Bachelor's project

NTNU
Norwegian University of Science and Technology
Department of Mathematical Sciences

Nora Røhnebæk Aasen

An Introduction to Copula Theory

Bachelor's project in mathematics
Supervisor: Sigrid Grepstad
Co-supervisor: Thea Bjørnland

May 2021

NTNU
Norwegian University of
Science and Technology

Nora Røhnebæk Aasen

An Introduction to Copula Theory

Bachelor's project in mathematics
Supervisor: Sigrid Grepstad
Co-supervisor: Thea Bjørnland
May 2021

Norwegian University of Science and Technology
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abstract

The goal of this thesis is to give a brief introduction to a class of functions called copulas. A major part of the thesis is devoted to understanding and proving Sklar's theorem. The remainder of the thesis presents other basic concepts and properties that are relevant to the studies of copulas.

Acknowledgments

I would like to thank my supervisors, Thea Bjørnland and Sigrid Grepstad. They have been available throughout the entire process, and their feedback and explanations was always thorough and helpful. In particular I want to thank them for finding such an interesting topic for me to write about, and thus allowing me to cultivate my interests in both analysis and statistics.

Contents

Chapter 1

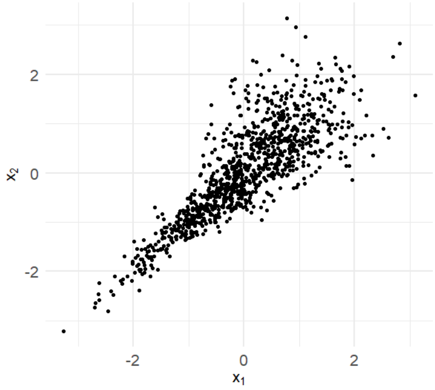
A motivating example

This thesis aims to introduce a class of functions called copulas and the most basic theory concerning these functions. The thesis is by no means a complete presentation of the topic and intends only to serve as an introduction. Before presenting the formal theory behind copulas, we will present an example that demonstrates what purpose these functions might serve in a statistical setting. We will not spend time on definitions in this chapter, as the concepts should be familiar to anyone who has taken an introductory course in statistics. The following example is a reconstruction of the motivating example from chapter 1 in the book *Elements of Copula Modeling with R* [1]. All simulations have been done in R [2], and visualised using the packages "ggplot2" [3] and "Reshape2" [4].

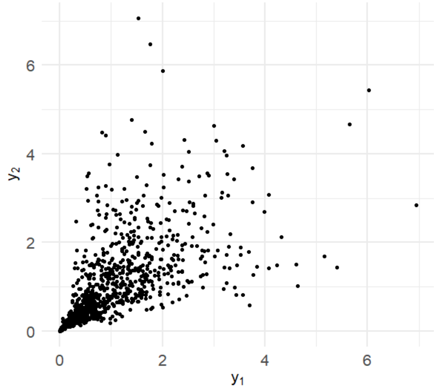
Assume that you are presented two data sets of paired observations (see Figure 1.1), and you are asked to find out if they have anything in common. The data sets consist of 1000 independent realisations of the *bivariate random vectors* (X_1, X_2) and (Y_1, Y_2). The joint distribution functions of these random vectors are unknown.

The first thing you notice is that there is dependence between X_1 and X_2, and between Y_1 and Y_2. You ask yourself: "How does a change in X_1 affect X_2, and is the effect stronger, weaker, or the same for Y_1 and Y_2?". The *linear correlation coefficient*, also known as *Pearson's correlation coefficient*, is a measure of linear dependence between two random variables. From examining the plots, you conclude that there appears to be a positive correlation for both X_1 and X_2, and Y_1 and Y_2. However, you suspect that the variables might not have the same correlation. The calculations of the empirical correlation coefficients confirms this, as $\mathrm{Cor}(X_1, X_2) \approx 0.83$ and $\mathrm{Cor}(Y_1, Y_2) \approx 0.64$.

Next, you want to evaluate the marginal densities of X_1, X_2, Y_1, and Y_2, as the two scatter plots do appear to differ a great deal more than what can be explained from the difference in correlation alone. From a plot of the empirical densities (see Figure 1.2), it is natural to suggest that X_1 and X_2 are both normally distributed,



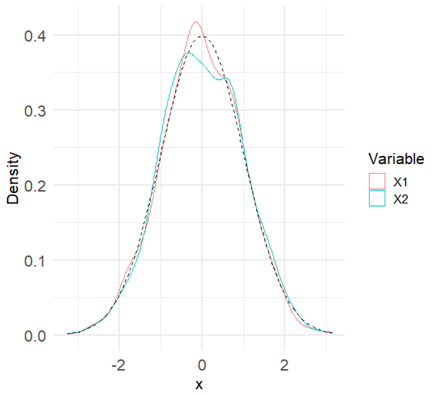
(a) 1000 realisations of (X_1, X_2).



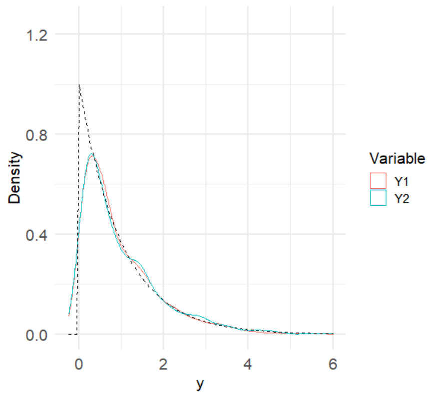
(b) 1000 realisations of (Y_1, Y_2).

Figure 1.1: Scatter plots of the two data sets.

whereas Y_1 and Y_2 are exponentially distributed.



(a) The density of X_1 and X_2, along with a dotted representation of the standard normal distribution.



(b) The density of Y_1 and Y_2, along with a dotted representation of the exponential distribution with $\lambda = 1$.

Figure 1.2: Plot of the estimated marginal density for the data sets.

To conclude, the two data sets appear to come from joint distributions with different marginal densities, and the normally distributed data has a stronger linear dependence than the exponentially distributed data. However, the linear correlation coefficient is only able to capture the strength of linear dependence of the underlying random variables. The different marginal distributions of the data sets might have affected how the dependence is perceived. You decide to transform the data so that they have the same marginal distribution. Then the comparison of dependence would be fairer.

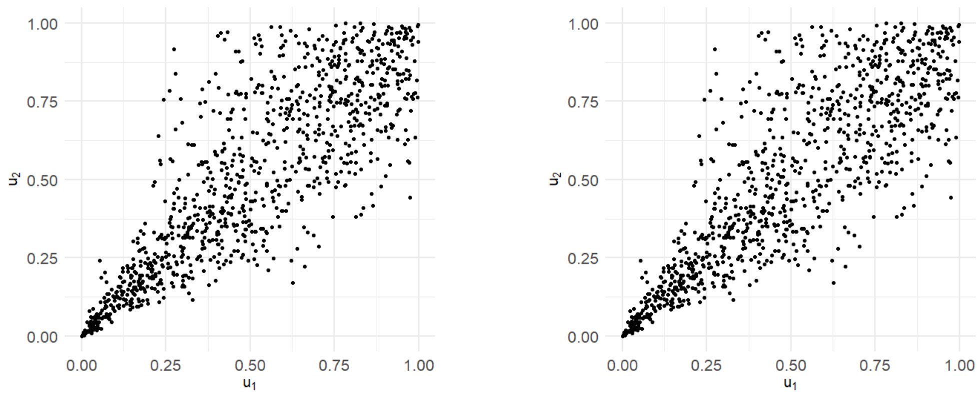
Lemma 1.1. *Let X be a random variable and let F be its continuous distribution function, i.e. $X \sim F$. Then $F(X) \sim U[0,1]$, where $U[0,1]$ is the standard uniform* *distribution on* $[0,1]$.

The proof follows from the observation that, given $Y = F(X)$,

$$P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

when the inverse F^{-1} exists, which it does for both the cumulative normal and exponential distribution function. [1]

Hence, you can calculate $U_1 = F(X_1)$ and $U_2 = F(X_2)$, where F in this case is the standard normal distribution, and similarly $V_1 = G(Y_1)$ and $V_2 = G(Y_2)$, where G is the exponential distribution with $\lambda = 1$. After looking at the plots of the transformed data (see Figure 1.3) you conclude that the two data sets actually have equal dependence, and the only difference was the marginal distributions.



(a) The transformed data $(F(X_1), F(X_2))$.

(b) The transformed data $(G(Y_1), G(Y_2))$.

Figure 1.3: Plot of the two data sets after transforming the marginal distributions.

The subject of this thesis is a class of functions called copulas. These functions represent the dependence between variables in multivariate distributions. In our case, instead of saying that (X_1, X_2) and (Y_1, Y_2) have the same dependence, one could say that they share the same copula. This illustrates that, in contrast to the well-known Pearson correlation coefficient, copulas serve as a more flexible tool for describing the dependence of random variables, separately from their marginal distributions.

A final note to the observant reader. The scatter plot of the two normally distributed variables (Figure 1.1a) does not look like what you would expect from a regular plot of the bivariate normal distribution. That is because the copula for X_1 and X_2 is the *Clayton copula* (see section 4.2). This copula has a stronger dependence in the left tail than the right tail, which explains the discrepancy between this scatter plot and what we would expect from a bivariate normal scatter plot.

[1] A similar argument can be made using a *quasi-inverse* (see definition 2.5) to generalize for all continuous functions.

Chapter 2

Preliminaries and the definition of copulas

In this chapter we introduce notation, cover certain preliminaries regarding distribution functions and present the copula function with some examples of copulas. The concepts presented in this chapter can be retraced in chapters 2.1-2.3 of *An Introduction to Copulas* [5]. The included illustrations were made using the packages "copula" [6] and "lattice" [7] in R .

2.1 Notation

Unless specified otherwise, X is a random variable and $\mathbf{X} = (X_1, \ldots, X_n)$ is a random vector in n dimensions, where each X_j is a random variable.

The unit interval is denoted $\mathbb{I} = [0, 1]$, and $\mathbb{I}^n = \mathbb{I} \times \mathbb{I} \times \cdots \times \mathbb{I}$, i.e. $\mathbb{I}^n$ is the unit cube in n dimensions. By $\times$ we mean the Carthesian product.

We say that a function $f : S_1 \to S_2$ has domain $\mathrm{Dom} f = S_1$ and range $\mathrm{Ran}\, f = S_2$. A function f is said to be nondecreasing if $f(x) \leq f(y)$ for all $x, y \in S_1$ such that $x < y$, and strictly increasing if $f(x) < f(y)$ for all $x, y \in S_1$ such that $x < y$.

2.2 Preliminaries

Before we present what a copula is, it is important to understand what we mean with a *distribution function*, and what properties these functions have. The following definitions will aid with this.

Definition 2.1. Let H be a function defined on $A \subseteq \mathbb{R}^n$. The *H-volume* of a box $B = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subseteq A$ is given by

$$V_H(B) := \sum_{\mathbf{v}} \mathrm{sgn}(\mathbf{v}) H(\mathbf{v}),$$

where $\mathbf{v}$ are the vertices of the box B, and

$$\operatorname{sgn}(\mathbf{v}) = \begin{cases} 1 & \text{if } v_j = a_j \text{ for an even number of indices.} \\ -1 & \text{if } v_j = a_j \text{ for an odd number of indices.} \end{cases}$$

When $n = 2$, the H-volume of a box $B = [a_1, b_1] \times [a_2, b_2]$ is

$$V_H(B) = H(b_1, b_2) - H(a_1, b_2) - H(b_1, a_2) + H(a_1, a_2).$$

Remark. The definition of H-volume might not be intuitive at first, but its use is mainly for functions that appear in probability theory. We include an additional example of H-volume after presenting distribution functions, as this might shed some light on the intuitive understanding of this concept.

Definition 2.2. We say that a function H is *n-increasing* if

$$V_H(B) \geq 0$$

for all boxes B with vertices in $\operatorname{Dom} H$.

For a function of one variable, being n-increasing is equivalent to being non-decreasing. However, for functions of multiple variables these two properties are not equivalent, as the following two examples, which can be found on page 8 in [5], demonstrate.

Example 2.1. Let $H(x, y) = \max(x, y)$ be defined on $\mathbb{I}^2$. Then it is clear that H is non-decreasing in both arguments. However,

$$V_H(\mathbb{I}^2) = 1 - 1 - 1 + 0 = -1$$

which shows that H is not 2-increasing.

Example 2.2. Let $H(x, y) = (2x - 1)(2y - 1)$ be defined on $\mathbb{I}^2$ and let $B = [x_1, x_2] \times [y_1, y_2] \subseteq \mathbb{I}^2$. Then

$$\begin{aligned} V_H(B) &= (2x_2 - 1)(2y_2 - 1) - (2x_1 - 1)(2y_2 - 1) \\ &\quad - (2x_2 - 1)(2y_1 - 1) + (2x_1 - 1)(2y_1 - 1) \\ &= ((2y_2 - 1) - (2y_1 - 1))((2x_2 - 1) - (2x_1 - 1)) \\ &= (2(y_2 - y_1))(2(x_2 - x_1)) \geq 0 \end{aligned}$$

This means that our function is 2-increasing. However, for $y \in [0, \frac{1}{2}]$ we get that H is a decreasing function of x. Similarly, H is a decreasing function of y when $x \in [0, \frac{1}{2}]$.

Lemma 2.1. *Assume that $H : S_1 \times \ldots \times S_n \to \mathbb{R}$ is an n-increasing function. Furthermore, assume that for each $S_j \subseteq \mathbb{R}$ there exists a smallest element a_j, $j = 1, 2, \ldots, n$, such that*

$$H(a_1, x_2, \ldots, x_n) = H(x_1, a_2, \ldots, x_n) = \ldots = H(x_1, x_2, \ldots, a_n) = 0.$$

Then H is non-decreasing in each argument.

Definition 2.3. A function f is *right-continuous* if for every $\varepsilon > 0$, there exists $\delta > 0$ such that when $x_0 < x < x_0 + \delta$ we have that $|f(x) - f(x_0)| < \varepsilon$.

Less formally, we can say that f is *right-continuous at a point* x_0 if it is continuous when x_0 is approached from the right, and f is a *right-continuous function* if this holds for every point in $\operatorname{Dom} f$.

2.3 Distribution functions

Copulas are of interest to us when viewed in applications alongside distribution functions. In this section, we define these functions and state an important property of them.

Definition 2.4. A *distribution function* $F : \mathbb{R}^n \to [0, 1]$ has the following properties:

1. The function
$$x_j \mapsto F(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)$$
is right-continuous for any $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \in \mathbb{R}$ and $j = 1, 2, \ldots, n$.
2. F is n-increasing.
3. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then
$$F(\mathbf{x}) \to 0 \text{ if } x_j \to -\infty$$
for at least one x_j and
$$\lim_{\mathbf{x} \to \infty} F(\mathbf{x}) = 1$$
where by $\mathbf{x} \to \infty$ we mean that $x_j \to \infty$ for $j = 1, 2, \ldots, n$.

Remark. Given a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the distribution function F of $\mathbf{X}$ is defined as the probability

$$F(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n), \tag{2.1}$$

for $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, and we write $\mathbf{X} \sim F$ to indicate that $\mathbf{X}$ has this distribution.

Many people are first introduced to distribution functions in an introductory statistics class, usually in one dimension and often under the term *cumulative distribution function*. Below we see an example of a one-dimensional distribution function.

Example 2.3. A random variable X is said to have a *standard uniform distribution* if, for $x \in [0,1]$,

$$P(X \leq x) = x.$$

We write this as $X \sim \mathrm{U}[0,1]$.

Example 2.4. For the purpose of understanding H-volume, as defined in definition 2.1, we include an example that should be familiar to anyone who has taken an introductory course in statistics. Let X be a random variable with distribution function F. Then, for an interval $B = [a,b]$,

$$V_F(B) = F(b) - F(a) = P(X \leq b) - P(X \leq a) = P(a < X \leq b).$$

Hence, the H-volume, or in this case F-volume, is the probability given by a distribution function on a restricted subset B of the domain. This probability is often visualized as the area under a graph or, for higher dimensions, a volume.

Note that from lemma2.1, it follows that distribution functions are non-decreasing functions. However, they are not necessarily strictly increasing, and they do not necessarily have an inverse. Therefore, it is useful to define a *quasi-inverse*, which does exist for any distribution function.

Definition 2.5. Let $f : [a,b] \to [c,d]$ be a non-decreasing function. Then the *quasi-inverse* $f^{(-1)}$ of f is defined as follows:

1. if $t \in \operatorname{Ran} f$ then $f^{(-1)}(t) = x$ such that $f(x) = t$, that is

$$f(f^{(-1)}(t)) = t.$$

2. if $t \notin \operatorname{Ran} f$ then

$$f^{(-1)}(t) = \inf\{x | f(x) > t\} = \sup\{x | f(x) < t\}$$

Note that the quasi-inverse of f will not necessarily be unique, as there might be multiple choices of x in 1.

Remark. If f is strictly increasing, we have that $f^{(-1)} = f^{-1}$, meaning that the regular inverse and the quasi-inverse of f coincide.

Multivariate distributions are joint distribution functions of two or more random variables. In the motivating example in chapter 1 we looked at bivariate vectors, and the vectors had bivariate distributions. We also considered the univariate distributions of each component of the vectors. The univariate distributions were, in fact, *marginal distributions*, as they were the distribution of the elements of a random vector with a multivariate distribution.

Definition 2.6. The *marginal distribution functions* F_j of a multivariate distribution function $H : \mathbb{R}^n \to [0,1]$ are defined as

$$F_j(x_j) = \lim_{N \to \infty} H(N, \ldots, N, x_j, N, \ldots, N).$$

where $j = 1, 2, \ldots, n$ and $\mathrm{Dom}F_j = \mathbb{R}$ for each j. We will call the functions F_j *marginals* for short.

Remark. When we say marginals we will mean the univariate marginal distributions. However, it is possible to look at k-dimensional marginals, $k < n$, by letting $x_j \to \infty$ for fewer indices j in H.

An important implication of the next theorem is that the continuity of multivariate distribution functions follows from the continuity of its marginals.

Theorem 2.1. *Let H be an n-dimensional distribution function and let $F_1, F_2, \ldots, F_n$ be its marginals. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have*

$$|H(\mathbf{y}) - H(\mathbf{x})| \leq \sum_{j=1}^{n} |F_j(\mathbf{y}_j) - F_j(\mathbf{x}_j)|$$

The n-dimensional proof is somewhat intricate and can be found in chapter 6 of [8]. Below we prove the theorem for two dimensions.

Proof. Assume H is a 2-dimensional distribution function with marginals F_1 and F_2. Let $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, and assume that $x_1 \leq y_1$ and $x_2 \leq y_2$. We first note that for some real value $M \geq y_2$,

$$V_H([x_1, y_1] \times [y_2, M]) = H(y_1, M) - H(x_1, M) - H(y_1, y_2) + H(x_1, y_2) \geq 0.$$

This implies that

$$H(y_1, y_2) - H(x_1, y_2) \leq H(y_1, M) - H(x_1, M),$$

and by letting $M \to \infty$ we get that

$$H(y_1, y_2) - H(x_1, y_2) \leq F_1(y_1) - F_1(x_1).$$

Furthermore, since distribution functions are non-decreasing in each argument and we assumed $x_1 \leq y_1$, we have

$$|H(y_1, y_2) - H(x_1, y_2)| \leq |F_1(y_1) - F_1(x_1)|. \tag{2.2}$$

With a similar argument we can show that $|H(x_1, y_2) - H(x_1, x_2)| \leq |F_2(y_2) - F_2(x_2)|$. Then

$$\begin{aligned} |H(y_1, y_2) - H(x_1, x_2)| &\leq |H(y_1, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, x_2)| \\ &\leq |F_1(y_1) - F_1(x_1)| + |F_2(y_2) - F_2(x_2)|. \end{aligned}$$

We use the triangle inequality first, and secondly what we saw in equation (2.2). An identical argument will work for all other size-orderings of the variables. □

2.4 Copulas

The copula function that we briefly mentioned in the introductory example is itself a distribution function.

Definition 2.7. Let C be a distribution function in $\mathbb{R}^n$ restricted to the unit cube, with standard uniform marginals. Then C is an *n-copula*, or simply a *copula*.

Equivalently, a copula is a function $C : \mathbb{I}^n \to \mathbb{I}$ with the following properties,

1. $C(\mathbf{u}) = 0$ if $u_j = 0$ for at least one j.
2. $C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$.
3. the C-volume of a box $B \subset \mathbb{I}^n$ is non-negative, i.e. $V_C(B) \geq 0$.
4. The marginals of C are standard uniform distribution functions (see example 2.3).

As a direct consequence of theorem 2.1, we have the following corollary concerning the continuity of copulas.

Corollary 2.1. *Let $C : \mathbb{I}^n \to \mathbb{I}$ be a copula. Then C is Lipschitz continuous, and the inequality*

$$|C(\mathbf{u}) - C(\mathbf{v})| \leq \sum_{j=1}^{n} |u_j - v_j|$$

holds for all $\mathbf{u}, \mathbf{v} \in \mathbb{I}^n$.

As we have briefly mentioned earlier, these functions are of interest to us due to their ability to describe the dependence between random variables. We return to this property in chapter 4. Below we present some examples of functions that are copulas.

Example 2.5. Let $\mathbf{u} \in \mathbb{I}^n$. Then the function $\Pi : \mathbb{I}^n \to \mathbb{I}$ given by

$$\Pi(\mathbf{u}) = u_1 u_2 \ldots u_n$$

is an n-dimensional copula called the *product copula*. See Figure 2.1 and Figure 2.4 for visualisations of this function when $n = 2$.

Example 2.6. Let $\mathbf{u} \in \mathbb{I}^n$. Then the function $M : \mathbb{I}^n \to \mathbb{I}$ given by

$$M^n(\mathbf{u}) = \min(u_1, u_2, \ldots, u_n).$$

is an n-dimensional copula called the *M copula* or the *upper Fréchet-Hoeffding bound*. See Figure 2.2 and Figure 2.4 for visualisations of this copula when $n = 2$.

Why the M copula is called the upper Fréchet-Hoeffding bound stems from the following theorem.

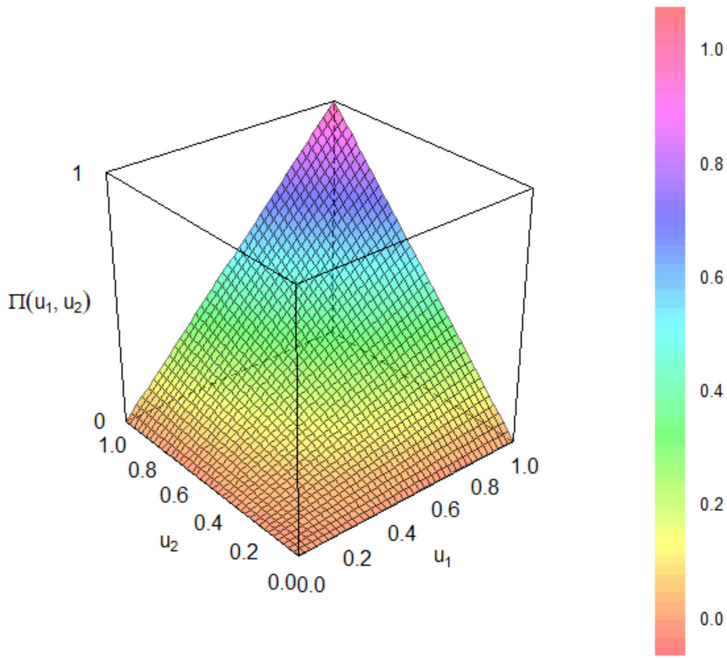


Figure 2.1: The product copula: $\Pi(\mathbf{u}) = u_1 u_2$.

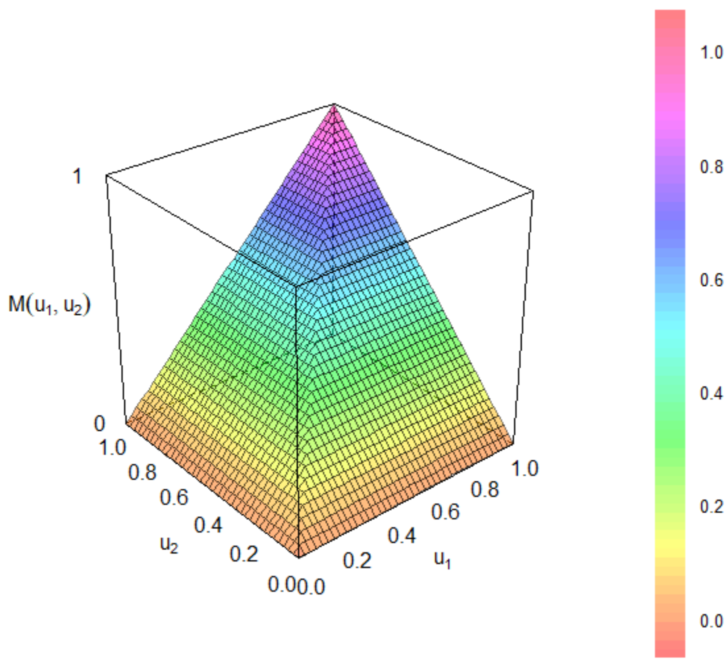


Figure 2.2: The upper Fréchet-Hoeffding bound: $M(\mathbf{u}) = \min(u_1, u_2)$.

Theorem 2.2. *For every copula $C : \mathbb{I}^n \to \mathbb{I}$ and point $\mathbf{u} \in \mathbb{I}^n$ the following inequality holds:*

$$W^n(\mathbf{u}) \leq C(\mathbf{u}) \leq M^n(\mathbf{u}).$$

Here M^n is the M copula from example 2.6 and

$$W^n(\mathbf{u}) = \max(1 + u_1 + \ldots + u_n - n, 0).$$

These bounds are called the *Fréchet-Hoeffding bounds*, hence M^n is called the *upper Fréchet-Hoeffding bound* and W^n is called the *lower Fréchet-Hoeffding bound*.

Proof. We show first that $C(\mathbf{u}) \leq M^n(\mathbf{u})$, and secondly that $W^n(\mathbf{u}) \leq C(\mathbf{u})$.

Start by noting that $C(\mathbf{u}) \leq C(1, \ldots, u_j, \ldots, 1) = u_j$ for all $j = 1, 2, \ldots n$. In particular, $C(\mathbf{u}) \leq \min(u_1, u_2, \ldots, u_n) = M^n(\mathbf{u})$.

For the other inequality, we use the fact that copulas are Lipschitz continuous (see corollary 2.1).

$$\begin{aligned}
|C(1, 1, \ldots, 1) - C(\mathbf{u})| &\leq \sum_{j=1}^{n} |1 - u_j| \\
\implies 1 - C(\mathbf{u}) &\leq \sum_{j=1}^{n} (1 - u_j) \\
\implies 1 - C(\mathbf{u}) &\leq n - \sum_{j=1}^{n} u_j \\
\implies 1 + \sum_{j=1}^{n} u_j - n &\leq C(\mathbf{u}).
\end{aligned}$$

We could safely remove the absolute value on both sides of the initial inequality as $C(\mathbf{u}) \in [0, 1]$ and $u_j \in [0, 1]$ for all j. Finally, since $C(\mathbf{u}) \geq 0$, we conclude that $W^n(\mathbf{u}) \leq C(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{I}^n$. □

It turns out that M^n is a copula for all n, whereas W^n is a copula only when $n = 2$ (see Figure 2.3 and Figure 2.4 for visualizations of the copula W). It is easy to check that W does not meet the requirements of a copula when $n \geq 3$.

Example 2.7. Let $B = \left[\frac{1}{2}, 1\right]^n \subset \mathbb{I}^n$. Then the W^n-volume (definition 2.1) of B is given by

$$V_{W^n}(B) = \sum_{k=0}^{n} \binom{n}{k} \max(1 + \frac{k}{2} + (n - k) - n, 0) = 1 - \frac{n}{2},$$

which is clearly negative for $n \geq 3$.

However, W^n is still the best lower bound that can be found.

Theorem 2.3. *For every point* $\mathbf{u} \in \mathbb{I}^n$*, there exist a copula* $C : \mathbb{I}^n \to \mathbb{I}$*, depending on* $\mathbf{u}$*, such that,*

$$C(\mathbf{u}) = W^n(\mathbf{u}).$$

A proof can be found on page 48 in [5].

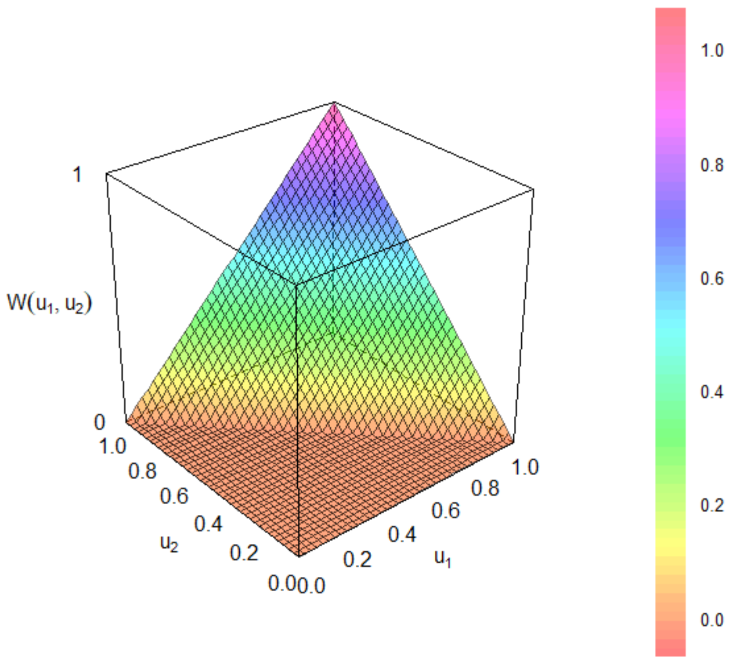


Figure 2.3: The lower Fréchet-Hoeffding bound: $W(\mathbf{u}) = \max(u_1 + u_2 - 1, 0)$.

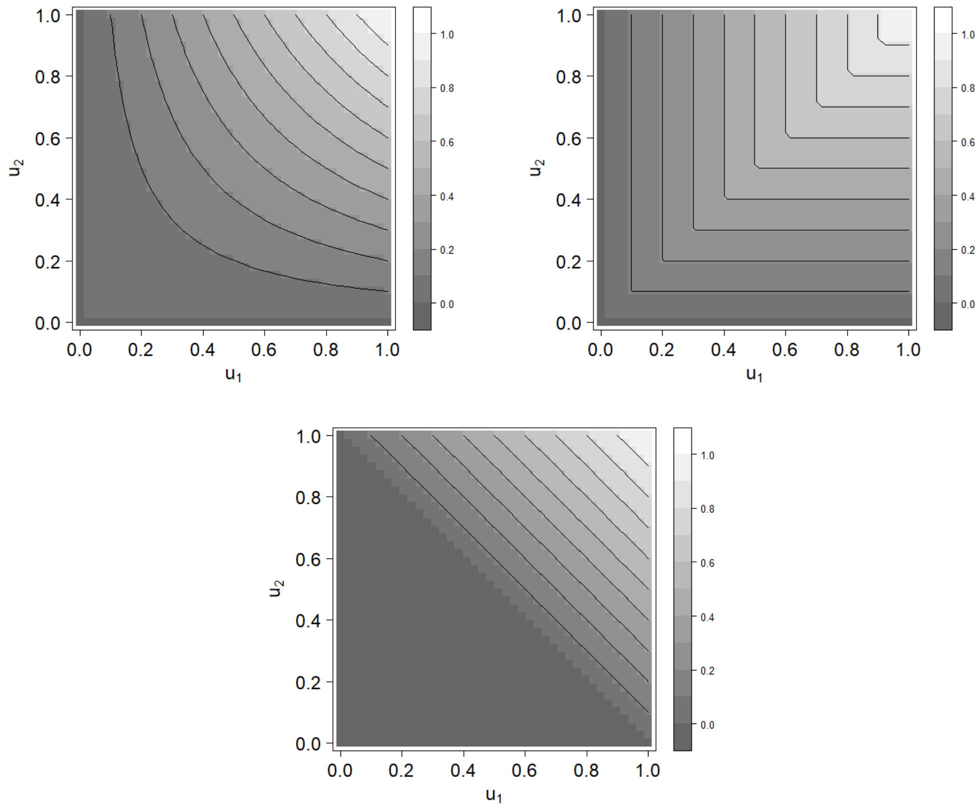


Figure 2.4: Contour plots of the product copula (upper-left), the *M* copula (upper-right), and the *W* copula (bottom).

Chapter 3

Sklar's theorem

The most central theorem within the theory of copulas is Sklar's theorem. This theorem states that every multivariate distribution function can be expressed through its univariate marginals and a copula that describes the dependence between the random variables. It was first presented in an article by Abe Sklar in 1959 [9]. We devote this section to stating and proving this theorem. The proof we present does not follow the original proof by Sklar, but rather the more recent approach of Durante, Fernándes-Sánchez, and Sempi [10]. Although what is presented in this chapter is based on their work, we have done some modifications to the proof.

Theorem 3.1 (Sklar's theorem)**.** *Let H be an n-dimensional distribution function and let $F_1, F_2, \dots, F_n$ be its marginals. Then there exists an n-copula C such that*

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \tag{3.1}$$

for all $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. If all the marginals are continuous, then C is unique. If not, then C can be uniquely determined on $\operatorname{Ran} F_1 \times \operatorname{Ran} F_2 \times \dots \times \operatorname{Ran} F_n$*. Conversely, given a copula C and univariate distribution functions $F_1, F_2, \dots, F_n$, the function H defined by* (3.1) *is an n-dimensional distribution function with marginals $F_1, F_2, \dots, F_n$.*

Remark. The name *copula* actually comes from the Latin word for "link", due to its ability to link together, or "couple", marginal distributions and joint distributions.

The last part of theorem 3.1, which states that given C and $F_1, \dots, F_n$, the function H defined in (3.1) must be a joint distribution function, is a matter of straightforward verification. Moreover, the former part of theorem 3.1 follows immediately from the result below when the marginals of the distribution function H are all continuous.

Corollary 3.1. *Let H be an n-dimensional distribution function and assume that its marginals $F_1, F_2, \dots, F_n$ are continuous. Then the copula C satisfying* (3.1) *is* *determined, for all* $\mathbf{u} \in \mathbb{I}^n$*, by*

$$C(\mathbf{u}) = H(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \ldots, F_n^{(-1)}(u_n)),$$

were $F_j^{(-1)}$ *is the quasi-inverse of* F_j*.*

Verifying theorem 3.1 when H has marginals with discontinuities is far more intricate. The remainder of this chapter is devoted to this task.

3.1 Approximation to the identity

Let $C(\mathbb{I}^n)$ be the set of continuous functions on the unit cube. Then the space $(C(\mathbb{I}^n), \|\cdot\|_\infty)$ is a Banach space or, equivalently, a complete normed space. Here, $\|\cdot\|_\infty$ denotes the supremum norm on $\mathbb{I}^n$. Furthermore, denote by $\mathscr{C}^n$ the set of n-dimensional copulas.

Theorem 3.2. *The set of n-dimensional copulas* $\mathscr{C}^n$ *is a compact subset of the Banach space* $(C(\mathbb{I}^n), \|\cdot\|_\infty)$*.*

Proof of theorem 3.2 will not be given here, but can be found in the article by Durante et al. [10].

We remind the reader of two important properties of compact subsets of a Banach space:

1. A compact subset of a Banach space is bounded and closed, meaning every convergent sequence converges to an element in the space.
2. For every sequence in a compact subset of a Banach space we can find a convergent subsequence.

Now let us assume that H is a multivariate distribution function, where at least one of its marginals is discontinuous. We can find a smooth function closely related to H by taking the convolution of H with an approximation to the identity. Such approximations are sometimes called mollifiers in literature, and they have certain specific properties.

Definition 3.1. A function $\varphi_\varepsilon : \mathbb{R}^n \to \mathbb{R}$ is a *mollifier* if

i) $\int_{\mathbb{R}^n} \varphi_\varepsilon(\mathbf{x}) d\mathbf{x} = 1$,
ii) the support of φ_ε is the closed ball $\overline{B_\varepsilon(\mathbf{0})}$,
iii) φ_ε is infinitely differentiable.

Example 3.1. In this example, we construct a function that fulfills the criteria of a mollifer. Let $B_1(\mathbf{0})$ be an open ball around the origin with radius 1. Then we can define a mollifier $\varphi : \mathbb{R}^n \to \mathbb{R}$ by

$$\varphi(\mathbf{x}) := k \exp\left(\frac{1}{|\mathbf{x}|^2 - 1}\right) \mathbf{1}_{B_1(\mathbf{0})}(\mathbf{x}). \tag{3.2}$$

This corresponds to the case $\varepsilon = 1$ in the definition 3.1. Furthermore, for any $\varepsilon > 0$, we now define

$$\varphi_\varepsilon(\mathbf{x}) := \frac{1}{\varepsilon^n}\varphi\left(\frac{\mathbf{x}}{\varepsilon}\right). \tag{3.3}$$

This allows us to construct a sequence of mollifiers by setting $\varepsilon = 1/m$, such that

$$\lim_{m\to\infty} \varphi_{1/m}(\mathbf{x}) = \delta(\mathbf{x}),$$

where $\delta(\mathbf{x})$ is the Dirac delta function. It follows from the definition above that every element in the sequence $\{\varphi_{1/m}\}_m$ is also a mollifier.

We now define the function H_m by convolution with $\varphi_{1/m}$ as

$$H_m(\mathbf{x}) := (H*\varphi_{1/m})(\mathbf{x}) = \int_{\mathbb{R}^n} H(\mathbf{x}-\mathbf{y})\varphi_{1/m}(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^n} H(\mathbf{y})\varphi_{1/m}(\mathbf{x}-\mathbf{y})d\mathbf{y}. \tag{3.4}$$

The function H_m is a continuous approximation to H, and below we state some important properties for H_m and its marginals $F_{m,j}$. We skip the proofs concerning the marginals of H_m, as this follows from analogous proofs.

Lemma 3.1. *The function H_m, as defined in* (3.4)*, is a continuous, n-dimensional distribution function.*

Proof. We have $H_m = (H * \varphi_{1/m})$, where H is a distribution function. Hence, there exists $M \in \mathbb{N}$ such that $H(\mathbf{x}) > 1-\varepsilon$ when $x_j > M$ for all j, as $H(\mathbf{x})$ tends to 1 when all the elements of $\mathbf{x}$ tend to infinity. Therefore, for $\mathbf{x}$ where $x_j > M + \frac{1}{m}$ for all j,

$$\begin{aligned} H_m(\mathbf{x}) &= \int_{\mathbb{R}^n} H(\mathbf{x}-\mathbf{y})\varphi_{1/m}(\mathbf{y})d\mathbf{y} \\ &= \int_{B_{1/m}(\mathbf{0})} H(\mathbf{x}-\mathbf{y})\varphi_{1/m}(\mathbf{y})d\mathbf{y} \\ &\geq (1-\varepsilon)\int_{\mathbb{R}^n} \varphi_{1/m}(\mathbf{y})d\mathbf{y} = 1-\varepsilon. \end{aligned}$$

Here we use property 1. and 2. of definition 3.1. By a similar argument, we can show that H_m is bounded from below by 0, and thus it satisfies the boundary conditions for a distribution function.

Now, for any box $B \subseteq \mathbb{I}^n$, the H_m-volume of B is

$$
\begin{aligned}
V_{H_m}(B) &= \sum_{\mathbf{v}} \operatorname{sgn}(\mathbf{v}) H_m(\mathbf{v}) \\
&= \sum_{\mathbf{v}} \operatorname{sgn}(\mathbf{v}) \int_{\mathbb{R}^n} H(\mathbf{v}-\mathbf{y}) \varphi_{1/m}(\mathbf{y}) d\mathbf{y} \\
&= \int_{\mathbb{R}^n} \sum_{\mathbf{v}} \operatorname{sgn}(\mathbf{v}) H(\mathbf{v}-\mathbf{y}) \varphi_{1/m}(\mathbf{y}) d\mathbf{y} \\
&= \int_{\mathbb{R}^n} V_H(B^*) \varphi_{1/m}(\mathbf{y}) d\mathbf{y},
\end{aligned}
$$

where B^* is the box B with vertices shifted by the vector $\mathbf{y}$. Clearly, the last integral is non-negative, as $V_H(B^*) \geq 0$ since H is a distribution function.

The last condition we need to check is that H_m is continuous. We already know that $\varphi_{1/m}$ is uniformly continuous on its support $\overline{B_{1/m}(\mathbf{0})} \subset \overline{B_1(\mathbf{0})}$ for $m \geq 1$. Let $\varepsilon > 0$, and choose $\delta > 0$ such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{y})| < \varepsilon$ whenever $|\mathbf{x}-\mathbf{y}| < \delta$. We then get

$$
\begin{aligned}
|H_m(\mathbf{x}) - H_m(\mathbf{y})| &= \left| \int_{\mathbb{R}^n} H(\mathbf{u}) \varphi_{1/m}(\mathbf{x}-\mathbf{u}) - H(\mathbf{u}) \varphi_{1/m}(\mathbf{y}-\mathbf{u}) d\mathbf{u} \right| \\
&\leq \int_{\mathbb{R}^n} |H(\mathbf{u})| |\varphi_{1/m}(\mathbf{x}-\mathbf{u}) - \varphi_{1/m}(\mathbf{y}-\mathbf{u})| d\mathbf{u} \\
&\leq \int_{B_1(\mathbf{x}) \cup B_1(\mathbf{y})} \varepsilon d\mathbf{u} \\
&= 2\varepsilon \lambda_n(B_1(\mathbf{0}))
\end{aligned}
$$

Note that we used the fact that $\sup |H| = 1$ in the second inequality. □

Lemma 3.2. *The marginals $F_{m,1}, F_{m,2}, \ldots, F_{m,n}$ of the function H_m defined in* (3.4) *are continuous, univariate distribution functions.*

Lemma 3.3. *Let $\mathbf{x}$ be a point of continuity for H. Then*

$$
\lim_{m \to \infty} H_m(\mathbf{x}) = H(\mathbf{x}).
$$

Proof. Assume H is continuous at a point $\mathbf{x} \in \mathbb{R}$. Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|H(\mathbf{x}) - H(\mathbf{x}-\mathbf{y})| < \varepsilon$ whenever $\mathbf{y} \in B_\delta(\mathbf{0})$. Assume now that $m > 1/\delta$. Then

$$\begin{aligned}
|H_m(\mathbf{x}) - H(\mathbf{x})| &= \left| \int_{\mathbb{R}^n} H(\mathbf{x}-\mathbf{y})\varphi_{1/m}(\mathbf{y})d\mathbf{y} - H(\mathbf{x}) \right| \\
&= \left| \int_{\mathbb{R}^n} H(\mathbf{x}-\mathbf{y})\varphi_{1/m}(\mathbf{y})d\mathbf{y} - \int_{\mathbb{R}^n} H(\mathbf{x})\varphi_{1/m}(\mathbf{y})d\mathbf{y} \right| \\
&\le \int_{\mathbb{R}^n} |H(\mathbf{x}-\mathbf{y}) - H(\mathbf{x})|\varphi_{1/m}(\mathbf{y})d\mathbf{y} \\
&= \int_{B_\delta(\mathbf{0})} |H(\mathbf{x}-\mathbf{y}) - H(\mathbf{x})|\varphi_{1/m}(\mathbf{y})d\mathbf{y} \\
&\le \varepsilon \int_{\mathbb{R}^n} \varphi_{1/m}(\mathbf{y})d\mathbf{y} = \varepsilon
\end{aligned}$$

which shows that $H_m(\mathbf{x})$ tends to $H(\mathbf{x})$ at points $\mathbf{x}$ where H is continuous. □

Lemma 3.4. *Let $F_1, F_2, \dots, F_n$ be the marginals of the function H, and let F_j be continuous at the point x_j for $j = (1, 2, \dots, n)$. Then*

$$\lim_{m\to\infty} F_{m,j}(x_j) = F_j(x_j),$$

where $F_{m,j}$ is the j'th marginal of H_m.

3.2 Proof of Sklar's theorem

We now have all the tools we need to prove Sklar's theorem when the distribution function H has marginals with discontinuities.

Proof of Sklar's theorem (theorem 3.1). Given a distribution function H, we construct the continuous function $H_m = H * \varphi_{1/m}$. Since H_m is continuous, we know from corollary 3.1 that there exists a copula C_m such that

$$H_m(\mathbf{x}) = C_m(F_{m,1}(x_1), \dots, F_{m,n}(x_n)) \tag{3.5}$$

for all $\mathbf{x} \in \mathbb{R}^n$, where $F_{m,j}$ are the continuous marginals of H_m. The compactness of $\mathscr{C}^n$ guarantees that for every sequence $(C_m)_m$, there exists a convergent subsequence $(C_{m(k)})_k$. In other words, for all $\varepsilon > 0$ there exists $K \in \mathbb{N}$, and a $C \in \mathscr{C}^n$, such that

$$\sup_{\mathbf{u}\in\mathbb{I}^n} |C_{m(k)}(\mathbf{u}) - C(\mathbf{u})| < \varepsilon \tag{3.6}$$

whenever $k \ge K$. Since equation (3.5) holds for all m, it holds in particular for the subsequence $m(k)$,

$$H_{m(k)}(\mathbf{x}) = C_{m(k)}(F_{m(k),1}(x_1), \dots, F_{m(k),n}(x_n)). \tag{3.7}$$

Consider first all continuity points $\mathbf{x}$ of H. At these points, it follows from lemma 3.3 that $H_m(\mathbf{x}) \to H(\mathbf{x})$ when $m \to \infty$. Furthermore, for any subsequences of H_m, we have $H_{m(k)}(\mathbf{x}) \to H(\mathbf{x})$ when $k \to \infty$.

Similarly, we can show that the right side of equation (3.7) converges as well. From lemma 3.4 we know that, for any $\varepsilon > 0$ we can find $k \geq K \in \mathbb{N}$ such that $|F_{m(k),j}(x_j) - F_j(x_j)| < \frac{\varepsilon}{2n}$. Furthermore, using corollary 2.1 and the convergence of copulas seen in (3.6) we have that

$$\begin{aligned}
&|C_{m(k)}(F_{m(k),1}(x_1),\dots,F_{m(k),n}(x_n)) - C(F_1(x_1),\dots,F_n(x_n))| \\
\leq &|C_{m(k)}(F_{m(k),1}(x_1),\dots,F_{m(k),n}(x_n)) - C_{m(k)}(F_1(x_1),\dots,F_n(x_n))| \\
+&|C_{m(k)}(F_1(x_1),\dots,F_n(x_n)) - C(F_1(x_1),\dots,F_n(x_n))| \\
\leq &\sum_{j=1}^{n} |F_{m(k),j}(x_j) - F_j(x_j)| + \frac{\varepsilon}{2} \\
&\leq n\frac{\varepsilon}{2n} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

We can thus conclude that the right hand side of equation (3.5) converges to $C(F_1(x_1),\dots,F_n(x_n))$. Hence, we have shown that, for all points of continuity $\mathbf{x} \in \mathbb{R}$, we have

$$H(\mathbf{x}) = C(F_1(x_1),\dots,F_n(x_n)). \tag{3.8}$$

Assume now that $\mathbf{x}$ is a point of discontinuity for H, which means that at least one of the marginals F_j is discontinuous at x_j. Then we can make a sequence of continuity points $(\mathbf{x}^i)_{i\in\mathbb{N}}$ with $x_j^i > x_j$ and such that $\lim_{i\to\infty} x_j^i = x_j$ for all $j = 1,\dots,n$. Since marginals are right-continuous by definition, such a sequence exists, and for any $\varepsilon > 0$ we can find some $i_0(\varepsilon) \in \mathbb{N}$ such that when $i \geq i_0$ we have

$$|F_j(x_j^i) - F_j(x_j)| < \frac{\varepsilon}{n}$$

for all $j = 1,\dots,n$. Furthermore, we have already shown that equation (3.8) holds for all continuity points of H, hence, for all i,

$$H(\mathbf{x}^i) = C(F_1(x_1^i),\dots,F_n(x_n^i)) \tag{3.9}$$

holds. That $H(\mathbf{x}^i) \to H(\mathbf{x})$ as $i \to \infty$ follows from the fact that the marginals are right-continuous. Finally, for the right hand side in (3.9) we can see that, when $i \geq i_0(\varepsilon)$, we have

$$\begin{aligned}
&|C(F_1(x_1^i),\dots,F_n(x_n^i)) - C(F_1(x_1),\dots,F_n(x_n))| \\
\leq &\sum_{j=1}^{n} |F_j(x_j^i) - F_j(x_j)| < n\frac{\varepsilon}{n} = \varepsilon,
\end{aligned}$$

where we again apply corollary 2.1 for the inequality. Hence, the right-hand side of equation (3.9) converges to $C(F_1(x_1),\dots,F_n(x_n))$. This confirms that equation (3.8) holds for all points $\mathbf{x} \in \mathbb{R}^n$, and concludes the proof of theorem 3.1. □

Chapter 4

Further properties of the copula

In chapter 1, we presented a motivating example where the dependence appeared to be different for two random vectors. However, when the two data sets were transformed to have the same marginal distributions (e.g. standard uniform distributions), the dependence was clearly the same. In chapter 2, we defined copulas as distribution functions with standard uniform marginals. We also showed that copulas are bounded by the upper and lower Fréchet-Hoeffding bounds. In chapter 3, we stated and proved Sklar's theorem. Sklar's theorem is particularly important because it shows how the copula "couples" marginal distribution functions with a multivariate distribution function. A multivariate distribution function can be expressed in terms of its univariate marginal distributions and their dependence structure, which is given by the copula.

In this chapter, which is the final chapter of this thesis, we will further explain some basic properties of copulas, and see these properties applied with the help of two examples: the Clayton family and the Marshall-Olkin family of copulas. These two examples can be retraced on pages 114-115 and 52-54 in [5], respectively. We limit ourselves to two dimensions in this chapter. However, most concepts can be generalised to higher dimensions (see in particular [11] or chapter 2.10 in [5] for details).

4.1 Measuring dependence between random variables

Copulas are primarily of interest when they act as the link between random variables and their joint distribution function, and this is what we turn our attention to in this section. More precisely, we will answer two questions that naturally arise when discussing dependence of random variables: "Which copula corresponds to the situation where the random variables are independent?" and "If we know the copula of two random variables, can we then say something about the copula of transformations of the random variables?".

Independence between random variables should be a familiar concept, and so should the joint distribution function H of two independent variables $X \sim F$ and $Y \sim G$, which is defined as $H(x, y) = F(x)G(y)$. From Sklar's theorem 3.1, the next theorem follows immediately.

Theorem 4.1. *Let X and Y be random variables. Then X and Y are independent if and only if the copula that describes their dependence is the product copula (see example 2.5), given by*

$$\Pi(u, v) = uv. \tag{4.1}$$

Transformations of random variables is a well-known concept in statistics. It turns out that copulas are well-behaved under monotone transformations of the random variables, as we will see below. We first include a small lemma that we state without proof.

Lemma 4.1. *Let f be a strictly monotonic function. Then*

i) the inverse f^{-1} of f exists on $\operatorname{Ran} f$.
ii) if f is strictly increasing, then f^{-1} is also strictly increasing.
iii) if f is strictly decreasing, then f^{-1} is also strictly decreasing.

Theorem 4.2. *Let X and Y be continuous random variables, and denote by C_{XY} the copula of X and Y. If α and β are strictly increasing functions on* $\operatorname{Ran} X$ *and* $\operatorname{Ran} Y$, *respectively, then $C_{\alpha(X)\beta(Y)} = C_{XY}$.*

Proof. Let F_X, F_Y, $F_{\alpha(X)}$ and $F_{\beta(Y)}$ be the distribution functions for X, Y, $\alpha(X)$ and $\beta(Y)$, respectively. From lemma 4.1 it follows that

$$F_{\alpha(X)}(x) = P(\alpha(X) \le x) = P(X \le \alpha^{-1}(x)) = F_X(\alpha^{-1}(x)).$$

Then, again using lemma 4.1,

$$\begin{aligned} C_{\alpha(X)\beta(Y)}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)) &= P(\alpha(X) \le x, \beta(Y) \le y) \\ &= P(X \le \alpha^{-1}(x), Y \le \beta^{-1}(y)) \\ &= C_{XY}(F_X(\alpha^{-1}(x)), F_Y(\beta^{-1}(y))) \\ &= C_{XY}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)) \end{aligned}$$

and we conclude that $C_{\alpha(X)\beta(Y)} = C_{XY}$. □

Note that Pearson's correlation coefficient, which we mentioned in chapter 1, is invariant under linear transformations, but not under all strictly increasing transformations. On the other hand, some linear transformations can alter the copula.

Theorem 4.3. *Let X and Y be continuous random variables, and let C_{XY} be the copula of X and Y. If α and β are strictly monotone functions on* $\operatorname{Ran} X$ *and* $\operatorname{Ran} Y$, *respectively, we have the following:*

1. *if α is strictly increasing and β is strictly decreasing, then*
$$C_{\alpha(X)\beta(Y)}(u,v) = u - C_{XY}(u, 1-v).$$
2. *if α is strictly decreasing and β is strictly increasing, then*
$$C_{\alpha(X)\beta(Y)}(u,v) = v - C_{XY}(1-u, v).$$
3. *if α and β are both strictly decreasing, then*
$$C_{\alpha(X)\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v).$$

We will not prove this, as the proof has a very similar structure to the proof of theorem 4.2.

Since copulas are functions it is not necessarily trivial to determine how they relate to each other. However, since they are measures of dependence it is useful to have a way of comparing them, and the next definition provides a method for doing so.

Definition 4.1. Let C_1 and C_2 be copulas. We say that C_1 is *smaller than* C_2, writing $C_1 \prec C_2$, if $C_1(u,v) \leq C_2(u,v)$ for all u, v in $\mathbb{I}$. Similarly, we say that C_1 is *larger than* C_2, writing $C_1 \succ C_2$, if $C_1(u,v) \geq C_2(u,v)$ for all u, v in $\mathbb{I}$. This ordering is called a *concordance ordering*.

The concordance ordering is only a partial ordering, since not all copulas are comparable in this sense. Note also that the Fréchet-Hoeffding lower bound, W, is smaller than every other copula, while the Fréchet-Hoeffding upper bound, M, is larger than every other copula.

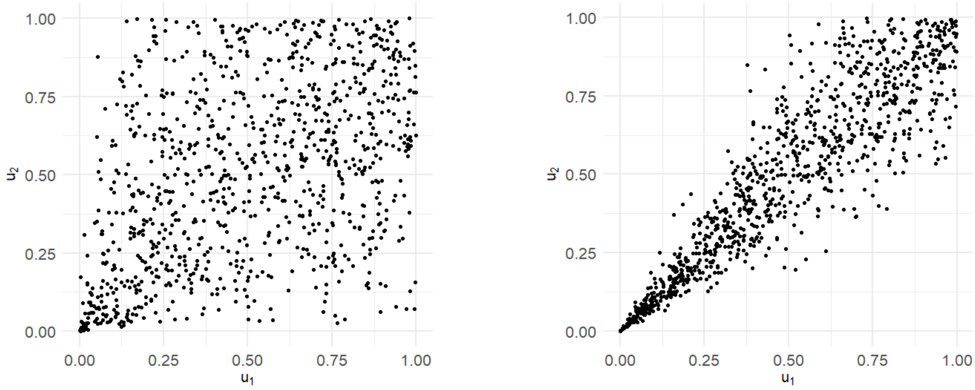
4.2 Example: the Clayton family

For this section, we will look at a family of copulas called the *Clayton family*, and present a setting where this copula naturally appears. We have in fact seen this copula already. In the introductory example, we presented two data sets with equal, nonlinear dependence. In order to construct this data set, we drew 1000 samples from a joint distribution function given by a copula belonging to the *Clayton family*. This is a one-parameter family of copulas with the general form

$$C_\theta = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-1/\theta}, \quad \theta \in [-1, \infty) \backslash \{0\}. \tag{4.2}$$

A visualization of this copula for two different choices of the parameter θ is shown in Figure 4.1. These plots were made by taking 1000 independent realizations of a bivariate vector (U_1, U_2) with joint distribution given by the Clayton copula in equation (4.2).

In Figure 4.1 it appears that the Clayton copula is bounded by $\Pi \prec C_\theta \prec M$ when $\theta \in [0, \infty)$, which is a correct observation. Furthermore, for one-parameter families of copulas in general, denoted $\{C_\theta\}$, we have that their concordance ordering is $C_{\theta_1} \prec C_{\theta_2}$ if $\theta_1 \leq \theta_2$.



(a) parameter $\theta = 1$

(b) parameter $\theta = 6$

Figure 4.1: 1000 samples drawn from the distribution given by the Clayton copula in (4.2).

Remark. An important feature of copulas is their ability to measure *tail dependence*. This is useful for extreme value theory and modeling. The Clayton copula is lower tail dependent, which can be seen in Figure 4.1.

Let us now present a statistical problem, and a method for finding a suitable copula to this problem. Assume $X_1, X_2, \ldots, X_n$ is a random sample of continuous independent random variables, and assume $X_j \sim F$ for all $j = 1, 2, \ldots, n$. Furthermore, let $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ and $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$. We want to find the copula $C_{1,n}$ that describes how $X_{(1)}$ and $X_{(n)}$ depend on each other.

It is well known that

$$\begin{aligned} F_1(x) &= P(X_{(1)} \leq x) = 1 - [1 - F(x)]^n, \text{ and} \\ F_n(x) &= P(X_{(n)} \leq x) = [F(x)]^n. \end{aligned}$$

We start by finding a joint distribution function $\tilde{H}$ of $-X_{(1)}$ and $X_{(n)}$:

$$\begin{aligned} \tilde{H}(s,t) &= P(-X_{(1)} \leq s, X_{(n)} \leq t) \\ &= P(-s \leq X_{(1)}, X_{(n)} \leq t) \\ &= P(X_i \in [-s,t] \text{ for all } i) \\ &= \begin{cases} [F(t) - F(-s)]^n, & -s \leq t \\ 0, & -s > t \end{cases} \\ &= [\max(F(t) - F(-s), 0)]^n. \end{aligned}$$

Let $G(x) = [1 - F(-x)]^n$ denote the distribution function of $-X_{(1)}$. We use Sklar's theorem 3.1 to find $\tilde{C}(u,v) = \tilde{H}(G^{(-1)}(u), F_n^{(-1)}(v))$. Observe that

$$u = [1 - F(-s)]^n \implies F(-s) = 1 - u^{1/n}, \text{ and}$$
$$v = [F(t)]^n \implies F(t) = v^{1/n},$$

and hence

$$\tilde{C}(u,v) = [\max(v^{1/n} + u^{1/n} - 1, 0)]^n.$$

The copula $\tilde{C}$ belongs to the *Clayton family*, with parameter $\theta = -1/n$. Since $\tilde{C}$ is the copula for $-X_{(1)}$ and $X_{(n)}$, we use the transformation $X_{(1)} = -(-X_{(1)})$ to find the copula $C_{1,n}$ that we are interested in. Theorem 4.3, which we covered in the previous section, allows us to find

$$C_{1,n}(u,v) = v - \tilde{C}(1-u,v) = v - \left[\max((1-u)^{1/n} + v^{1/n} - 1, 0)\right]^n.$$

Intuitively we understand that $X_{(1)}$ and $X_{(n)}$ are not independent, and since $C_{1,n} \neq \Pi$ they are indeed dependent on each other in some way. However, they are asymptotically independent. Note that the Clayton copula $\tilde{C}_\theta$ is Π when $\theta = 0$, and when the sample size grows towards infinity, we get

$$\lim_{n\to\infty} C_{1,n}(u,v) = \lim_{n\to\infty} v - \tilde{C}_{-1/n}(1-u,v) = v - \Pi(1-u,v) = v - v + uv = uv$$

which means that the copula $C_{1,n} \to \Pi$ as $n \to \infty$.

The copula $C_{1,n}$ says something about how the smallest and largest element of a set depend on each other. When the set is small, $X_{(1)}$ and $X_{(n)}$ are effectively bounding each other. Say, for instance, that we draw five random samples from a standard uniform distribution. It is not unlikely that $X_{(n)}$ lies in the lower half of the interval. However, if it turns out that $X_{(1)} = 0.4$, then $X_{(n)}$ is suddenly bounded below by 0.4. Assume now that we draw 15 random samples from a standard uniform distribution, and that $X_{(1)} = 0.2$. This does not really tell us much about $X_{(15)}$, because the probability that $X_{(15)}$ even lies in the lower half of the interval is as small as $P(X_{(15)} < 0.5) = 0.5^{15} = 0.00003052$. In other words, for larger sample sizes, knowing $X_{(1)}$ will tell us close to nothing about what values $X_{(n)}$ will take.

4.3 Decomposition of the copula

Let us return to the Fréchet-Hoeffding bounds, presented in section 2.4, and in particular the two plots in Figure 4.2 below. They show 100 independent realizations of a bivariate vector (U_1, U_2) with standard uniform marginals and joint distribution given by the copulas M and W. A question that arises immediately is "Why do the data points accumulate along the diagonals?". From the plots it appears that $U_1 = U_2$ when their joint distribution function is M and that $U_1 = 1 - U_2$ when their joint distribution function is W. This is, in fact, correct.

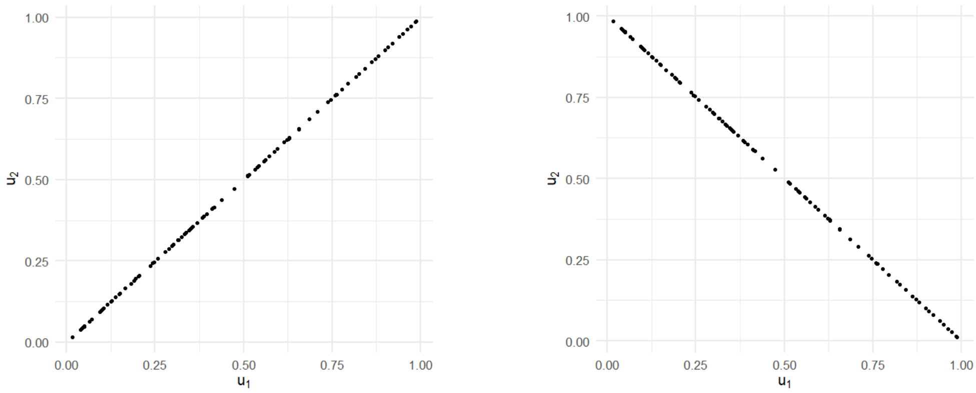


Figure 4.2: 100 samples drawn from the distribution given by the copulas M (left) and W (right).

Theorem 4.4. *Assume that $\mathbf{U} = (U_1, U_2)$ is a bivariate vector with $U_1, U_2 \sim U[0,1]$. Then the joint distribution of $\mathbf{U}$ is given by the copula M if and only if $U_1 = U_2$.*

Theorem 4.5. *Assume that $\mathbf{U} = (U_1, U_2)$ is a bivariate vector with $U_1, U_2 \sim U[0,1]$. Then the joint distribution of $\mathbf{U}$ is given by the copula W if and only if $U_1 = 1 - U_2$.*

In other words, the copulas M and W imply something important about the dependence between U_1 and U_2, namely that they determine each other with absolute certainty. This relationship is what we call *perfect dependence*, and the copulas M and W describe *perfect positive* and *perfect negative dependence*, respectively. An analogous concept is when the Pearson correlation coefficient is 1 or -1, respectively, for a set of multivariate normally distributed variables.

Remark. Keep in mind that W is not a copula for $n > 2$, as we saw in example 2.7, and furthermore, note that the idea of perfect negative dependence does not make sense for dimensions greater than two. Hence, theorem 4.5 can not be generalized to the case $n > 2$.

Perfect dependence, and how this relates to the copulas M and W, is not restricted to random variables with standard uniform distributions, nor linear relationships. Let X and Y be two random variables. If the copula of X and Y is M, then $Y = \alpha(X)$, where α is an increasing function. Similarly, if the copula of X and Y is W, then $Y = \beta(X)$, where β is a decreasing function. The general idea is that if the copula of two random variables is either M or W, the random variables are completely determined by each other.

Let us now justify the claims in theorems 4.4 and 4.5. It is straightforward to verify that, for some $U \sim \mathrm{U}[0,1]$,

$$(U, U) \sim M \quad \text{and} \quad (U, 1-U) \sim W.$$

However, to complete the argument for the converse implications, we need some results and terminology from probability theory. We briefly introduce these concepts formally, before explaining them in a more casual manner below.

Given a probability space, which satisfies certain measure-theoretic properties, it is possible to define a *probability measure*. This is a real-valued function defined on a set of events in a probability space. One important thing to note is that, opposed to general measures, probability measures must assign the value 1 to the entire probability space.

It is well known that any two-dimensional distribution function H induces a probability measure on $\mathbb{R}^2$. Since H-volume upholds the properties of a probability measure, this is done by assigning the measure $H(x,y) = V_H(B)$ to the box $B = (-\infty, x] \times (-\infty, y]$. This measure can be extended to any Borel subset of $\mathbb{R}^2$, and thus create a set of events in $\mathbb{R}^2$ that are measurable. Since copulas are also joint distribution functions, we can in a similar way define a C-measure on $\mathbb{I}^2$ by assigning the measure $C(u,v) = V_C(B)$ to the box $B = [0,u] \times [0,v]$.

Intuitively, the C-measure of a subset $A \subset \mathbb{I}^2$ is the probability that a random vector (U_1, U_2) with standard uniform marginals and joint distribution function C assumes values in that subset, that is $P((U_1,U_2) \in A) = V_C(A)$. With the notion of measure established, we are now ready to introduce *support*, as well as a useful decomposition of copulas.

Definition 4.2. Let $C : \mathbb{I}^2 \to \mathbb{I}$ be a copula. The complement of the union of all open sets in $\mathbb{I}^2$ that have C-measure zero is called the *support* of C, and is denoted $\mathrm{supp}(C)$. If $\mathrm{supp}(C) = \mathbb{I}^2$, we say that C has *full support*.

Again, let (U_1, U_2) be a random vector with standard uniform marginals and joint distribution function C. The support of C is the subset of $\mathbb{I}^2$ in which (U_1, U_2) can take values. Conversely, given a subset $A \subset \mathbb{I}^2$ which satisfies $A \cup \mathrm{supp}(C) = \emptyset$, then $P((U_1, U_2) \in A) = 0$.

Definition 4.3. Let $C : \mathbb{I}^2 \to \mathbb{I}$ be a copula. Then C can be split into an *absolutely continuous component* A_C and a *singular component* S_C, where

$$A_C(u,v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} C(s,t)\, ds dt \tag{4.3}$$

and $S_C(u,v) = C(u,v) - A_C(u,v)$.

If $C \equiv A_C$ on $\mathbb{I}^2$ then C is *absolutely continuous*. If $C \equiv S_C$ then C is *singular*. Note that C is singular if and only if the Lebesgue measure of its support is zero. Additionally, the C-measure of the absolutely continuous component and singular component is given by $A_C(1,1)$ and $S_C(1,1)$, respectively.

Example 4.1. The upper Fréchet-Hoeffding bound is given by $M(u,v) = \min(u,v)$.

Computing the absolutely continuous component of C yields

$$A_M(u,v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} \min(s,t)\, ds dt.$$

It is obvious that $\frac{\partial^2}{\partial s \partial t} \min(s,t)$ is zero everywhere, except on the diagonal $t = s$ where it is undefined, thus $A_M = 0$ and $S_M = M$. We could have come to the same conclusion by observing that $\text{supp}(M)$ is the diagonal $u_1 = u_2$, which has Lebesgue measure zero, and therefore M must be singular.

We can relate this back to the plot in Figure 4.2, and the question concerning why the data points accumulated in such a specific pattern. Example 4.1 concludes that M is singular, with the diagonal $u_1 = u_2$ as its support. This implies that (U_1, U_2) has joint distribution function M only if $U_1 = U_2$. The case is similar for the lower Fréchet-Hoeffding bound.

Example 4.2. The product copula is given by $\Pi(u,v) = uv$. Hence

$$A_\Pi(u,v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} st\, ds dt = \int_0^u \int_0^v 1\, ds dt = uv,$$

which shows that the product copula is absolutely continuous. Furthermore, assume $B = [a_1, b_1] \times [a_2, b_2] \subseteq \mathbb{I}^2$, with $a_1 < b_1$ and $a_2 < b_2$. Then the Π-volume of B is

$$V_\Pi(B) = (b_2 - a_2)(b_1 - a_1) > 0.$$

Hence, the product copula has full support.

The next section covers an example of a copula that has full support, but has both an absolutely continuous and a singular component.

4.4 Example: the Marshall-Olkin family

Consider a two-component system, which at all times is subject to “shocks” to either one of the components, or both simultaneously. These shocks always cause the components to fail, and they appear according to three independent Poisson processes with parameters λ_1, λ_2 and λ_{12} for failure in component 1, 2, or both, respectively. The times until failure, denoted T_1, T_2 and T_{12}, are therefore independent, exponential random variables with parameters λ_1, λ_2 and λ_{12}.

We are interested in finding the distribution of the lifetime of our system. Let $X = \{\text{Lifetime of component 1}\}$, $Y = \{\text{Lifetime of component 2}\}$ and $\bar{H}(x,y) = P(X > x, Y > y)$[1]. Then

$$X = \min(T_1, T_{12}), \quad Y = \min(T_2, T_{12}),$$

[1] $\bar{H}$ is called a survival function.

and

$$\begin{aligned}\bar{H}(x,y) &= P(T_1 > x)P(T_2 > y)P(T_{12} > \max(x,y)) \\ &= \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)).\end{aligned}$$

The marginals of $\bar{H}$ are given by

$$\begin{aligned}\bar{F}(x) &= P(X > x) = \exp(-(\lambda_1 + \lambda_{12})x), \text{ and} \\ \bar{G}(y) &= P(Y > y) = \exp(-(\lambda_2 + \lambda_{12})y).\end{aligned} \tag{4.4}$$

We can now, using the fact that $\max(x,y) = x + y - \min(x,y)$, find

$$\begin{aligned}\bar{H}(x,y) &= \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)) \\ &= \exp(-\lambda_1 x - \lambda_{12}x - \lambda_2 y - \lambda_{12}y - \lambda_{12}\min(x,y)) \\ &= \bar{F}(x)\bar{G}(y)\min(\exp(\lambda_{12}x), \exp(\lambda_{12}y)).\end{aligned}$$

Note that

$$\bar{F}^{(-1)}(u) = \frac{-1}{\lambda_1 + \lambda_{12}}\log(u) \text{ and } \bar{G}^{(-1)}(v) = \frac{-1}{\lambda_2 + \lambda_{12}}\log(v),$$

and define

$$\alpha = \frac{\lambda_{12}}{\lambda_1 + \lambda_{12}} \text{ and } \beta = \frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}. \tag{4.5}$$

We now use Sklar's theorem 3.1 to find

$$C(u,v) = \bar{H}(\bar{F}^{(-1)}(u), \bar{G}^{(-1)}(v)) = uv\min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}). \tag{4.6}$$

Remark. Note that this is not direct use of Sklar's theorem, since $\bar{F}$ and $\bar{G}$ are not distribution functions, but rather what is known as *survival functions*. However, it can be shown that the relationship between univariate and joint survival functions is analogous to that of regular distribution functions, and the function C obtained in equation (4.6) is in fact a copula. For further details see pages 32-33 in [5].

The copula in (4.6) belongs to a two-parameter family known as the *Marshall-Olkin family*, which is given by,

$$C_{\alpha,\beta}(u,v) = \min\left(u^{1-\alpha}v, uv^{1-\beta}\right) = \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta \end{cases} \tag{4.7}$$

for $0 < \alpha, \beta < 1$. If we extend the parameter range to $0 \leq \alpha, \beta \leq 1$, then $C_{\alpha,0} = C_{0,\beta} = \Pi$ and $C_{1,1} = M$.

The copulas in this family have full support, but they are neither absolutely continuous nor singular. We compute the absolutely continuous component by first finding the partial derivatives

$$\frac{\partial^2}{\partial u \partial v} C_{\alpha,\beta}(u,v) = \begin{cases} (1-\alpha)u^{\alpha}, & u^{\alpha} > v^{\beta} \\ (1-\beta)v^{-\beta}, & u^{\alpha} < v^{\beta} \end{cases}$$

and then evaluating the double integral in (4.3) to get

$$A_{\alpha,\beta}(u,v) = C_{\alpha\beta}(u,v) - \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \left[\min\left(u^{\alpha}, v^{\beta}\right)\right]^{(\alpha+\beta-\alpha\beta)/\alpha\beta}.$$

Hence, the singular component is given by

$$S_{\alpha,\beta}(u,v) = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \left[\min\left(u^{\alpha}, v^{\beta}\right)\right]^{\frac{\alpha+\beta-\alpha\beta}{\alpha\beta}} = \int_0^{\min\left(u^{\alpha}, v^{\beta}\right)} t^{\frac{1}{\alpha}+\frac{1}{\beta}-2} dt$$

and is supported on the line $u^{\alpha} = v^{\beta}$. The $C_{\alpha,\beta}$-measure of the singular component is given by

$$S_{\alpha,\beta}(1,1) = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta}.$$

If we return to the initial setup, the singular component corresponds to the case when a shock kills both components simultaneously, which is when $X = Y$. To see this, recall that

$$\bar{H}(x,y) = C(\bar{F}(x), \bar{G}(y)) = A_C(\bar{F}(x), \bar{G}(y)) + S_C(\bar{F}(x), \bar{G}(y)).$$

The singular component, S_C, has its support on the line $v = u^{\alpha/\beta}$. Recalling what we defined in equation (4.5), we get that $v = u^{\alpha/\beta}$ evaluated at $u = \bar{F}(x)$ and $v = \bar{G}(y)$ corresponds to

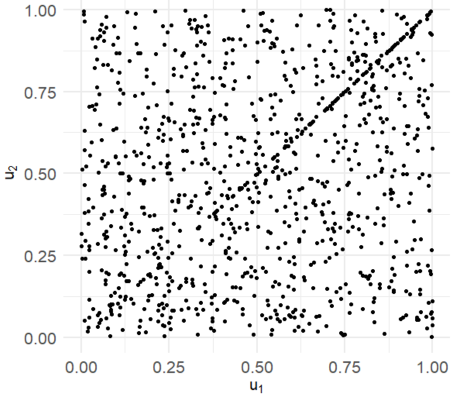
$$\bar{G}(y) = (\bar{F}(x))^{\alpha/\beta} = (\bar{F}(x))^{\frac{\lambda_2+\lambda_{12}}{\lambda_1+\lambda_{12}}}.$$

Furthermore, using (4.4), we get

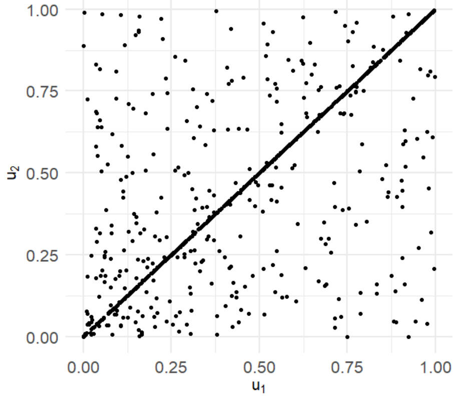
$$\begin{aligned} \bar{G}(y) &= (\bar{F}(x))^{\frac{\lambda_2+\lambda_{12}}{\lambda_1+\lambda_{12}}} \\ \exp\left(-(\lambda_2+\lambda_{12})\,y\right) &= \exp\left(-(\lambda_1+\lambda_{12})\,x\right)^{\frac{\lambda_2+\lambda_{12}}{\lambda_1+\lambda_{12}}} \\ -(\lambda_2+\lambda_{12})\,y &= \left(\frac{\lambda_2+\lambda_{12}}{\lambda_1+\lambda_{12}}\right)(-(\lambda_1+\lambda_{12})\,x) \\ y &= x \end{aligned}$$

To conclude, we have

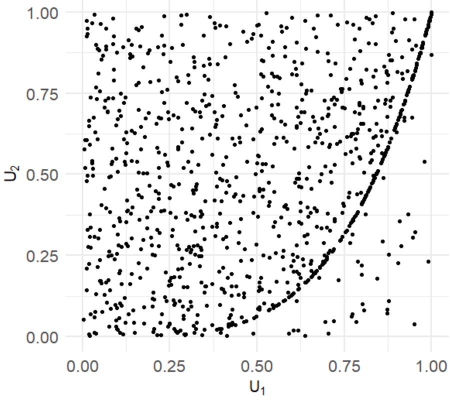
$$P(T_{12} < \min(T_1, T_2)) = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} = \frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}},$$



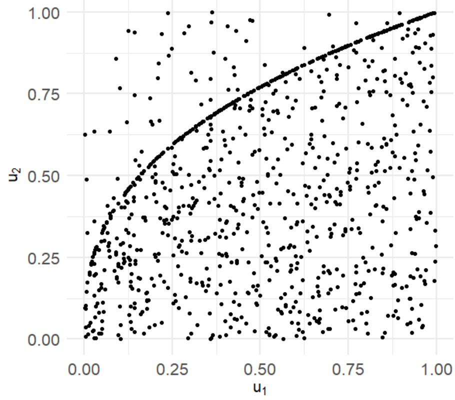
(a) $(\alpha, \beta) = (0.2, 0.2)$



(b) $(\alpha, \beta) = (0.8, 0.8)$



(c) $(\alpha, \beta) = (0.8, 0.2)$



(d) $(\alpha, \beta) = (0.3, 0.8)$

Figure 4.3: 1000 samples drawn from the distribution given by the Marshall-Olkin copula defined in equation (4.7).

which is coherent with what we would get from manually computing this probability based on the three independent exponential distributions.

Each plot in Figure 4.3 is made from 1000 simulations of the bivariate vector (U_1, U_2) with joint distribution given by the Marshall-Olkin copula from equation (4.7), with parameter values for α and β indicated in the caption below each plot. In terms of our example, we can think of $U_1 = \bar{F}(X)$ and $U_2 = \bar{G}(Y)$[2]. We could have obtained similar plots by simulating the failure times t_1, t_2, t_{12} of three independent exponential distributions, calculated the point (x, y) for each simulation, and transformed the data point as $u_1 = F(x)$, $u_2 = G(y)$.

The singular component is clearly visible in all four plots. Recall that this corres-

[2]This follows from the same logic that allowed us to transform the variables in the motivating example in chapter 1. However, keep in mind that here we are not actually transforming anything, but only arguing for the link that exists between the variables in our example and the data in our plots.

ponds to the case $T_{12} < \min(T_1, T_2)$. By comparison of Figure 4.3a and 4.3b we conclude that as α and β get close to 1, a shock that destroys both components is more likely to occur before shocks that only strike one of them.

Furthermore, from 4.3c and 4.3d we can deduce that the plots seem to be separated into distinct areas. In fact, the area above the singular component corresponds to the situation $T_2 < \min(T_1, T_{12})$ and the area below the singular component corresponds to the situation $T_1 < \min(T_2, T_{12})$.

Hence, in Figure 4.3c the low value of the β-parameter along with the high value of the α-parameter result in many simulations where component two breaks first, and quite few situations were component one breaks first. This is reasonable, since $\beta < 0.5 < \alpha$ implies that $\lambda_2 < \lambda_{12} < \lambda_2$, based on how we defined α and β in equation (4.5). In particular, for β close to one and α close to zero, λ_2 is much greater than λ_1, and hence T_2 is on average significantly smaller than T_1. For Figure 4.3d the situation is opposite.

Bibliography

[1] M. Hofert, I. Kojadinovic, M. Mächler and J. Yan, *Elements of Copula Modeling with R*, ser. Use R! Springer, Cham, 2018, ISBN: 978-3-319-89635-9.

[2] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: `https://www.R-project.org/`.

[3] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN: 978-3-319-24277-4. [Online]. Available: `https://ggplot2.tidyverse.org`.

[4] H. Wickham, 'Reshaping data with the reshape package,' *Journal of Statistical Software*, vol. 21, no. 12, pp. 1–20, 2007. [Online]. Available: `http://www.jstatsoft.org/v21/i12/`.

[5] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed. Springer-Verlag New York, 2006, ISBN: 978-0-387-28678-5.

[6] M. Hofert, I. Kojadinovic, M. Maechler and J. Yan, *Copula: Multivariate dependence with copulas*, R package version 1.0-1, 2020. [Online]. Available: `https://CRAN.R-project.org/package=copula`.

[7] D. Sarkar, *Lattice: Multivariate Data Visualization with R*. New York: Springer, 2008, ISBN 978-0-387-75968-5. [Online]. Available: `http://lmdvr.r-forge.r-project.org`.

[8] B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*, ser. North Holland series in probability and applied mathematics. North Holland, 1983, ISBN: 9780444006660.

[9] A. Sklar, 'Fonctions de répartition à n dimensions et leurs marges,' *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959.

[10] F. Durante, J. Fernández-Sánchez and C. Sempi, 'Sklar's theorem obtained via regularization techniques,' *Nonlinear Anal*, vol. 75, no. 2, pp. 769–774, 2012.

[11] P. Embrechts, F. Lindskog and A. McNeil, 'Chapter 8 – modelling dependence with copulas and applications to risk management,' Available: `http://homepage.math.uiowa.edu/~lwang/finance/DependenceWithCopulas.pdf`, 2003.

NTNU
Norwegian University of
Science and Technology