both) and from 9 a.m. to 4 p.m. and from 8p.m. until midnight (including all boundary hours) and under the condition that the weather situation is cloudy but without precipitation (labeled as factor 2 in the data column about weather situation). The rationale behind this is that user behaviour on non-working days is totally different from working-days, and using bicycle as the means of transportation is not natural to winter months due to unfavourable conditions (snowy and icy roads, low temperatures). Furthermore, it is likely that registered users use bicycles to commute to and from work which can be seen in a massive surge in number of registered users from 7 a.m. to 8 a.m. and from 5 p.m. to 7 p.m., and as for the restriction on weather situation, no possible logical explanation is provided but that no other conditions (including no condition at all) regarding weather situation allowed the hourly number of registered and casual users to follow standard discrete distributions (Poisson and negative binomial), so it is an artificial restriction. Hopefully, the reader will forgive us such a blatant data manipulation given that the purpose of this section is to solely demonstrate copula modelling on “some” discrete data, and the author decided from real-world rather than simulated data in order to introduce one possible field of application of discrete copula modelling. This restricted data set has 1025 entries, which makes it also more computationally convenient.

3.2 Data characteristics

Narrowing our focus to the restricted data set, which have 1025 observations, we are ready to provide basic characteristics and do a basic exploratory analysis of observed data of our interest. First our all, both number of registered and casual users are discrete random variables with infinite support, i.e. their support is $\mathbb{N}_0$. Hence, it is natural to assume that this random variables have either Poisson or negative binomial distribution. The following table, scatterplot and boxplot provide us with both numerical and visual characteristics of observed data.

Users	min	max	1st Q	m_X	3rd Q	μ	σ	σ^2
Registered	17	454	112	158	209	169.082	78.009	6085.405
Casual	2	158	20	35	54	39.213	24.841	617.058

Table 3.1: Descriptive characteristics of observed data

It is visible from Table 3.1 that there is significantly more registered users than casual users (on average almost 3 times more) and the hourly number of registered users is more volatile. Moreover, we can observe from Figure 3.1 that the observed data from both user categories that upper whisker is significantly longer than lower one and also that all outliers of both categories lies beyond upper whisker.

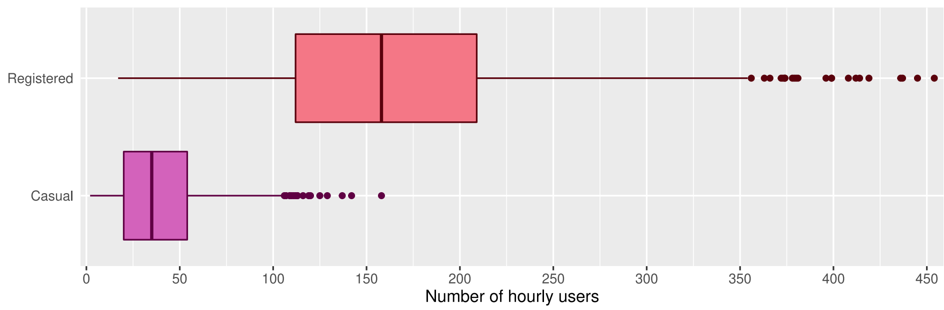

Figure 3.1: Boxplot: hourly numbers of registered and casual users

As for the scatterplot of data (Figure 3.2) which plots observed hourly pairs of registered and casual users, we can observe an apparent link between larger numbers of registered users with larger numbers of casual users, and the same hold also for lower numbers of both categories. This implies that we should expect a positive dependence between these two random variables. Because of this, our 3 selected copula families remain a viable choices for modelling as all can capture and model a positive dependence structure. Furthermore, we can notice the clustering of observed pairs in the bottom right corner, whereas observed pairs become more dispersed as we move to the region with a higher number of users from both categories. The clustering has a natural interpretation given that the number of bicycle users can be only non-negative, so clustering near zero makes sense and can be interpreted as a lower demand for bicycles from both user categories.

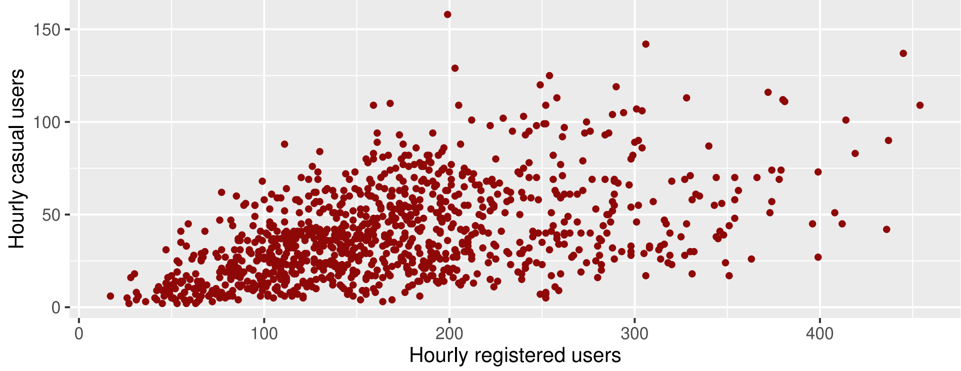

Figure 3.2: Scatterplot: hourly data of registered and causal users

3.3 Methodology

In order to fit copula model for the given discrete data, we will proceed as follows. First, we will compute number of concordant pairs, number of discordant pairs and number of ties, from where can directly determine sample values of Kendall's tau, both its original version $t^{(K)}$ as well as the version adjusted for ties $t_b^{(K)}$, from formulas in Definition 5, Definition 6, respectively. Afterwards, we obtain the sample probability of a tie, denoted $P(T_e)$ from the equation

$$P(T_e) = \frac{\binom{n}{2} - N_c - N_d}{\binom{n}{2}}.$$

Having calculated $P(T_e)$, we can determine possible bounds for $\tau^{(K)}$ obtained from the sting of inequalities (2.8), and evaluate how problematic is the narrower range for $\tau^{(K)}$ in our case. Note that we are not able to obtain the probabilistic version of the probability of a tie $P(T)$ since $P(T) = P[X_1 = X_1^{(c)}] + P[X_2 = X_2^{(c)}] - P[X_1 = X_1^{(c)}, X_2 = X_2^{(c)}]$ given that we do not know the value of $P[X_1 = X_1^{(c)}, X_2 = X_2^{(c)}]$. Having obtained empirical values of Kendall's tau and tau-b, i.e. $t^{(K)}, t_b^{(K)}$, we proceed with calculating the empirical value of Kendall's tau for upper and lower Carley's bounds. The implementation of the algorithm is provided in Appendix D. It is implemented in a straightforward fashion from Theorem 12.

As for the estimation of a selected copula, we will used so-called *inference functions for margins* (IFM) method proposed by Joe (1997), which is a 2-step method. Restricting ourselves to single-parameter copula families, suppose that we have a copula C from a parametric family of copulas C_θ which we want to fit to observed data $((x_1^{(1)}, x_1^{(2)}), \ldots, (x_n^{(1)}, x_n^{(2)}))$ originating from random vector (X_1, X_2), and that there exists a copula density, denoted $c = \frac{\partial^2 C(u,v)}{\partial u \partial v}, \forall C \in C_\theta$, i.e. C_θ is absolutely continuous (Nelsen, 2006). First, we estimate parameters of marginal distributions using standard maximum likelihood estimation (MLE) method. After the first step, when we estimated and tested the hypothesis (by Pearson χ^2 goodness of fit test) that indeed $X_1 \sim F,\ X_2 \sim G$, we will maximize the copula log-likelihood function with respect to θ. Provided that the joint density (or in the sense of discrete random variable joint pmf) of X_1 and X_2 is given as

$$h(x^{(1)}, x^{(2)}; \boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \theta) = c(F_1(x^{(1)}; \boldsymbol{\alpha_1}), F_2(x^{(2)}; \boldsymbol{\alpha_2}; \theta) f_1(x^{(1)}; \boldsymbol{\alpha_1}) f_2(x^{(2)}, \boldsymbol{\alpha_2}),$$

where $f_1,\ f_2$ are respective densities (pmfs) of marginal distributions. We have obtained by standard MLE that

$$\boldsymbol{\hat{\alpha}_i} = \underset{\boldsymbol{\alpha_i} \in \boldsymbol{\Omega}}{\arg\max} \sum_{j=1}^{n} \log(f_i(x_j^{(i)}; \boldsymbol{\alpha_i})),\ i = 1, 2$$

where $\boldsymbol{\hat{\alpha}_i}$ represents vector of estimated parameters of a given distribution. The second step is to set $\boldsymbol{\alpha_1} := \boldsymbol{\hat{\alpha}_1},\ \boldsymbol{\alpha_2} := \boldsymbol{\hat{\alpha}_2}$ and compute

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max} \sum_{j=1}^{n} h(x_j^{(1)}, x_j^{(2)}; \boldsymbol{\hat{\alpha}_1}, \boldsymbol{\hat{\alpha}_2}, \theta).$$

Note. It is also possible to do so-called full maximum likelihood estimation by maximizing $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \theta$ at the same time, however, this approach is much more computationally difficult. For details regarding the efficiency of IFM and its comparison to full-likelihood method refer to Joe (2005).

3.4 Preparatory work

Using the R language for computation, we will make use of inbuilt functions *ConDisPairs*, *KendallTauA* and *KendallTauB* from the R package *DescTools* to compute $t^{(K)}$, $t^{(K)}$, $P(T)$ (Andri, S. et mult. al., 2021). Having $n = 1025$ and using above functions we get $\binom{1025}{2} = 524800$, $N_c = 351813, N_d = 164362$,

$$P(T_e) = 0.01643, \tag{3.1}$$

where sample probability of a tie was rounded to 5 decimal places. Moreover, we obtain

$$t^{(K)} = 0.35719, \tag{3.2}$$

$$t_b^{(K)} = 0.36016 \tag{3.3}$$

$$\left[t^{(K)}(C_H^-), t^{(K)}(C_H^+)\right] = [0.33944, 0.37424] \tag{3.4}$$

Note. All non-integer results are rounded to 5 decimal places.

We can see that there is indeed positive dependence between hourly number of registered and casual users. High heterogeneity of data ensures that even though we have discrete data, the probability of a tie is considerably small, which means that possible bounds for $\tau^{(K)}$ remain wide enough for practical application, i.e. $-1 + P(T_e) = -0.98357 \geq \tau^{(K)} \geq 0.98357 = 1 - P(T_e)$. Moreover, thank to the low probability of a tie the adjusted version of Kendall's τ (Kendall's τ_b) is only slightly different from the original version. Further both sample versions of Kendall's τ are contained within bounds (3.4) are are close to each other, which allows for considerably good interpretability of dependence structure captured by copulas for this case. Although only sample version of bounds is provided, it is intuitive to expect that probabilistic version would provided even narrower range of $[\tau^{(K)}(C_H^-, \tau^{(K)}(C_H^+))]$ provided that the joint distribution lies almost surely on more atoms (eventually on infinite number of atoms) with lower probability mass per single atoms than observed finite values. However, calculation of the probabilistic version is beyond the scope of this work.

Given the nature of data (hourly number of registered and casual users), Poisson and negative binomial distributions have been selected as the potential candidate distributions for our random variables. We make use of the function *fitdist* from R package *fitdistrplus* in order to estimate their parameters using MLE method (Delignette-Muller and Dutang, 2015). The following table provides estimated parameters obtained from MLE and result of χ^2 goodness of fit (GoF) test (with significance level $\alpha = 0.05$) for our selected distribution candidates.

rv	dist	params	df	χ^2 test	$\chi^2_{df}(0.95)$	p-value
reg	Pois	$\hat{\lambda} = 169.082$	454	$> 10^{40}$	504.675	$< 2.2^{-16}$
cas	Pois	$\hat{\lambda} = 39.213$	158	$> 10^{40}$	188.332	$< 2.2^{-16}$
reg	NB	$\hat{n} = 4.72679,\ \hat{p} = 0.02719$	454	450.202	504.675	0.542
cas	NB	$\hat{n} = 2.48957,\ \hat{p} = 0.05970$	158	151.881	188.332	0.622

Table 3.2: Results of MLE and χ^2 GoF Test

Note. Estimated parameter of Poisson distribution for both random variables were rounded to 3 decimal places, whereas estimated parameters of negative binomial were rounded to 5 decimal places. $\chi^2_{df}(0.95)$ is 95th percentile of χ^2_{df} distribution with degrees of freedom being equal to "df."

Note. We work with standard parametrization of Poisson distribution, i.e. given that $X \sim Pois(\lambda),\ \lambda > 0$ we have

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!},\ x = 0, 1, 2, \dots .$$

Analogously, if $X \sim NB(n, p),\ 0 < p < 1,\ r > 0$ we have

$$P[X = x] = \frac{\Gamma(x + n)}{\Gamma(n)x!} p^n (1 - p)^x,\ x = 0, 1, 2, \dots$$

From Table 3.2 we can immediately see that none of random variables follows Poisson distribution as hypothesis is unanimously rejected. On the other hand, we cannot reject that random variables of hourly number of register and casual users originate from $NB(4.72679, 0.02719),\ NB(2.48957, 0.05970)$, respectively.

3.5 Fitting copula models

Finally, we proceed with estimation of selected copula models though IFM, provided that $X_1 \sim NB(4.72679, 0.02719),\ X_2 \sim NB(2.48957, 0.05970)$. R package *copula* and its function *fitCopula* was used to estimate dependence parameter θ of Clayton, Gumbel and Frank copulas (Hofert et al., 2020). In order for *fitCopula* to fit a copula to observed data, observations needed to be transformed by their respective distributions functions that have been estimated above. The scatterplot after this transformation is provided below. It can be noticed that due to the discrete nature of data, hence presence of ties, we can observe that points are scatter on a mesh structure, i.e. several points could be seen lying in horizontal and vertical lines. It is clearly visible in areas of the plot that have lower concentration of points, e.g. the bottom-right quadrant.

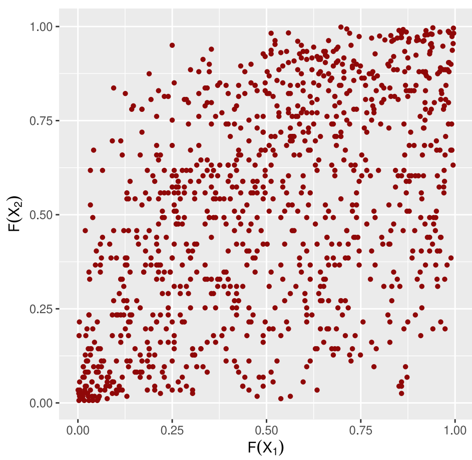

Figure 3.3: Scatterplot: transformed observed data

We can notice significantly higher concentration of points in the bottom-right corner, indicating that observed data (resp. its dependence) may have a slight right-tail dependence. Because of this Clayton copula should fit the observed data best, as its has lower-tail dependence, whereas Gumbel copula has upper-tail dependence, although less pronounced than Clayton, and Frank copula is radially symmetric and experiences no tail-dependence (McNeil et al., 2015). The following table confirm to an extent this notion, when we compare log-likelihood values and AIC of fitted copulas.

copula family	$\hat{\theta}$	log-likelihood	AIC	$\tau_{inv}^{(K)}$
Clayton	0.89132	170.078	-338.156	0.30827
Gumbel	1.46358	137.450	-272.900	0.31674
Frank	3.52720	152.578	-303.157	0.35148

Table 3.3: Fitted copula parameters and characteristics

Note. $\hat{\theta}$ and $\tau_{inv}^{(K)}$ are rounded to 5 decimal places, whereas AIC and log-likelihood are rounded to 3 decimal places. Given that the number of estimated parameters is same for all copula models, the relationship between log-likelihood and AIC is straightforward and log-likelihood is enough for our interpretation purposes.

The intuition that Clayton copula would fit the observed data best (and also that the worst fit should be Gumbel copula) was confirmed by log-likelihood values in the table, given slight right-tail dependence observed in data. This is also visually demonstrated by Figure 3.4 which compares simulation of 5000 observations from each fitted copula model against transformed observed data $\left((F(x_1^{(1)}), G(x_1^{(2)})), \ldots, (F(x_n^{(1)}), G(x_n^{(2)}))\right)$. Concerned scatterplots have points that are simulated from a given copula colored in dark red and these points are set only half-opaque, which makes it easier to distinguish between regions with a higher concentration of simulated points (dark red) and regions with a lower concentration (pale red). The author encourages reader to refer to Appendix C where the same plot are displayed but in full-page size. The software implementation utilised function *rCopula* from the R package *copula* (Hofert et al., 2020).

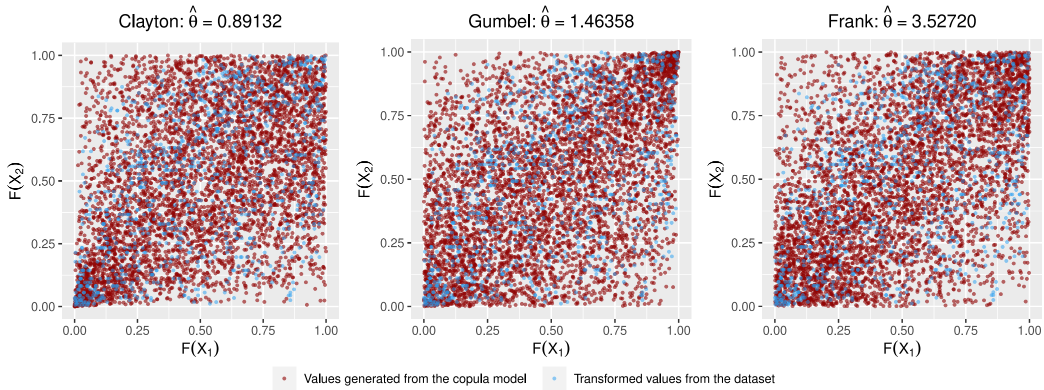

Figure 3.4: Simulated data from fitted copulas vs transformed observed data

However, from log-likelihood (or AIC) we only know which copula fits the observed data best, but we do not know whether the fitted copula represents an adequate model. Given that a standard goodness of fit test for copulas may not be appropriate in non-continuous settings, Yang et al. (2019) proposed a non-parametric estimator to tackle this issue, however, it is beyond the scope of this paper as it is concerned with kernel estimation.

We conclude this chapter with the remark about $\tau^{(K)}$ (denoted $\tau_{inv}^{(K)}$) obtained from the dependence parameter θ through functional relationships stated in Table 2.3; this method is also called as "inversion method." As we can see in Table 3.3 values of $\tau_{inv}^{(K)}$ tend to differ from sample values (3.2),(3.3), and $\tau_{inv}^{(K)}$ for Clayton and Gumbel copulas do not even lie within the range of values of Kendall's tau for Carley's bounds (although only sample bounds were calculated, but probabilistic bounds are likely to lie within sample bounds). This shows that the estimation of $\tau^{(K)}$ from copula model for discrete data is prone to bias.

Note. Note that Frank copula's $\tau_{inv}^{(K)}$ lies (as the only one) within the interval (3.4), and it can also be observed that apart from right-tail dependence, it fits better to observed data than Clayton copula since the density of observations generated from Frank copula is higher around diagonal suggesting higher positive dependence than Clayton copula, which can be also observed from comparing their respective $\tau_{inv}^{(K)}$, although we should be beware of possible bias.

Genest and Nešlehová (2007) argue that it is caused by the fact that $\tau^{(K)} : \theta \mapsto \tau^{(K)}(C_\theta)$ is no longer 1-to-1 relation, which is caused by non-uniqueness issue. Indeed, the bias issue is demonstrated on our data set when method of inversion is concerned, and therefore we should refrain from considering values $\tau_{inv}^{(K)}$ as true values of $\tau^{(K)}$ and rather make concession and reconcile that the true value of $\tau^{(K)}$ remains unknown to us and the best measure we have in terms of the dependence structure is an interval $[\tau^{(K)}(C_H^-), \tau^{(K)}(C_H^+)]$ which still is of a certain interpretation value as long as it is narrow enough, which was also our case (although we only managed to compute sample bounds).

4. Conclusion

The objective of this work was to introduce the reader to the concept of copulas and the theory that makes it an attractive tool for modelling of dependence structures as long as we have continuous settings. Primary sources of difficulty in non-continuous settings were introduced along with possible remedies that prevented us from perceiving copula modelling for non-continuous data a priori invalid. Although rigorous theoretical background guaranteeing copula modelling for is yet to be presented, several empirical findings enable us to apply copulas to non-continuous data when additional care and caution is applied in terms of modelling and inference. The discussion regarding pitfalls and remedies of copula modelling and inference for non-continuous data was demonstrated on discrete real-world data set along with possible methodology framework how to proceed when discrete data are concerned. The demonstration on real-world discrete data has shown that copula modelling remain a viable option as long as we are able to make concessions in terms of inference, and approach each data set case-by-case with additional care and caution.

Furthermore, the author deems it is appropriate to give an account of his contribution, no matter how minuscule. The author of this work provided several modified, more detailed versions of proofs that he believed deserved to be stated in a more detailed way. Especially, he relishes the provision of detailed proofs of preliminary lemmas whose proof were either non-existent or not well-developed and provision of well-structured and easy-to-follow proofs that were of a mere reference nature in original texts. The coherence of the work is reinforced by a uniform mathematical notation contrary to the notation in the cited literature which ranged vastly from one author to the other. Last but not least, the section concerned with non-continuous issues provides a compact discussion about each of the encountered issues and their possible remedies from various authors complementing their findings, which may serve as a terminus a quo for a reader who wishes to delve deeper in the universe of copulas for non-continuous random variables.

As for the practical part, the author provided his own implementation for the computation of Kendall's τ for Carley's bounds. Nonetheless the algorithm was implemented rather inefficiently, using only heuristics to make it slightly more efficient. Moreover, the algorithm was only implemented for the sample version. Hence, another possible future improvement would be to implement the more robust function which is able to calculate both the sample and probabilistic version. As for possible future research the author deems beneficial to examine other concordance measures, their properties in non-continuous settings, and their relationship with Kendall's τ under such conditions. Furthermore, other possible topics for future research stemming directly from this work include expanding theory from discrete to non-continuous settings (mixture distributions), expanding theory to higher dimensions, improving methodology framework based on future theoretical findings and real-world applications to higher dimensions and/or mixture data.

Copulas have enjoyed wide-spread use in real-world applications when continuous random variables are concerned. This work provides a glimpse into issues we encounter when we leave continuous settings, and provides an account on seriousness of these issues both theoretically and practically (demonstrated on real data). Moreover, it contemplates that when additional care and caution is applied, the copula modelling for non-continuous random variables represents a valid option in terms of modelling and inference, despite the fact that the proper theoretical groundwork is yet to be presented. As soon as this obstacle is overcome, the use of copulas in fields such as insurance, where count data are dealt with, and operational risk management, where use of mixture distributions is popular, is set to became as popular and widespread as the use of copulas in the continuous environment such as credit or market risk management (Genest and Nešlehová, 2007).

A. Calculating Limiting Cases for Clayton Copula Depending on θ

Detailed calculation of the limiting cases of Clayton copula, i.e.

$$C_\theta(u,v) = (\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}},\ u, v \in [0,1]$$

for $\theta \to -1,\ 0,\ \infty$. One by one we get:

1. $\theta \to -1$:

$$\lim_{\theta \to -1} C_\theta(u,v) = \lim_{\theta \to -1} (\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}} =$$

$$= \max\{u + v - 1, 0\} = W(u,v)$$

2. $\theta \to 0$: We only need to consider the case

$$(\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}} = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}.$$

The limit is computed as:

$$\lim_{\theta \to 0} C_\theta(u,v) = \lim_{\theta \to 0} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} = \lim_{\theta \to 0} \exp\{-\frac{1}{\theta}\log(u^{-\theta} + v^{-\theta} - 1)\} =$$

$$= \exp\{\lim_{\theta \to 0} -\frac{1}{\theta}\log(u^{-\theta} + v^{-\theta} - 1)\} \overset{\frac{0}{0}}{=} exp\{\lim_{\theta \to 0} \frac{u^{-\theta}\log(u) + v^{-\theta}\log(v)}{u^{-\theta} + v^{-\theta} - 1}\} =$$

$$= exp\{\log(u) + \log(v)\} = uv = \Pi(u,v)$$

The "$\frac{0}{0}$" denotes that we encountered 0 over 0 limit. Hence, L'Hopital's rule was used. Note that we had to apply $A \mapsto \exp(\log(A))$ transformation in order to solve the limit.

3. $\theta \to \infty$: Analogously as above, we assume

$$(\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}} = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}.$$

The limit then yields:

$$\lim_{\theta \to \infty} C_\theta(u,v) = \lim_{\theta \to \infty} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} = \lim_{\theta \to \infty} \exp\{-\frac{1}{\theta}\log(u^{-\theta} + v^{-\theta} - 1)\} =$$

$$= \exp\{\lim_{\theta \to \infty} -\frac{1}{\theta}\log(u^{-\theta} + v^{-\theta} - 1)\} \overset{\frac{\infty}{\infty}}{=} \exp\{\frac{u^{-\theta}\log(u) + v^{-\theta}\log(v)}{u^{-\theta} + v^{-\theta} - 1}\} =$$

$$= \exp\{\lim_{\theta \to \infty} (\frac{u^{-\theta}\log(u)}{u^{-\theta} + v^{-\theta} - 1} + \frac{v^{-\theta}\log(v)}{u^{-\theta} + v^{-\theta} - 1})\} =$$

$$= \exp\{\lim_{\theta \to \infty} (\frac{\log(u)}{1 + (\frac{u}{v})^\theta - u^\theta} + \frac{\log(v)}{1 + (\frac{v}{u})^\theta - v^\theta})\}$$

Now, let us focus on the expression inside the exponential and 3 possible cases of u and v:

(a) $u = v$:

$$\lim_{\theta\to\infty}(\frac{\log(u)}{1+(\frac{u}{v})^\theta - u^\theta} + \frac{\log(v)}{1+(\frac{v}{u})^\theta - v^\theta}) = \lim_{\theta\to\infty} 2\frac{log(u)}{2-u^\theta} = \log(u)$$

from which we have that

$$\lim_{\theta\to\infty} C_\theta(u,v) = u = v.$$

(b) $u < v$:

$$\lim_{\theta\to\infty}(\frac{\log(u)}{1+(\frac{u}{v})^\theta - u^\theta} + \frac{\log(v)}{1+(\frac{v}{u})^\theta - v^\theta}) = \log(u)$$

from which we have that

$$\lim_{\theta\to\infty} C_\theta(u,v) = u$$

since $u < v \implies (\frac{u}{v})^\theta \to 0, u^\theta \to \infty, (\frac{v}{u})^\theta \to \infty$, and either for $v < 1$ we have $v^\theta \to 0$ or for $v = 1$ we have $v^\theta \to 1$.

(c) $u > v$:

$$\lim_{\theta\to\infty} C_\theta(u,v) = v$$

because we have analogous situation as in (b).

Combining (a), (b) and (c), we get

$$\lim_{\theta\to\infty} C_\theta(u,v) = \min(u,v) = M(u,v).$$

Similarly as the limiting case $\theta \to 0$, we used L'Hopital's rule (encountered limit expression "$\frac{\infty}{\infty}$"), and we also applied $A \mapsto \exp(\log(A))$ transformation in order to solve the limit.

B. Calculating Kendall's tau – Clayton and Gumbel Copula

We will make use of Theorem 11, which provides us with all sufficient theory related to copulas required for computation. The calculation of $\tau^{(K)}$ expressed in terms of copula dependence parameter θ for Clayton and copulas is provided below.

(a) Clayton copula: From Table 2.1 we get that $\varphi(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$, from which we directly obtain that

$$\varphi'(t) = -t^{-\theta-1},$$

and by the virtue of Theorem 11 we compute

$$\tau^{(K)} = 1 + \frac{4}{\theta}\int_0^1 \frac{t^{-\theta}-1}{-t^{-\theta-1}}\,dt = 1 + \frac{4}{\theta}\int_0^1 t\frac{1-t^{-\theta}}{t^{-\theta}}\,dt = 1 + \frac{4}{\theta}\int_0^1 t^{\theta+1} - t\,dt$$

$$= 1 + \frac{4}{\theta}\left[\frac{t^{\theta+2}}{\theta+2} - \frac{t^2}{2}\right]_0^1 = 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) = 1 + \frac{4}{\theta}\left(\frac{-\theta}{2(\theta+2)}\right) = 1 - \frac{2}{\theta+2}$$

$$= \frac{\theta}{\theta+2}$$

for $\theta \in [-1, \infty) \setminus \{0\}$

(b) Gumbel copula:

$$\varphi(t) = (-\log(t))^{\theta}$$

from Table 2.1. Therefore,

$$\varphi'(t) = -\frac{\theta(-\log(t))^{\theta-1}}{t}$$

given that $\theta \geq 1$ and

$$\tau^{(K)} = 1 + \frac{4}{\theta}\int_0^1 t\ \log(t)\,dt = 1 + \frac{4}{\theta}\left(\left[\frac{t^2\log(t)}{2}\right]_0^1 - \int_0^1 \frac{t}{2}\,dt\right) = 1 + \frac{4}{\theta}\ \left(-\frac{1}{4}\right)$$

$$= \frac{\theta-1}{\theta},$$

where in the second equality we used per partes with $u = \log(t)$, $v' = t$, and we evaluate

$$\left[\frac{t^2\log(t)}{2}\right]_0^1 = \frac{1^2\log(1)}{2} - \lim_{t\to 0^+}\frac{t^2\log(t)}{2} = 0 + \frac{1}{2}\lim_{t\to 0^+}\frac{-\log(t)}{\frac{1}{t^2}}$$

$$\overset{\frac{\infty}{\infty}}{=} \frac{1}{2}\lim_{t\to 0^+}\frac{\frac{1}{t}}{\frac{2}{t^3}} = \frac{1}{4}\lim_{t\to 0^+} t^2 = 0$$

thanks to L'Hopital's rule (viz "$\frac{\infty}{\infty}$").

C. Detailed Plots for Fitted Copula Models

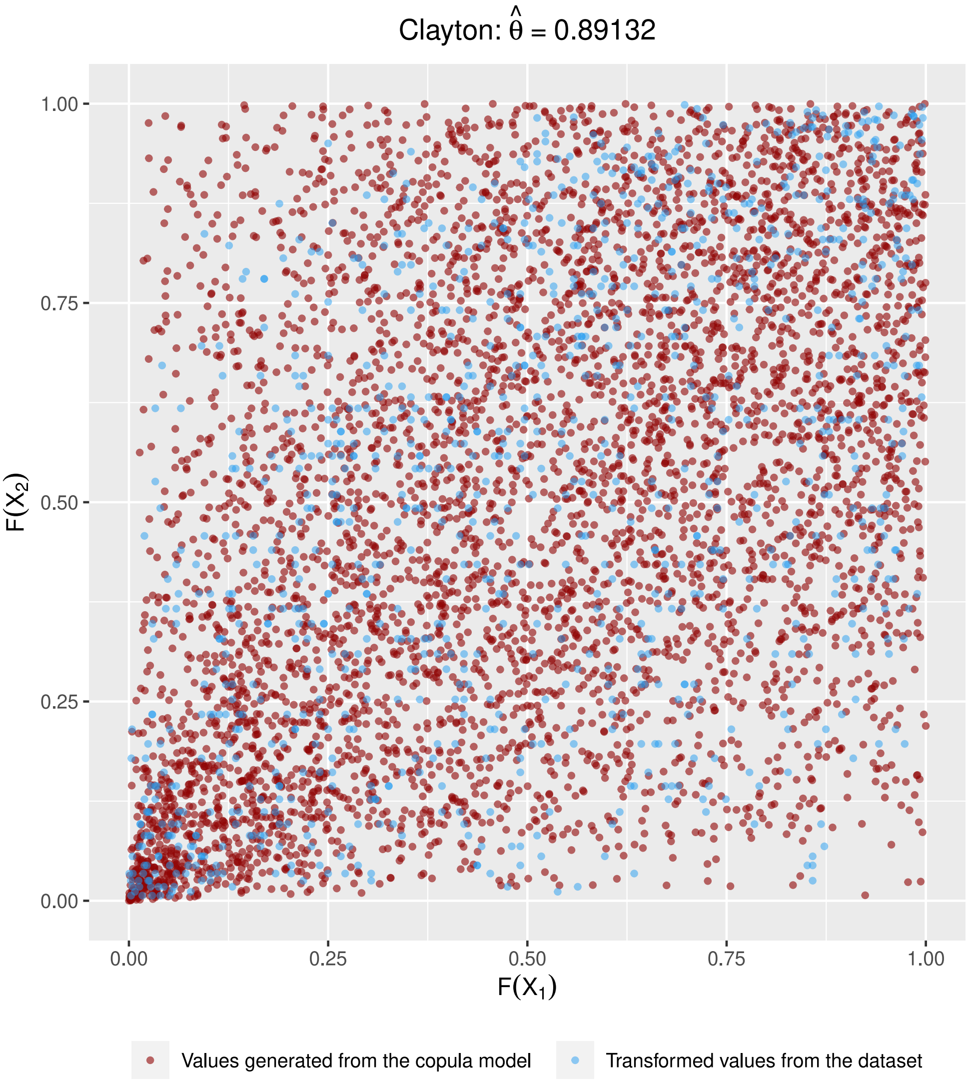

Figure C.1: 5000 simulated points from Clayton copula vs observed data

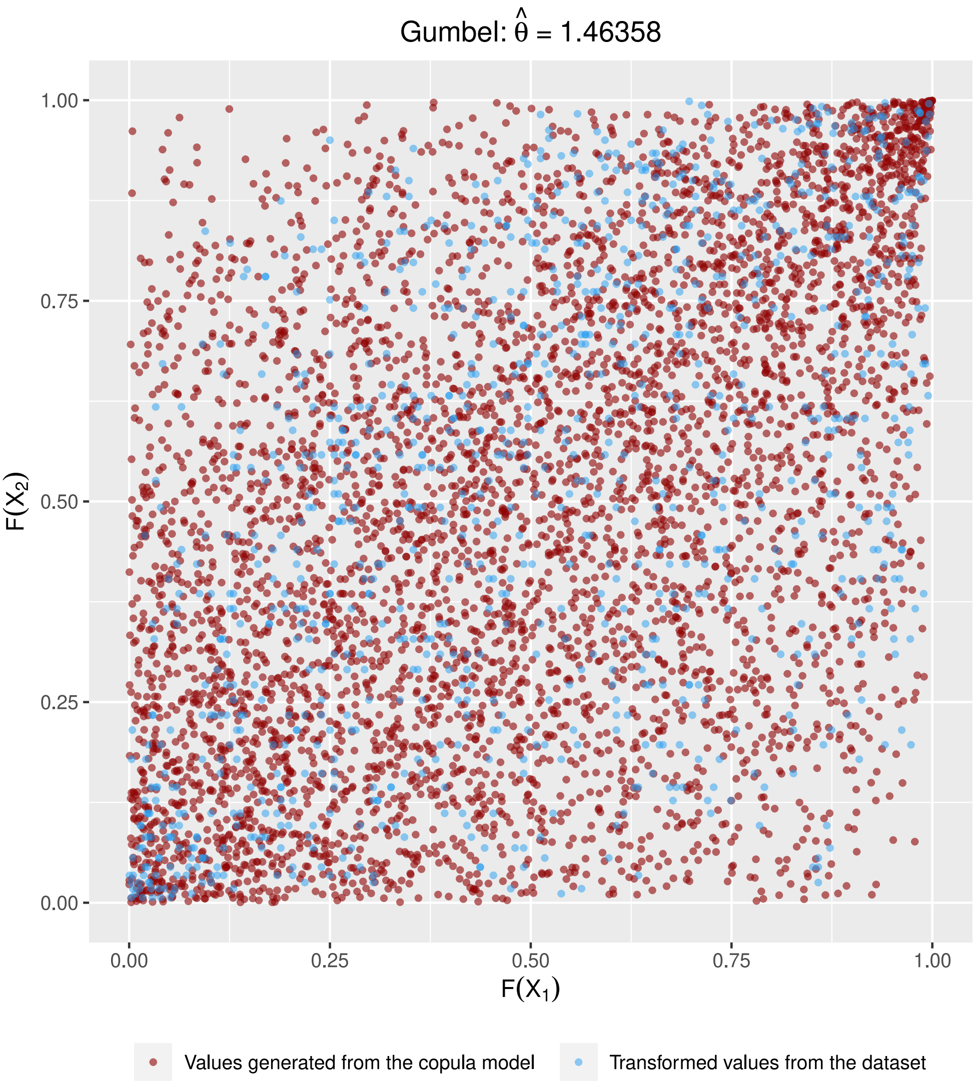

Figure C.2: 5000 simulated points from Gumbel copula vs observed data

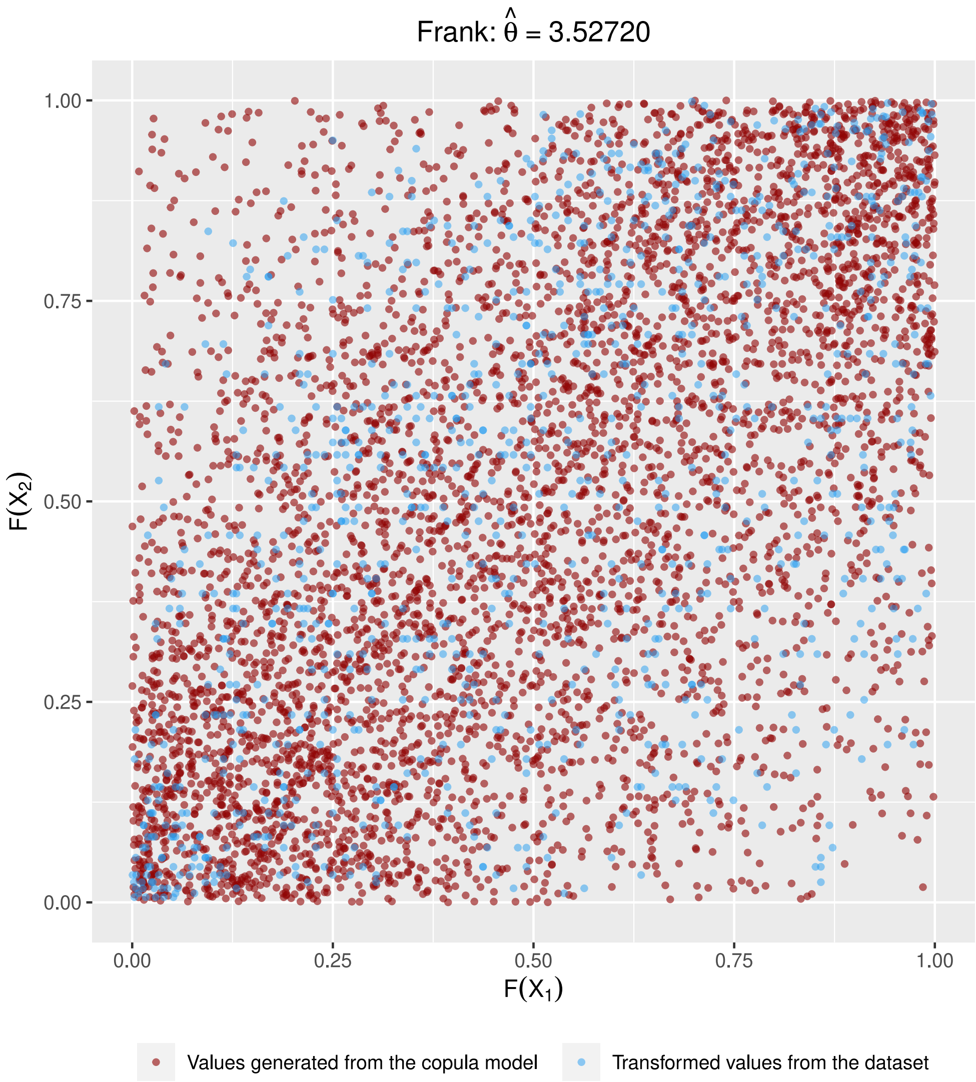

Figure C.3: 5000 simulated points from Frank copula vs observed data

D. Selected R Code Chunks

Note. Note that R indexes its data objects from 1, not from 0 (e.g. Python). Input for both functions computing $\tau^{(K)}$ value of upper/lower Carley's bound requires to be a "table" object with the structure of a contingency table. The subsequent output is a number of type "numeric."

```
tau.lowerCarley = function(tbl){
  #rescalling to empirical 'probability' contingency table (if
   not already)
  tbl = tbl / sum(tbl)
  x = dim(tbl)[1]
  y = dim(tbl)[2]
  s = 0
  for (i in 1:x) {
    for (j in 1:y){
      #tbl[i,j] "\neq" 0 (heuristics to make the code faster)
      if(i>1 & tbl[i,j] != 0) {
        for (k in 1:(i-1)){
          if (j>1) {
            for (l in (1:(j-1))){
              s = s + tbl[i,j]*tbl[k,l]
            }
          }
        }
      }
    }
  }
  res = -1+4*s
  return(res)
}
```

Listing D.1: Function for Kendall's tau of upper Carley's bound computation

Function for calculating the value of Kendall's tau for empirical lower Carey's bound.

```
tau.upperCarley = function(tbl){
  #rescalling to empirical 'probability' contingency table (if
    not already)
  tbl = tbl / sum(tbl)
  x = dim(tbl)[1]
  y = dim(tbl)[2]
  s=0
  for (i in 1:x) {
    for (j in 1:y){
      if (i>2 & tbl[i,j]!=0) {
        #tbl[i,j] "\neq" 0 (heuristics to make the code faster)
        for (k in 1:(i-1)){
          if (j <= (y-1)) {
            for (l in (j+1):y){
              s = s + tbl[i,j]*tbl[k,l]
            }
          }
        }
      }
    }
  }
  res = 1-4*s
  return(res)
}
```

Listing D.2: Function for Kendall's tau of lower Carley's bound computation

```
#probability-wise contingency table
cont.tbl = table(df2$registered, df2$casual) / length(df2$
    registered)
tau.up = tau.upperCarley(cont.tbl)
tau.lo = tau.lowerCarley(cont.tbl)
```

Listing D.3: Using custom functions to compute $[t^{(K)}(C_H^-), t^{(K)}(C_H^+)]$

```
#libraries:
library(copula) #fitting copula (IFM) & sample generation
library(fitdistrplus) #MLE

#loading data
rd.h = read.csv("hour.csv", header = T)
df2 = rd.h[rd.h$workingday==1 & (rd.h$mnth>=4 & rd.h$mnth <= 11)
    &
            (rd.h$hr>=20 | (rd.h$hr>=9 & rd.h$hr <=16)) & rd.h$
    weathersit==2, ]

#fitting data (MLE)
reg.nb = fitdist(df2$registered, "nbinom")
cas.nb = fitdist(df2$casual, "nbinom")
reg.p = fitdist(df2$registered, "pois")
cas.p = fitdist(df2$casual, "pois")

#converting data from (x_i, y_i) to  (F(x_i),G(y_i))
reg01 = pnbinom(df2$registered, size = coef(reg.nb)[1],
                mu = coef(reg.nb)[2])
cas01 = pnbinom(df2$casual, size = coef(cas.nb)[1],
                mu = coef(cas.nb)[2])
data01 = cbind(reg01,cas01)

#fitting copulas (IFM)
#various optimizers tried - all yielded the same result
cl = fitCopula(claytonCopula(), data = data01, method = "ml",
               lower = 0, upper = 10 , optim.method="Brent")
gu = fitCopula(gumbelCopula(), data = data01, method = "ml",
               lower = 0, upper = 10, optim.method="Brent")
fr = fitCopula(frankCopula(), data = data01, method = "ml",
               lower = 0, upper = 10, optim.method="Brent")

#sample generation from copula models
set.seed(123)
n.obs = 5000
cl.sample = rCopula(n.obs, cl@copula)
gu.sample = rCopula(n.obs, gu@copula)
fr.sample = rCopula(n.obs, fr@copula)
```

Listing D.4: Copula fitting and sample generation

Bibliography

Andri, S. et mult. al. (2021). *DescTools: Tools for Descriptive Statistics*. R package version 0.99.44. https://cran.r-project.org/package=DescTools.

Carley, H. (2002). Maximum and minimum extensions of finite subcopulas. *Communications in Statistics – Theory and Methods*, 31(12):2151–2166.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

Fanaee-T, H. and Gama, J. (2013). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2(2-3):113–127.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37(2):475–515.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). *copula: Multivariate Dependence with Copulas*. R package version 1.0-1. https://CRAN.R-project.org/package=copula.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC, New York, NY, First edition.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management*. Princeton University Press, Princeton, NJ.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, NY, Second edition.

Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, 98(3):544–567.

Nikoloulopoulos, A. K. (2013). Copula-based models for multivariate discrete response data. In Jaworski, P., Durante, F., and Härdle, W. K., editors, *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer, Berlin, Germany.

Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201–218.

Sklar, A. (1959). *Functions de Repartition an Dimension Set Leursmarges*. l'Institut de Statistique de l'Universite de Paris, Paris, France.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, 8(4):814–827.

Trivedi, P. and Zimmer, D. (2017). A note on identification of bivariate copulas for discrete count data. *Econometrics*, 5(1):10.

Yang, L., Frees, E. W., and Zhang, Z. (2019). Nonparametric estimation of copula regression models with discrete outcomes. *Journal of the American Statistical Association*, 115(530):707–720.