Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Sigurd Stenvik

Copula

Master's thesis in MSMNFMA
Supervisor: Øyvind Bakke
May 2020

NTNU
Norwegian University of Science and Technology

Sigurd Stenvik

Copula

Master's thesis in MSMNFMA
Supervisor: Øyvind Bakke
May 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Copula

Sigurd Stenvik

May 2020

Abstract

The copula is a very interesting tool in statistics. It's used in many setting from quantitative finance to climate models. Copulas are to a large extent useful because they can elegantly separate the dependence structure from marginal distributions in a multivariate distribution. We will explain this in detail in the thesis. If you for the first time look at the definition of a copula it might be difficult to understand what a copula actually is. Therefore, we also give a description of the copula function in terms of distribution function, which is intuitive for someone with some prior knowledge of statistics. We also write about Sklar's Theorem, which theoretically explains the connection between the bivariate distribution, its marginal distributions and the copula. Next we generalise the theory of the copula from 2 to n dimensions, and we also show how to estimate the parameters of a copula. Finally, we show how you can use the copula to simulate samples from a bivariate distribution.

Sammendrag

Copulaen er et interessant verktøy i statistikken. Den er brukt i mange forskjellige områder fra finans til klimamodeller. En stor grunn til at copulaen er nyttig er hvordan man kan bruke copulaen til å splitte en bivariat fordeling opp i avhengighetsstrukturen og selve marginalfordelingene. Vi vil forklare dette i denne oppgaven. Hvis du for første gang ser på definisjonen til copulaen kan det være vanskelig å forstå hva copulaen egentlig er. Derfor har vi gitt en foklaring på hva en copula er i form av sansynlighetsfordelinger, som burde være intuitiv for en person som allerde har litt kunnskap om statistikk. Vi skriver også om sklars teorem, som teoretisk forklarer denne sammenhengen mellom den bivariate fordelingen, dens marginalfordelinger og avhengighetsstrukturen mellom marginalfordelingene. Vi generaliserer også denne teorien fra 2 til n dimensjoner, og vi forklarer hvordan man kan estimere parameterne til en copula. Vi avslutter oppgaven med å vise hvordan man kan bruke en copula til å simulere fra en bivariat fordeling.

Preface

This thesis is part of my 2 year Master of Science degree in Mathematical Sciences with specialization in statistics at NTNU. My supervisor has been professor Øyvind Bakke, and the subject of study was the copula. In the process of writing the thesis I've got a lot of help, and I would like to especially thank Øyvind for all the help and guidance through writing this thesis.

Contents

Chapter 1

Introduction

If we know the marginal distributions of a multivariate distribution the reader might already know that this is not enough to fully describe the multivariate distribution. What is lacking to fully describe the multivariate distribution, is the dependence between the marginal distributions. It is here the copula comes into play as the copula is a tool to show the dependence structure between marginal distributions in a multivariate distribution.

On the applied side the copula is a very useful tool in finance where modeling of joint distribution is needed. For example, if you want to make a multivariate distribution function of different asset return Roncalli [1] says that you can use the copula to split up the problem into two parts. Part one is modeling the marginal distribution of the individual assets returns. Part two is figuring out a copula that describe the dependence structure between the different assets returns.

Copulas were not used a lot in finance before year 2000, but this changed after Li [4] published his article in 1999. His paper led to use of the Gaussian copula "to price and manage the risk of Collatarised Debt Obligations" [5, p. 1]. After the finance crisis in 2007–2008 the Gaussian copula was target of some criticism. One of these critics was Felix [6], who called the Gaussian copula "The Formula that Killed Wall Street". However, Watts [5] states that changing the Gaussian copula with another copula would not have changed the outcome of the crisis. Watts also says that the crisis was more a product of poor estimation of the correlation between assets rather than the choice of the Gaussian copula.

Chapter 2

Definition and basic properties

To define what a copula is we first need some preliminaries. We first define what a 2-increasing function is. Let $\mathbb{R}$ be the real line $(-\infty, \infty)$ and let $\overline{\mathbb{R}}$ be the real line included $\pm\infty$, that is $[-\infty, \infty]$. We then define the extended real plane as the Cartesian product $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$. We also use the notation $\mathbb{I}$ for the subset $[0, 1]$ of the of the real line $\mathbb{R}$. Sometimes we will talk about a box B which could be of some dimension n. If $n = 2$ we describe B as a Cartesian product $[x_1, x_2] \times [y_1, y_2]$ where $(x_1, y_1), (x_2, y_1), (x_1, y_2), (x_2, y_2)$ are called the vertices of the box B.

We will also introduce the notation $V_C(B)$, which is called the C-volume of the box B. It is defined as $C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1)$ where $(x_1, y_1), (x_2, y_1)$, $(x_1, y_2), (x_2, y_2)$ are the vertices of the box B.

Definition 1. *A copula $C(u, v)$ is a function that maps values from $\mathbb{I} \times \mathbb{I}$ to $\mathbb{I}$ and satisfies the following three properties:*
a) The copula is grounded, meaning $C(u, 0) = C(0, v) = 0$ for all u and v.
b) $C(u, 1) = u$ and $C(1, v) = v$ for all u and v.
c) The copula is a 2-increasing function, that is $C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \geq 0$ for all possible x_1, x_2, y_1, y_2 where $y_1 \leq y_2$ and $x_1 \leq x_2$.

Some of these properties might feel a bit arbitrary, especially that a copula has to be a 2-increasing function, and you might wonder if it is possible to make the definition of a copula simpler.

It can be shown that a 2-increasing function is non-decreasing in each argument if it is grounded. To prove this we start with a lemma from [2, p. 9].

Lemma 1. *Let S_1 and S_2 be nonempty subsets of $\overline{\mathbb{R}}$, and let H be a 2-increasing function with domain $S_1 \times S_2$. Let x_1, x_2 be in S_1 with $x_1 \leq x_2$, and let y_1, y_2 be in S_2 with $y_1 \leq y_2$. Then the function $t \mapsto H(t, y_2) - H(t, y_1)$ is nondecreasing on S_1, and the function $t \mapsto H(x_2, t) - H(x_1, t)$ is nondecreasing on S_2*

Proof. Since H is 2-increasing we know that

$$H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \geq 0,$$

so that

$$H(x_2, y_2) - H(x_1, y_2) \geq H(x_2, y_1) - H(x_1, y_1)$$

since $y_2 \geq y_1$, $t \mapsto H(x_2, t) - H(x_1, t)$ must be a nondecreasing function. The proof for the function $t \mapsto H(t, y_2) - H(t, y_1)$ is exactly same. □

We will generalize the definition of *grounded*. Assume S_1 has a least element a_1 and that S_2 has a least element a_2. Then a function H from $S_1 \times S_2$ to $\mathbb{R}$ is *grounded* if $H(u, a_2) = 0 = H(a_1, v)$ for all (u, v) in $S_1 \times S_2$. Notice that this still means that a copula is *grounded* if $C(u, 0) = 0 = C(0, v)$ for all (u, v) in $\mathbb{I}$ since a copula goes by definition from $\mathbb{I}^2$ to $\mathbb{I}$. If we now add the additional requirement that H is grounded we get next lemma.

Lemma 2. *Let S_1 and S_2 be nonempty subsets of $\overline{\mathbb{R}}$, and let H be a grounded 2-increasing function with domain $S_1 \times S_2$. Then H is nondecreasing in each argument.*

Proof. We let x_1 and y_1 be equal to the least element in S_1 and S_2 and since H is grounded the result follows immediately. □

This proof does not mean that a 2-increasing function on its own implies that the function is non-decreasing in each argument, or the other way, that a function which is non-decreasing in each argument implies that it is a 2-increasing function. Two counterexamples of this taken from [2, p. 8] shows this.

Example 1. Let C be defined on $\mathbb{I}^2$ by $C(x,y) = \max(x,y)$. Then C is obviously non decreasing in each argument. However, $V_C([0,1]\times[0,1]) = C(1,1) - C(1,0) - C(0,1) + C(0,0) = 1 - 1 - 1 + 0 = -1$, which means that the function is not a 2-increasing function.

For the next example we need a lemma first

Lemma 3. $(2x-1)(2y-1)$ *is a 2-increasing function.*

Proof. To show that $(2x-1)(2y-1)$ is a 2 increasing function. we have to show that $C(x_2,y_2) - C(x_1,y_2) - C(x_2,y_1) + C(x_1,y_1) \geq 0$ for all $0 \leq x_1 \leq x_2 \leq 1$ and $0 \leq y_1 \leq y_2 \leq 1$ We calculate the value of $V_C([x_1,x_2]\times[y_1,y_2])$ and we get

$$\begin{aligned}
(2x_2-1)(2y_2-1) - (2x_2-1)(2y_1-1) - (2x_1-1)(2y_2-1) + (2x_1-1)(2y_1-1) \\
&= 4x_2y_2 - 2x_2 - 2y_2 + 1 - 4x_2y_1 + 2x_2 + 2y_1 - 1 \\
&\quad - 4x_1y_2 + 2x_1 + 2y_2 - 1 + 4x_1y_1 - 2x_1 - 2y_1 + 1 \\
&= 4(x_2y_2 - x_2y_1 - x_1y_2 + x_1y_1) \\
&= 4(x_2[y_2 - y_1] + x_1[y_1 - y_2]) \\
&= 4(x_2 - x_1)(y_2 - y_1) \geq 0
\end{aligned}$$

□

Now we use the previous lemma in this example.

Example 2. Let C be defined on $\mathbb{I}^2$ by $C(x,y) = (2x-1)(2y-1)$. Then C is a 2-increasing function, however it is a decreasing function of x for each y in $(0,\frac{1}{2})$ and a decreasing function of y for each x in $(0,\frac{1}{2})$, since $2y-1$ and $2x-1$ is negative when x and y is in the interval $(0,\frac{1}{2})$.

We proceed with bounds for copulas. We already know that $0 \leq C(u,v) \leq 1$ for all (u,v) in $\mathbb{I}^2$, but tighter bounds exist.

Theorem 4. *Let $C(u,v)$ be a copula. Then for all (u,v) in $\mathbb{I}^2$*

$$\max(u+v-1,0) \leq C(u,v) \leq \min(u,v).$$

Proof. Let (u, v) be a point in $\mathbb{I}^2$. Since a copula is increasing in each argument we have that $C(u, v) \leq C(u, 1) = u$. A similar argument gives $C(u, v) \leq C(1, v) = v$, and we obtain that $C(u, v) \leq \min(u, v)$. For the next inequality

$$0 \leq V_C\left([u, 1] \times [v, 1]\right) = C(1, 1) - C(u, 1) - C(1, v) + C(u, v) = 1 - u - v + C(u, v),$$

so that

$$C(u, v) \geq u + v - 1.$$

And since $0 \leq C(u, v)$ we conclude that $\max(u + v - 1, 0) \leq C(u, v)$. □

An interesting question is if these bounds are actually copulas. It turns out they actually are, and we will denote them as $M(u, v) = \min(u, v)$ and $W(u, v) = \max(u + v - 1, 0)$. Another copula that is of special interest is the product copula $\Pi(u, v) = uv$ as it has a link to independence. We will come back to this copula later, but first we prove that they are all copulas.

Lemma 5. *$M(u, v) = \min(u, v)$, $W(u, v) = \max(u + v - 1, 0)$ and $\Pi(u, v) = uv$ are copulas*

Proof. (1) We start with $M(u, v)$. $M(u, v)$ is obviously grounded and condition 1b) from Definition 1 holds. What is left is showing that that $M(u, v)$ is a 2-increasing function, or in other words that

$$\min(x_2, y_2) - \min(x_1, y_2) - \min(x_2, y_1) + \min(x_1, y_1) \geq 0$$

for $0 \leq x_1 \leq x_2 \leq 1$ and $0 \leq y_1 \leq y_2 \leq 1$. We start by assuming that $x_1 \leq y_1$, which means that x_1 is the least number which simplifies our earlier inequality to

$$\min(x_2, y_2) - \min(x_2, y_1) \geq 0.$$

Now there are three possibilities, $x_2 \leq y_1 \leq y_2$, $y_1 \leq x_2 \leq y_2$ or $y_1 \leq y_2 \leq x_2$. We start by assuming $x_2 \leq y_1 \leq y_2$ which gives us

$$\min(x_2, y_2) - \min(x_2, y_1) = x_2 - x_2 = 0 \geq 0.$$

The second inequality gives

$$\min(x_2, y_2) - \min(x_2, y_1) = x_2 - y_1 \geq 0$$

and the third gives

$$\min(x_2, y_2) - \min(x_2, y_1) = y_2 - y_1 \geq 0.$$

We continue with $W(u, v)$. $W(0, v) = \max(v - 1, 0) = 0$ since $v - 1 \leq 1 - 1 = 0$, the argument for $W(u, 0)$ is similar. $W(u, 1) = \max(u, 0) = u$ and similarly we have $W(1, v) = \max(v, 0) = v$. Next we show that $W(u, v)$ is a 2-increasing function, that is

$$\max(x_2+y_2-1, 0)-\max(x_1+y_2-1, 0)-\max(x_2+y_1-1, 0)+\max(x_1+y_1-1, 0) \geq 0$$

with $0 \leq x_1 \leq x_2 \leq 1$ and $0 \leq y_1 \leq y_2 \leq 1$. We first look at the the case when $x_2 + y_2 < 1$, when we get that

$$\begin{aligned}\max(x_2 + y_2 - 1, 0) - \max(x_1 + y_2 - 1, 0) - \max(x_2 + y_1 - 1, 0)+ \\ \max(x_1 + y_1 - 1, 0) = 0 - 0 - 0 + 0 = 0 \geq 0,\end{aligned}$$

so we can safely assume that $x_2 + y_2 \geq 1$ in the rest of the cases. We now look at the case that $x_1 + y_1 \geq 1$

$$\begin{aligned}\max(x_2+y_2-1, 0)-\max(x_1+y_2-1, 0)-\max(x_2+y_1-1, 0)-\max(x_1+y_1-1, 0) \\ = (x_2 + y_2 - 1) - (x_1 + y_2 - 1) - (x_2 + y_1 - 1) + (x_1 + y_1 - 1) \\ = 0 \geq 0.\end{aligned}$$

This means we will further assume $x_1 + y_1 < 1$. After all this we only have 4 cases left to check. Each of $x_1 + y_2$ and $x_2 + y_1$ can be ≥ 1 or < 1. We start with the case where they both are ≥ 1:

$$\begin{aligned}\max(x_2+y_2-1, 0)-\max(x_1+y_2-1, 0)-\max(x_2+y_1-1, 0)+\max(x_1+y_1-1, 0) \\ = (x_2 + y_2 - 1) - (x_1 + y_2 - 1) - (x_2 + y_1 - 1) + 0 \\ = 1 - x_1 - y_1 \geq 0.\end{aligned}$$

Next we assume $x_1 + y_2 \geq 1$ and $x_2 + y_1 < 1$

$$\begin{aligned}\max(x_2+y_2-1, 0)-\max(x_1+y_2-1, 0)-\max(x_2+y_1-1, 0)+\max(x_1+y_1-1, 0) \\ = (x_2 + y_2 - 1) - (x_1 + y_2 - 1) - 0 + 0 \\ = x_2 - x_1 \geq 0.\end{aligned}$$

We now assume $x_1 + y_2 < 1$ and $x_2 + y_1 \geq 1$

$$\begin{aligned}\max(x_2+y_2-1,0)-\max(x_1+y_2-1,0)-\max(x_2+y_1-1,0)+\max(x_1+y_1-1,0)\\ = (x_2+y_2-1)-0-(x_2+y_1-1)+0\\ = y_2-y_1 \geq 0.\end{aligned}$$

For the last case let $x_1 + y_2 < 1$ and $x_2 + y_1 < 1$

$$\begin{aligned}\max(x_2+y_2-1,0)-\max(x_1+y_2-1,0)-\max(x_2+y_1-1,0)+\max(x_1+y_1-1,0)\\ = (x_2+y_2-1)-0-0+0 \geq 0.\end{aligned}$$

Now for the last copula $\Pi(u, v)$. Showing that $\Pi(u, v)$ is grounded and that 1b) holds is rather straightforward. Again we are left with showing that our copula is a 2-increasing function.

$$x_2y_2 - x_2y_1 - x_1y_2 + x_1y_1 = x_2(y_2 - y_1) - x_1(y_2 - y_1) = (x_2 - x_1)(y_2 - y_1) \geq 0$$

This is true because of our requirement that $0 \leq x_1 \leq x_2 \leq 1$ and $0 \leq y_1 \leq y_2 \leq 1$ which concludes our proof. □

There are also more bounds on copulas, but to prove these we first need to introduce *margins*. Let S_1 and S_2 have the a greatest element b_1 and b_2. Then a function H from $S_1 \times S_2$ into $\mathbb{R}$ has *margins*, and those margins are defined as $x \mapsto H(x, b_1)$ with domain S_1 and $y \mapsto H(b_2, y)$ with domain S_2. We will often define these margins as F and G, respectively. We continue with a lemma considering grounded 2-increasing functions with margins.

Lemma 6. *Let H be a 2-increasing function from $S_1 \times S_2$ into $\mathbb{R}$ where S_1 and S_2 are nonempty subsets of $\overline{\mathbb{R}}$ and F and G are H's margins. Let (x_1, y_1) and $(x_2.y_2)$ be any points in $S_1 \times S_2$. 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. From the triangle inequality, we have

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

If we now assume that $x_1 \leq x_2$ we have from Lemma 1 that $H(x_2, y) - H(x_1, y) \leq F(x_2) - F(x_1)$. We have from Lemma 2 that $0 \leq H(x_2, y) - H(x_1, y)$. Combining these two we get $0 \leq H(x_2, y) - H(x_1, y) \leq F(x_2) - F(x_1)$. If we now assume $x_2 \leq x_1$ we get similar inequalities. Hence we have that $|H(x_2, y_2) - H(x_1, y_2)| \leq |F(x_2) - F(x_1)|$ for all x_1 and x_2 in S_1. Combining this with a similar process for any y_1 and y_2 in S_2 we complete the proof. □

Since a copula is a 2-increasing function with margins we directly get this theorem from the previous lemma.

Theorem 7. *Let C be a copula. Then for every (u_1, u_2), (v_1, v_2) in $\mathbb{I}^2$,*

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

This gives us some limitations on how fast a copula can increase or decrease in any direction.

We have now seen that if a function is nondecreasing in each argument, it is not implied that it is a 2-increasing function. And a function being 2-increasing does not imply that is is nondecreasing in each argument. So the last question is: if you have a function that is grounded and definition 1b holds and the function is nondecreasing in each argument, will this imply that the function is also a 2-increasing? The answer is no and we have the following counterexample taken from [2, p. 16]

Example 3. Let

$$Q(u,v) = \begin{cases} \min\left(u, v, \frac{1}{3}, u+v-\frac{2}{3}\right), & \frac{2}{3} \leq u+v \leq \frac{4}{3} \\ \max\left(u+v-1, 0\right), & \text{otherwise} \end{cases}$$

(see Figure 1). We want to show that Q (1) is grounded, (2) definition 1b) holds (3) $W(u,v) \leq Q(u,v) \leq M(u,v)$, (4) is continuous, (5) is increasing in each argument, (6) is not a 2-increasing function, and (7) satisfies Theorem 7.

Proof. (1) Assume that $u = 0$. Then we have two cases to check: When $\frac{2}{3} \leq v \leq 1$, and otherwise. In both cases it is easy to check that $Q(0, v) = 0$. The argument is the same for $Q(u, 0)$.

(2) We first assume that $u = 1$. Here we also have two cases: When $0 \leq v \leq \frac{1}{3}$ and otherwise. Assume first that $0 \leq v \leq \frac{1}{3}$. Then

$$Q(1, v) = \min\left(1, v, \frac{1}{3}, v + \frac{1}{3}\right) = v$$

In the second case we get that $Q(1, v) = \max(v, 0) = v$.

(3) We remind the reader that $0 \leq u \leq 1$ and $0 \leq v \leq 1$. We start by showing that $W(u, v) \leq Q(u, v)$. Since $Q(u, v) = W(u, v)$ when $u + v < \frac{2}{3}$ or $u + v > \frac{4}{3}$, the only case we have to check is when $\frac{2}{3} \leq u + v \leq \frac{4}{3}$. In this case we must show that

$$\min\left(u, v, \frac{1}{3}, u + v - \frac{2}{3}\right) \geq \max\left(u + v - 1, 0\right).$$

We check that each argument is greater than or equal to each argument of the maximum. First $u \geq u - (1 - v) = u + v - 1$ and we also have $u \geq 0$. The argument for v is exactly the same. $\frac{1}{3} > 0$ and $u + v - \frac{2}{3} \geq \frac{2}{3} - \frac{2}{3} = 0$.

We now take a look at the claim that $Q(u, v) \leq M(u, v)$. We first check the case when $\frac{2}{3} \leq u + v \leq \frac{4}{3}$, for which me must show that

$$\min\left(u, v, \frac{1}{3}, u + v - \frac{2}{3}\right) \leq \min(u, v).$$

This is true because a minimum of u and v will be greater than or equal to a minimum of u, v and more arguments. We are now left with showing that $Q(u, v) \leq M(u, v)$ when $u + v < \frac{2}{3}$ or $\frac{4}{3} < u + v$. But in this case $Q(u, v) = W(u, v)$ by definition of Q and we have already seen that $W(u, v) \leq M(u, v)$ from Theorem (4), and we conclude that $Q(u, v) \leq M(u, v)$.

(4) We remind the reader of the definition of Q

$$Q(u, v) = \begin{cases} \min\left(u, v, \frac{1}{3}, u + v - \frac{2}{3}\right), & \frac{2}{3} \leq u + v \leq \frac{4}{3} \\ \max\left(u + v - 1, 0\right), & \text{otherwise;} \end{cases}$$

First Q is continuous on the three regions of its domain, since a maximum or minimum of continuous functions is continuous. What we are left with showing is that Q is continuous on the boundary of the regions. First, when $u + v = \frac{2}{3}$

$$Q(u, v) = \min\left(u, v, \frac{1}{3}, u + v - \frac{2}{3}\right) = 0,$$

and when $u+v=\frac{4}{3}$

$$Q(u,v)=\min\left(u,v,\frac{1}{3},u+v-\frac{2}{3}\right)=\frac{1}{3}.$$

The limits when approaching the boarder from the other region are

$$Q(u,v)=\max\left(u+v-1,0\right)\to 0$$

when $u+v\to\frac{2}{3}$

$$Q(u,v)=\max\left(u+v-1,0\right)\to\frac{1}{3}$$

when $u+v\to\frac{4}{3}$. This means that $\lim_{(x,y)\mapsto(u,v)} Q(x,y)=Q(u,v)$ for all (u,v) in $\mathbb{I}^2$ and we conclude that Q is continuous on $\|x\|$.

(5) Assume that v is fixed. Then Q is non-decreasing both when $\frac{2}{3}\leq u+v\leq\frac{4}{3}$ and otherwise. Since Q is continuous it follows that $u\to Q(u,v)$ is non-decreasing. The proof that Q is non-decreasing in v when u is fixed is similar.

(6) We will show that Q is not a 2-increasing function by considering $V_Q\left(\left[\frac{1}{3},\frac{2}{3}\right]^2\right)$. We calculate

$$Q\left(\frac{2}{3},\frac{2}{3}\right)=\frac{1}{3},\quad Q\left(\frac{2}{3},\frac{1}{3}\right)=\frac{1}{3},\quad Q\left(\frac{1}{3},\frac{2}{3}\right)=\frac{1}{3},\quad Q\left(\frac{1}{3},\frac{1}{3}\right)=0$$

and we conclude that

$$V_Q\left(\left[\frac{1}{3},\frac{2}{3}\right]^2\right)=\frac{1}{3}-\frac{1}{3}-\frac{1}{3}+0=-\frac{1}{3}$$

□

(7) We want to show that $|Q(u_2,v_2)-Q(u_1,v_1)|\leq|u_2-u_1|+|v_2-v_1|$ for all u_1, u_2, v_1 and v_2. We assume without loss of generality that $Q(u_2,v_2)\geq Q(u_1,v_1)$ and we divide the proof into 3 cases where $4/3\leq u_2+v_2$, $2/3\leq u_2+v_2\leq 4/3$ and $u_2+v_2\leq 2/3$.

(I) Assume $4/3\leq u_2+v_2$: Then $Q(u_2,v_2)=u_2+v_2-1$ and we get

$$|Q(u_2,v_2)-Q(u_1,v_1)|\leq|u_2+v_2-1-(u_1+v_1-1)|=|u_2-u_1+v_2-v_1|\leq|u_2-u_1|+|v_2-v_1|$$

To justify the first inequality we have to check check that $u_1+v_1-1 \leq Q(u_1, v_1)$ for all the possible values of $Q(u_1, v_1)$. (a) $Q(u_1, v_1) = u_1+v_1-1$: Indeed $u_1+v_1-1 \leq u_1+v_1-1$. (b) $Q(u_1, v_1) = u_1$: $u_1 \geq u_1-(1-v_1) = u_1+v_1-1$. We have a similar argument when $Q(u_1, v_1) = v_1$. (c) $Q(u_1, v_1) = 1/3$: $u_1+v_1-1 \leq 1/3$ since $u_1+v_1 \leq 4/3$ when $Q(u_1, v_1) = 1/3$. (d) $Q(u_1, v_1) = u_1+v_1-2/3$: $u_1+v_1-1 \leq u_1+v_1-2/3$. (e) $Q(u_1, v_1) = 0$: $u_1+v_1-1 \leq -1/3 \leq 0$ since $u_1+v_1 \leq 2/3$.

(II) Assume $1/3 \leq u_2+v_2 \leq 4/3$. First, also $u_1+v_1 \leq 4/3$, since otherwise $Q(u_1, v_1) = u_1+v_1-1 > 1/3 \geq Q(u_2, v_2)$. We now check for all the possible values of $Q(u_1, v_1)$. (a) $Q(u_1, v_1) = 1/3$:

$$|Q(u_2, v_2) - Q(u_1, v_1)| \leq |1/3 - 1/3| = 0$$

The first inequality is true because when $1/3 \leq u_2+v_2 \leq 4/3$ we have that $Q(u_2, v_2) = \min(u_2, v_2, 1/3, u_2+v_2-2/3) \leq 1/3$. (b) $Q(u_1, v_1) = u_1+v_1-2/3$ or $Q(u_1, v_1) = 0$

$$\begin{aligned}|Q(u_2, v_2) - Q(u_1, v_1)| &\leq |u_2+v_2-2/3-(u_1+v_1-2/3)| \\ &= |u_2-u_1+v_2-v_1| \leq |u_2-u_1| + |v_2-v_1|\end{aligned}$$

We have the first inequality because $Q(u_2, v_2) = \min(u_2, v_2, 1/3, u_2+v_2-2/3) \leq u_2+v_2-2/3$ and $u_1+v_1-2/3 \leq Q(u_1, v_1)$ which is obviously true when $Q(u_1, v_1) = u_1+v_1-2/3$, and it is also true when $Q(u_1, v_1) = 0$ since $u_1+v_1 \leq 2/3$ by definition of Q when $Q(u_1, v_1) = 0$. (c) $Q(u_1, v_1) = u_1$:

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

We justify the first inequality by noticing that $Q(u_2, v_2) \leq Q(u_2, 1) = u_2$ since Q is increasing in each argument. (d) $Q(u_1, v_1) = v_1$: Similar argument as when $Q(u_1, v_1) = u_1$.

(III) $u_2+v_2 \leq 2/3$: $Q(u_2, v_2) = Q(u_1, v_1) = 0$ which means that

$$|Q(u_2, v_2) - Q(u_1, v_1)| = |0-0| \leq |u_2-u_1| + |v_2-v_1|.$$

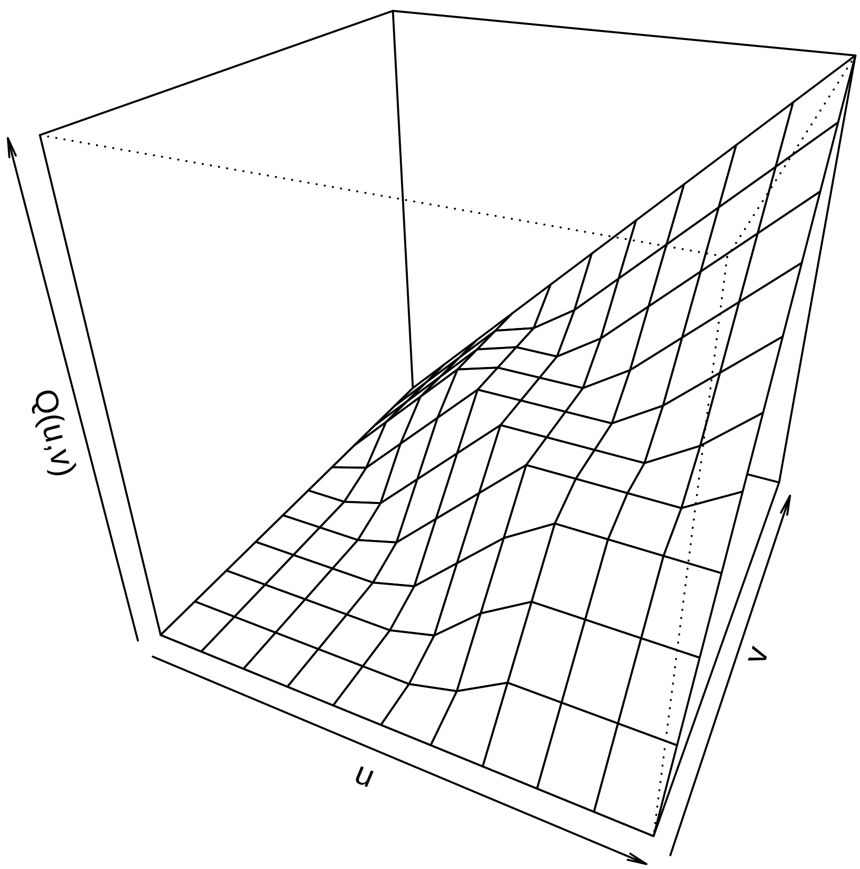

Figure 2.1: A 3D plot of the function $Q(u, v)$ of example 3.

Chapter 3

Another perspective on copulas

So far we have considered the three requirements for a function to be a copula in Definition 1. To understand why these are the requirements for a copula we can look at it from a completely different angle. But first we remind the reader of the definition of a distribution function.

Definition 2. *A function F is a distribution function if these two statements are true:*

(1) *F is nondecreasing.*
(2) $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$.

We also need the definition of the join distribution function.

Definition 3. *A function H of two variables is a joint distribution function if these two statements holds:*

(1) *H is a 2-increasing function.*
(2) $\lim\limits_{y\to-\infty} H(x,y) = 0$ *for all x,* $\lim\limits_{x\to-\infty} H(x,y) = 0$ *for all y and* $\lim\limits_{(x,y)\to(\infty,\infty)} H(x,y) = 1$.

This means that a joint distribution function is grounded, and has margins $x \mapsto \lim_{y\to\infty} H(x,y)$ and $y \mapsto \lim_{x\to\infty} H(x,y)$.

A special case is when (U, V) is a pair of two random variables where U and V both have the marginal distribution $U[0,1]$ (uniform distribution on $[0,1]$). If we now define H as the joint distribution, namely $H(u,v) = P(U \leq u \cap V \leq v)$ then H is a copula. With this in mind it's understandable why all copulas have to be grounded as $H(u,v) = 0$ if u or v is equal to 0. To see that $H(1,v) = v$ we use the fact that a joint distribution function with domain $\mathbb{I}^2$ will have the property that $H(1,v) = G(v)$ where G is the marginal distribution of V. Since $G \sim U[0,1]$ we have that $G(v) = v$. Conversely a copula C can be seen as a joint cdf with marginals uniformly distributed on $\mathbb{I}$.

To understand why copulas have to be 2-increasing it can be useful to have in mind that

$$H(u_2, v_2) - H(u_1, v_2) - H(u_2, v_1) + H(u_1, v_1) = P(u_1 \leq U \leq u_2 \cap v_1 \leq V \leq v_2) \geq 0.$$

We can use this new insight to show that $M(u,v)$ and $W(u,v)$ are copulas (see page 11). First consider the case there $U = V$ Then $H(u,v) = P(U \leq u \cap U \leq v) = P(U \leq \min(u,v)) = \min(u,v) = M(u,v)$. And since it's a joint distribution function with uniform marginals on $[0,1]$, it is a copula by the above remarks.

Next consider the case that $V = 1 - U$. Then $H(u,v) = P(U \leq u \cap 1 - U \leq v) = P(U \leq u \cap 1 - v \leq U) = P(1 - v \leq U \leq u) = \max(u + v - 1, 0) = W(u,v)$. Also $1 - U$ is uniform on $[0,1]$ and with same reasoning as in the last example we conclude that $W(u,v)$ is a copula.

Chapter 4

Sklar's Theorem

Theorem 8 (Sklar's Theorem)**.** *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x, y in $\overline{\mathbb{R}}$*

$$H(x, y) = C(F(x), G(y)).$$

If F and G are continuous then C is unique. Conversely, if C is a copula and F and G are distribution functions, then the function H is a joint distribution function with margins F and G.

Proof. We prove the Theorem in the case that F and G are continuous, both with range $\mathbb{I}$. The readers is referred to [2, p. 21] for the general case. The joint distribution H satisfies the conditions in Lemma 6, since H is a 2-increasing function from $\overline{\mathbb{R}}^2$ to $\mathbb{R}$. This gives us that

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

for all pairs of (x_1, y_1) and (x_2, y_2) in $\overline{\mathbb{R}}^2$. If $F(x_2) = F(x_1)$ and $G(y_2) = G(y_1)$ it follows that $H(x_2, y_2) = H(x_1, y_1)$. This implies that the function C defined by $C(F(x), G(y)) = H(x, y)$ is well defined with domain which is the range of $F \times$ range of G, which is equal to $\mathbb{I}^2$ when F and G are continuous.

To verify that C is a copula we have to check the conditions in Definition 1. We start with Definition 1 (a): Let $v \in \mathbb{I}$. Then since G is continuous, there exists y such that

$G(y) = v$, and

$$C(0, v) = C(F(-\infty), G(y)) = H(-\infty, y) = 0.$$

Similarly let $u \in \mathbb{I}$. Then since G is continuous, there exists x such that $F(x) = u$,

$$C(u, 0) = C(F(x), G(-\infty)) = H(x, -\infty) = 0.$$

Next is Definition 1(b): Again, for $v \in \mathbb{I}$, assume $G(y) = v$. Since $F(\infty) = 1$ we have

$$C(1, v) = C(F(\infty), G(y)) = H(\infty, y) = G(y)$$

and similarly, for $u \in \mathbb{I}$, assume $F(x) = u$. Since $G(\infty) = 1$

$$C(u, 1) = C(F(x), G(\infty)) = H(x, \infty) = F(x).$$

For Definition 1(c) we want to show that

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

when $u_1 \leq u_2$ and $v_1 \leq v_2$. We choose x_i and y_i such that $F(x_i) = u_i$ and $G(y_i) = v_i$, $i = 1, 2$. This translates our problem into showing that

$$H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \geq 0,$$

which is true because H is a 2-increasing function since it is a joint distribution function. And $u_1 \leq u_2$ if and only if $x_1 \leq x_2$, and $v_1 \leq v_2$ if and only if $y_1 \leq y_2$ which concludes one direction of our proof.

Now for the converse direction. We want to show that if C is a copula and F and G are distribution functions then H is a joint distribution function with F and G as its marginals. We start with showing that $C(F(x), G(y))$ is a joint distribution function.

Definition 3 (a): We have to show that H is 2-increasing, that is,

$$H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \geq 0$$

for all $x_1 \leq x_2$ and $y_1 \leq y_2$. We now define u_i and v_i as previously in this proof. Since we have already shown that $F(x_1) \leq F(x_2)$ and $G(y_1) \leq G(y_2)$ if and only if $u_1 \leq u_2$ and $v_1 \leq v_2$ our problem translates into showing

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

when $u_1 \leq u_2$ and $v_1 \leq v_2$, which is the definition 1 (c) of a copula.

Definition 3 (b): We check if the limits are correct

$$H(\infty, \infty) = C(F(\infty)), G(\infty)) = C(1, 1) = 1,$$
$$H(-\infty, y) = C(F(-\infty), G(y)) = C(0, G(y)) = 0$$

for all y, and

$$H(x, -\infty) = C(F(x), G(-\infty)) = C(F(x), 0) = 0$$

for all x.

We have now proved that $C(F(x), G(y))$ is a joint distribution function. What's left to prove is that F and G are H's marginal distributions,

$$H(\infty, y) = C(F(\infty), G(y)) = C(1, G(y)) = G(y)$$

for all y, and similarly

$$H(x, \infty) = C(F(x), G(\infty)) = C(F(x), 1) = F(x)$$

for all x, which completes our proof.

□

So a copula is a connection between marginal distribution and joint distributions. This is also the reason why copula is called copula as it "couples" marginal distributions together into a joint distribution. From Sklar's theorem we know that we can construct a joint distribution function if we have two marginal distributions F, G and a copula C. A question which might be asked is if you can construct a copula from a joint distribution and its marginals. The answer is yes, for continuous distributions.

Corollary 9. *Let H be a joint distribution function with its marginals F and G contínous with range $\mathbb{I}$ and let C be the unique copula such that $H(x, y) = C(F(x), G(y))$ for all $x, y \in \overline{\mathbb{R}}$. Then for all $u, v \in \mathbb{I}$, $C(u, v) = H(F^{-1}(u), G^{-1}(v))$, where $F^{-1}(u)$ denotes any x such that $F(x) = u$ and $G^{-1}(v)$ denotes any y such that $G(y) = v$.*

Proof. Let $u, v \in \mathbb{I}$. Assume $F(x) = u$ and $G(y) = v$. We then have from Sklar's Theorem that

$$H(F^{-1}(u), G^{-1}(v)) = H(x, y) = C(F(x), G(y)) = C(u, v).$$

□

We have earlier stated that the copula $\Pi(u,v) = uv$ has a link to independence. The reason is that the joint distribution of two random variables X and Y is $H(x,y) = F(x)G(y)$ if and only if X and Y are independent. It follows from corollary 9 that two random variables are independent if and only if their copula is the independent copula $\Pi(u,v) = uv$. We summarize this in the next corollary.

Corollary 10. *Let X and Y be continuous random variables with F and G their respective distribution functions. Then their copula is $C(u,v) = uv$ if and only if X and Y are independent random variables.*

Example 4. We can use Corollary 9 to create the Gaussian copula. We start with a random vector (X, Y) which we assume have a bivariate normal distribution H with the parameters μ_x, μ_y, σ_x, σ_y and ρ. Since we want to use that $C(u,v) = H(F^{-1}(u), G^{-1}(v))$ we have to figure out what $F^{-1}(u)$ and $G^{-1}(v)$ is. Since H is a bivariate normal distribution we know that the marginal distribution F G are normal distributions with parameters μ_x, μ_y, σ_x and σ_y. We now use this to calculate $F^{-1}(u)$. Firstly we have that

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu_x}{\sigma_x} \leq \frac{x - \mu_x}{\sigma_x}\right) = \Phi\left(\frac{x - \mu_x}{\sigma_x}\right)$$

where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ variable. We use this to find the inverse of the distribution function

$$\begin{aligned} F^{-1}(u) = x \Leftrightarrow u = F(x) \Leftrightarrow u = \Phi\left(\frac{x - \mu_x}{\sigma_x}\right) &\Leftrightarrow \Phi^{-1}(u) = \frac{x - \mu_x}{\sigma_x} \\ &\Leftrightarrow x = \mu_x + \sigma_x \Phi^{-1}(u) \\ &\Leftrightarrow F^{-1}(u) = \mu_x + \sigma_x \Phi^{-1}(u). \end{aligned}$$

Similarly,

$$G^{-1}(v) = \mu_y + \sigma_y \Phi^{-1}(v).$$

We also know that

$$\begin{aligned} H(x,y) = {} & \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{x} \int_{-\infty}^{y} \\ & \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(z-\mu_x)^2}{\sigma_x^2} + \frac{(w-\mu_y)^2}{\sigma_y} - \frac{2\rho(z-\mu_x)(w-\mu_y)}{\sigma_x\sigma_y}\right]\right) dz\, dw. \end{aligned}$$

We plug in $F^{-1}(u)$ for x and $G^{-1}(v)$ for y and we get

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{\mu_x+\sigma_x\Phi^{-1}(u)} \int_{-\infty}^{\mu_y+\sigma_y\Phi^{-1}(v)}$$
$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(z-\mu_x)^2}{\sigma_x^2} + \frac{(w-\mu_y)^2}{\sigma_y} - \frac{2\rho(z-\mu_x)(w-\mu_y)}{\sigma_x\sigma_y}\right]\right) dz\, dw.$$

Now we introduce a change of variables namely $s = \frac{z-\mu_x}{\sigma_x}$ and $t = \frac{w-\mu_y}{\sigma_y}$. This gives that $\sigma_x ds = dz$ and $\sigma_y dt = dw$. We also have that $z = -\infty \Rightarrow s = -\infty$, $z = \mu_x + \sigma_x\Phi^{-1}(u) \Rightarrow s = \Phi^{-1}(u)$, $w = -\infty \Rightarrow t = -\infty$ and $w = \mu_y + \sigma_y\Phi^{-1}(v)$. We then use Corollary 9 and get that

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left[\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right] ds\, dt.$$

We see that μ_x, μ_y, σ_x and σ_y has disappeared in the calculations. For that reason, we use notation $C_\rho(u,v)$ when we are talking about the Gaussion copula, since it only depends on the parameter ρ. The correlation structure is seperated from the marginal distributions.

Chapter 5

Multivariate Copulas

Until now we have focused specifically on the copula with dimension equal to 2. We shall soon define copulas with dimension $n \geq 2$ where n is an integer. We will start with some new notation taken from [2, p. 43].

Let $\overline{\mathbb{R}}^n$ denote the cartesian product $\overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \cdots \times \overline{\mathbb{R}}$. For vectors $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, $\boldsymbol{a} \leq \boldsymbol{b}$ means that $a_k \leq b_k$ for all k. We will also denote by $[\boldsymbol{a}, \boldsymbol{b}]$ be the n-dimensional box or an n-box $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$. The vertices of an n-dimensional box can be described as $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)$ where c_k is equal to a_k or b_k for all k. An n-place real function H is a function whose domain is a subset of $\overline{\mathbb{R}}^n$ and its range is a subset of $\mathbb{R}$. We are now ready for the definition of the H-volume of a box B.

Definition 4. *Let $S_1, S_2, \ldots S_n$ be nonempty subsets of $\overline{\mathbb{R}}$, and let H be an n-place real function with domain $S_1 \times S_2 \times \cdots \times S_n$. Let $B = [\boldsymbol{a}, \boldsymbol{b}]$ be an n-box with all vertices in the domain of H. Then the H-volume of B is given by*

$$V_H(B) = \sum \operatorname{sgn}(\boldsymbol{c}) H(\boldsymbol{c}),$$

where the sum is taken over all the vertices $\mathbf{c}$ of B. Notice that this means that there are 2^n parts in the sum. $\operatorname{sgn}(\boldsymbol{c})$ *is given by*

$$\operatorname{sgn}(\boldsymbol{c}) = \begin{cases} 1 & \text{if } c_k = a_k \text{ for an even number of } k\text{'s}, \\ -1 & \text{if } c_k = a_k \text{ for an odd number of } k\text{'s}. \end{cases}$$

This definition extends the previous definition naturally. Now that we have defined the H-volume we can give the requirements for an n-place real function to be an n-dimensional copula.

Definition 5. *An n-dimensional copula is a function C from $\mathbb{I}^d$ to $\mathbb{I}$ with the following properties: a) For every $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$ in $\mathbb{I}^d$, $C(\boldsymbol{u}) = 0$ if $u_k = 0$ for at least one k.*
b) If all $u_k = 1$ except u_t, then $C(\boldsymbol{u}) = u_t$.
c) For all $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{I}^d$ such that $\boldsymbol{a} \leq \boldsymbol{b}$, $V_C[\boldsymbol{a}, \boldsymbol{b}] \geq 0$.

As we can see the multivariate definition of a copula is similar to the original definition. Lots of previous results we have proved for copulas with dimensional equal to 2 also holds for copula with dimension n, such as Sklar's Theorem and the corollary of Sklar's Theorem. Because of it's importance we state Sklar's Theorem in the multivariate case here. But first we have to define what an n-dimensional distribution function is, and what margins are in the multivariate sense.

If each S_k is nonempty and has a greatest element b_k, then the one dimensional margins of H is defined as $H_k(x) = (b_1, \ldots, b_{k-1}, x, b_{k+1}, \ldots, b_n)$. Higher dimensional margins are defined by fixing fewer arguments in H.

Definition 6. *An n-dimensional distribution function is a function H with domain $\overline{\mathbb{R}}^n$ such that: a) H is n-increasing, meaning that $V_H(B) \geq 0$ for all boxes B with vertices that lie in the domain of H.*
b)$H(\boldsymbol{t}) = 0$ for all $\boldsymbol{t}$ in $\overline{\mathbb{R}}^n$ such that $t_k = -\infty$ for at least one k, and $H(\infty, \infty, \ldots, \infty) = 1$.

Theorem 11 (Sklar's Theorem)**.** *Let H be an n-dimensional distribution function with margins $F_1, F_2, \ldots, F_n$. Then there exists an n-dimensional copula C such that for all $\boldsymbol{x}$ in $\overline{\mathbb{R}}^n$,*

$$H(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

If $F_1, F_2, \ldots, F_n$ are all continuous then C is unique. Conversely, if C is an n-dimensional copula and $F_1, F_2, \ldots, F_n$ are distribution functions, then the function H defined by the previous equation is an n-dimensional distribution function with margins $F_1, F_2, \ldots, F_n$.

We say that a copula C admits a density c if

$$c(\boldsymbol{u}) = \frac{\partial^n}{\partial u_n \ldots \partial u_1} C(u_1, \ldots, u_n)$$

exists and is integrateable [3, p. 13]. This means that if we differentiate the equation of Theorem 11 using the chain rule and $c(\boldsymbol{u})$ exists we get

$$h(x_1, x_2, \ldots, x_n) = c(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \prod_{i=1}^{n} f_i(x_i)$$

where h is the density function of H and f_i is the density function of the distribution function F_i.

Chapter 6

Parametric estimation

The notation in this chapter and the next one is very similar to that of [3, ch. 4]. Let us say we have random sample from a continous multivariate distribution function H and we want to estimate its marginals $F_1, F_2, \ldots, F_d$ and the copula C. We can do this by parametric or nonparamtric estimation of the marginals. We first start with the MLE parametric estimation.

MLE

To do this we need to some assumptions:
1) We know the distributions $F_1, F_2, \ldots, F_d$ except for the parameters of F_i which we call γ_i which lies in a subset of $\mathbb{R}^{p_j}$ where $p_j \in \mathbb{Z}^+$.
2) C comes from a specific family of copulas that admits a density.

So say we have n iid realizations $X_1, \ldots, X_n$ which all have dimension d. We then try to maximize the log likelihood function ℓ which is defined as

$$\ell(\gamma_1, \ldots, \gamma_d, \theta) = \sum_{i=1}^{n} \log[c_\theta(F_1(x_{i1}), F_2(x_{i2}), \ldots, F_d(x_{id})] + \sum_{j=1}^{d} \sum_{i=1}^{n} \log[f_j(x_{ij})], \quad (6.1)$$

and use the argument $[\hat{\gamma}_1, \ldots, \hat{\gamma}_d, \hat{\theta}]$ of our maximum as our estimation of the parameters. Since the parameter space can have a very high dimension it can be quite hard to find the maximum of the likelihood function. If an easier computational estimation is wanted IFME might be a more optimal solution:

IFME

IFME stands for *inference function for margins estimator* and is a two-stage estimator. It starts by estimating the parameters γ_j by $\hat{\gamma}_j$ for all $j \in \{1, 2, \ldots, d\}$ where $\hat{\gamma}_j$ is defined as

$$\hat{\gamma}_j = \underset{\gamma_j}{\text{argsup}} \sum_{i=1}^{n} \log[f_j(x_{ij})],$$

which means that $\hat{\gamma}_j$ is the MLE for each marginal distribution. We now use $\hat{\gamma}_j$ in our estimate of the unknown parameter θ of the copula family, that is,

$$\hat{\theta} = \underset{\theta}{\text{argsup}} \sum_{i=1}^{n} \log[c_\theta(F_{\hat{\gamma}_1}(x_{i1}), F_{\hat{\gamma}_2}(x_{i2}), \ldots, F_{\hat{\gamma}_d}(x_{id}))].$$

The drawback of this method is that this is not a maximum likelihood estimator.

Nonparametric estimation

MPLE

If we have the same situation as in the last section, that is n iid realizations with dimension d, and want to estimate the parameter θ of the family of copula without assuming which distributions the margins $F_1, F_2, \ldots, F_d$ are from, we can do this by estimating the margins $F_j(x)$ by $\hat{F}_j(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(x_{ij} \leq x)$ where $\mathbb{1}$ is the indicator function.

If we have chosen a family of copula C_θ we can then estimate θ by maximum likelihood methods. If we are in the (unlikely) scenario that the margins $F_1, F_2, \ldots, F_d$ are known we estimate θ by

$$\hat{\theta} = \underset{\theta}{\operatorname{argsup}} \sum_{i=1}^{n} \log[c_\theta(\boldsymbol{U}_i)]$$

which is the MLE and $\boldsymbol{U}_i$ is defined as

$$\boldsymbol{U}_i = (F_1(x_{i1}), F_1(x_{i2}), \ldots, F_1(x_{id})).$$

However if we don't know the margins we can use the nonparametric estimation of the margins and do basically the same thing. We just put a hat on $\boldsymbol{U}_i$ and it's not an MLE anymore but a *pseudo-likelihood estimator* (MPLE). We then estimate $\boldsymbol{U}_i$ by

$$\hat{\boldsymbol{U}}_i = (\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2}), \ldots, \hat{F}_d(x_{id}))$$

for all $i \in \{1, 2, \ldots, n\}$, so that $\hat{\boldsymbol{U}}_i$ serves as an estimate of the argument of the copula density in (6.1). One observation is that if we define the rank R_{ij} as the rank of x_{ij} among $x_{1j}, x_{2j}, \ldots, x_{nj}$,

$$\hat{\boldsymbol{U}}_i = \frac{1}{n+1}(R_{i1}, R_{i2} \ldots, R_{id}).$$

Example 5. We now do some estimation where we have 10, 100, 1000 and 10000 data points from the bivariate normal distribution with mean vector $\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance matrix $\sum = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. We now want to use our three methods of estimating the copula parameter ρ, and also the means and standard deviations from the two marginal distributions from the MLE and IMFE methods. The experiment is repeated 100 times, and the results are shown in Table 6.1, 6.2, 6.3 and 6.4.

For the MLE and IFME cases, all the results are the same up to 4 digits in the two estimation processes, negative ρ values where also not included in the tables since the estimation of the negative ρ values is the same as the estimation of the positive ρ values except for a switched sign. The numbers in Table 6.1, 6.2, 6.3 and 6.4 represent the mean of the 100 experiments while the error term is the empirical standard deviation.

The estimation of the parameter ρ is fairly close to the real value in most of the estimations, with the exception of the MPLE method with $N = 10$ where the mean of the estimation where pretty far from the true value in most of the cases. In general we see that the empirical standard deviation becomes small as N becomes bigger,

MLE,IFME estimate	ρ				
	0	0.2	0.4	0.6	0.8
ρ	0.0000 ± 0.0096	0.1993 ± 0.0092	0.3994 ± 0.0080	0.5995 ± 0.0061	0.7997 ± 0.0035
μ_1	-0.0007 ± 0.0119	-0.0026 ± 0.0095	-0.0027 ± 0.0092	-0.0028 ± 0.0090	-0.0028 ± 0.0089
μ_2	-0.0028 ± 0.0091	-0.0017 ± 0.0111	-0.0019 ± 0.0108	-0.0022 ± 0.0104	-0.0024 ± 0.0100
σ_1	1.0001 ± 0.0075	0.9996 ± 0.0069	0.9996 ± 0.0069	0.9995 ± 0.0068	0.9994 ± 0.0068
σ_2	0.9994 ± 0.0068	0.9997 ± 0.0072	0.9997 ± 0.0072	0.9996 ± 0.0070	0.9995 ± 0.0069
MPLE estimate	0	0.2	0.4	0.6	0.8
ρ	0.0000 ± 0.096	0.1995 ± 0.0092	0.3998 ± 0.0081	0.5999 ± 0.0061	0.7999 ± 0.0034

Table 6.1: 100 MLE, IFME, and MPLE estimations where $N = 10000$ for each estimation

there is also more empirical standard deviation in the MPLE method compared to the MLE and IFME method in general when N is equal to 10 and 100. However this is not the case when $N = 10$ with $\rho = 0.6, 0.8$ and $N = 100$ with $\rho = 0.6$, although in all of those three cases the mean ρ value is closer to the true ρ value in the MLE, IFME estimation compared to MPLE estimation. When N is equal to 10000 and 1000 the empirical standard deviation is about the same for the MLE, IFME method compared to the MPLE method. An explanation for this could be that when the number of samples get really big, then the information from the samples becomes a lot more important for the estimation compared to the extra information from the assumptions in the MLE and IFME methods.

When we compare the empirical standard deviation of ρ across the three method MLE, IMFE and MPLE for different values of ρ we see that $\rho = 0.8$ is the value which leads to the least amount of variance in the estimation. A possible reason for this is that the parameter space of ρ is $[-1, 1]$ which means that $\rho = 0.8$ is the closest value to the boundary of the parameter space which could lead to less variance in the estimation.

When we look at the empirical standard deviation of the other parameters μ_x, μ_y, σ_x and σ_y in table 6.1, 6.2 , 6.3 and 6.4 we see the same trend with ρ, namely higher empirical standard deviation when we have a low value for the number of samples N. But changing the true value of ρ does not seem to impact the empirical standard deviation for μ_x, μ_y, σ_x and σ_y.

We want to investigate why we get identical estimates of ρ using the MLE and IMFE method. We start by finding the the maximum likelihood estimator of the parameter ρ of the bivariate normal distribution when the other parameters $\mu_x, \mu_y, \sigma_x, \sigma_y$ are known and the samples are independent. The likelihood function is defined as

MLE, IFME estimate	ρ				
	0	0.2	0.4	0.6	0.8
ρ	-0.0023 ± 0.0318	0.1991 ± 0.0333	0.3991 ± 0.0292	0.5992 ± 0.0223	0.7995 ± 0.0126
μ_1	-0.0008 ± 0.0294	0.0037 ± 0.0263	0.0041 ± 0.0265	0.0045 ± 0.0269	0.0049 ± 0.0275
μ_2	0.0054 ± 0.0294	0.0047 ± 0.0322	0.0050 ± 0.0320	0.0052 ± 0.0317	0.0054 ± 0.0312
σ_1	1.0015 ± 0.0226	1.000 ± 0.0233	1.0000 ± 0.0235	1.000 ± 0.0237	1.0001 ± 0.0239
σ_2	1.0007 ± 0.0237	1.0021 ± 0.0211	1.0020 ± 0.0214	1.0018 ± 0.0219	1.0015 ± 0.0225
MPLE estimate	0	0.2	0.4	0.6	0.8
ρ	-0.0021 ± 0.0319	0.2013 ± 0.0336	0.4022 ± 0.0292	0.6018 ± 0.0221	0.8006 ± 0.0125

Table 6.2: 100 MLE, IFME, and MPLE estimations where $N = 1000$ for each estimation

MLE, IFME estimate	ρ				
	0	0.2	0.4	0.6	0.8
ρ	0.0044 ± 0.0987	0.1966 ± 0.0949	0.3962 ± 0.0830	0.5965 ± 0.0634	0.7977 ± 0.0358
μ_1	0.0150 ± 0.0959	0.0122 ± 0.0838	0.0112 ± 0.0829	0.0099 ± 0.0823	0.0081 ± 0.0822
μ_2	0.0035 ± 0.0842	-0.0068 ± 0.0941	-0.0053 ± 0.0926	-0.0036 ± 0.0909	-0.0014 ± 0.0887
σ_1	0.9953 ± 0.0732	0.9955 ± 0.0656	0.9950 ± 0.0639	0.9945 ± 0.0623	0.9937 ± 0.0612
σ_2	0.9920 ± 0.0623	0.9912 ± 0.0680	0.9910 ± 0.0674	0.9909 ± 0.0667	0.9910 ± 0.0655
MPLE estimate	0	0.2	0.4	0.6	0.8
ρ	0.0030 ± 0.1091	0.2087 ± 0.1045	0.4147 ± 0.0860	0.6129 ± 0.0620	0.8031 ± 0.0362

Table 6.3: 100 MLE, IFME, and MPLE estimations where $N = 100$ for each estimation

$$L(\rho) = \prod_{i=1}^{n} f(x_i, y_i)$$
$$= \prod_{i=1}^{n} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$
$$= \left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right]\right).$$

To find the maximum of the log likelihood function we take the logarithm of the likelihood function take the derivative and put it equal to zero.

$$\ell(\rho) = \ln L(\rho) = -n\ln(2\pi) - n\ln(\sigma_x) - n\ln(\sigma_y) - \frac{1}{2}\ln(1-\rho^2)$$
$$-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right]$$

MLE, IFME estimate	ρ				
	0	0.2	0.4	0.6	0.8
ρ	-0.0375 ± 0.3341	0.1988 ± 0.3156	0.3913 ± 0.2784	0.5878 ± 0.2180	0.7896 ± 0.1298
μ_1	-0.0155 ± 0.3015	0.0235 ± 0.3293	0.0275 ± 0.3286	0.0316 ± 0.3267	0.0359 ± 0.3228
μ_2	0.0431 ± 0.3093	0.0432 ± 0.2812	0.0445 ± 0.2838	0.0455 ± 0.2877	0.0458 ± 0.2936
σ_1	0.9259 ± 0.2226	0.9105 ± 0.2202	0.9123 ± 0.2183	0.9152 ± 0.2154	0.9197 ± 0.2119
σ_2	0.9307 ± 0.2100	0.9453 ± 0.2161	0.9445 ± 0.2158	0.9428 ± 0.2154	0.9399 ± 0.2141
MPLE estimate	0	0.2	0.4	0.6	0.8
ρ	-0.0653 ± 0.4684	0.2678 ± 0.4245	0.4841 ± 0.3579	0.6784 ± 0.2129	0.8327 ± 0.1047

Table 6.4: 100 MLE, IFME, and MPLE estimations where $N = 10$ for each estimation

$$\frac{d\ell(\rho)}{d\rho} = \frac{n\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\sum_{i=1}^{n}\left(\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right) - \frac{1}{2(1-\rho^2)}\sum_{i=1}^{n}\left(\frac{-2(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right) = 0$$

We multiply both sides with $(1-\rho^2)^2$ and we get

$$\begin{aligned}\frac{d\ell(\rho)}{d\rho} &= n\rho(1-\rho^2) - \rho\sum_{i=1}^{n}\left(\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right) \\ &\quad + (1-\rho^2)\sum_{i=1}^{n}\left(\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}\right) \\ &= n\rho - n\rho^3 - \rho\sum_{i=1}^{n}\left(\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y^2}\right) \\ &\quad + \rho^2\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y} + \sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y} = 0.\end{aligned}$$

Then we multiply both sides with $\frac{-1}{n}$ and we get

$$\begin{aligned}\rho^3 - \rho^2\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y} + \rho\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)}{\sigma_y^2}\right] - 1\right) \\ - \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y} = 0.\end{aligned} \tag{6.2}$$

To make the equation simpler we define $\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{n\sigma_x^2}\sqrt{n\sigma_y^2}}=\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}=$ k and $\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(x_i-\mu_x)^2}{\sigma_x^2}+\frac{(y_i-\mu_y)^2}{\sigma_y^2}\right]=t$. This simplifies (6.2) into

$$\begin{aligned}\rho^3-\rho^2k+\rho(t-1)-k&=0\\ \rho(\rho^2-\rho k+(t-1))&=k.\end{aligned}\tag{6.3}$$

We want to check if $\rho=k$ is a solution of (6.3), and we see that k is a solution in the equation if $t=2$. If we now let $\mu_x=\bar{x}$, $\mu_y=\bar{y}$, $\sigma_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$, which are the well-known estimates when estimating the complete parameter vector by maximal likelihood, and $\sigma_y=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$ we get

$$t=\frac{1}{n}\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}+\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)=\frac{1}{n}(n+n)=2.$$

This means that the maximum likelihood estimator of ρ is $\sum_{i=1}^{n}\frac{(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{n\sigma_x^2}\sqrt{n\sigma_y^2}}$ if the other parameters μ_x, μ_y, σ_x and σ_y are equal to the MLE of the marginal distributions. Thus, the MLE and the IFME methods give the same estimates of ρ in the binormal case.

Chapter 7

Transformations

Another important property of the copula is the fact that it is invariant under increasing transformations. To explain what we mean by this we have to introduce some new notation. If we have two random variables X and Y we will use the notation C_{XY}, meaning the copula that couples the distribution functions of X and Y together. We also remind the reader that a continuous random variable means that the distribution function to the random variable is continuous. In the next theorem we will use the fact that if an increasing function α with domain that contains the range of a random variable X, then the distribution function of $\alpha(X)$ is continuous [2, p. 25]. With this fact we introduce the next theorem and its proof [2, p. 25].

Theorem 12. *Let X and Y be continuous random variables with copula C_{XY}. If α and β are increasing functions on the range of X and Y, respectively, then $C_{\alpha(X)\beta(Y)} = C_{XY}$*

Proof. Let F_1, G_1, F_2, and G_2 denote the distribution functions of X, Y, $\alpha(X)$, and $\beta(Y)$, respectively. Since α and β are increasing functions $F_2(x) = P(\alpha(X) \leq x) = P(X \leq \alpha^{-1}(x)) = F_1(\alpha^{-1}(x))$. With a similar procedure we get that $G_2(y) =$

$G_1(\beta^{-1}(y))$. This gives that for any x and y in the range of X and Y, respectively,

$$\begin{aligned}
C_{\alpha(X)\beta(Y)}(F_2(x), G_2(y)) &= P(\alpha(X) \leq x, \beta(Y) \leq y) \\
&= P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) \\
&= C_{XY}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y)) \\
&= C_{XY}(F_2(x), G_2(y)).
\end{aligned}$$

Since X and Y are continuous random variables, the range of F_2 and G_2 is $\mathbb{I}$, this gives that $C_{\alpha(X)\beta(Y)} = C_{XY}$ on $\mathbb{I}^2$. □

In the theorem above we were only considering the case where α and β are increasing functions. We have similar cases when α and/or β are decreasing functions [2, p. 26]:

Theorem 13. *Let X and Y be continuous random variables with copula C_{XY}. Let α and β be strictly monotone on the range of X and the range of Y, respectively.*

(1) *If α is increasing and β is decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v).$$

(2) *If α is decreasing and β is increasing, then*

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v).$$

(3) *If α and β are both decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v).$$

Chapter 8

Simulation

Assume that we want to simulate a continuous random variable X having distribution function F. One way to do this is to use the inverse of the distribution function, F^{-1}. The algorithm is as follows [2, p. 41]:

1. Simulate U from $U[0, 1]$.
2. Set $X = F^{-1}(U)$.

Now we prove that X really is a simulation from the distribution F. This statement is equivalent to proving that $P(X \leq x)$ is equal to the distribution function $F(x)$.

Proof. Note that for U distributed $U[0, 1]$, $P(U \leq u) = u$ for $u \in [0, 1]$. Then $P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$, as required. □

The same idea can be used to simulate from a copula, but in this case we have to introduce the conditional distribution of a copula. We define $c_u(v) = P(V \leq v \mid U = u)$ where U and V are uniform random variables $U[0, 1]$. The joint pdf of (U, V) is $\frac{\partial^2 C(u,v)}{\partial u \partial v}$, which is also the conditional pdf of V given $U = u$ since the pdf of U is uniform and has pdf equal to 1. If we integrate $\frac{\partial^2 C(u,v)}{\partial u \partial v}$ from $-\infty$ to v with respect to the second argument, which is the same as taking the partial derivative of $C(u,v)$ with respect to, u we obtain $c_u(v) = \frac{\partial C(u,v)}{\partial u}$. The idea is now to simulate U, then V is conditioned on U using the previous algorithm [2, p. 41]:

1. Simulate (U,T) from $U[0,1]$ random variables.
2. Set $V = c_u^{-1}(T)$.
3. Now (U,V) is a simulation from the copula C.

We now show an example of the calculation of $c_u(v)$ and $c_u^{-1}(t)$.

Example 6. From Lemma 5 we know that $\Pi(u,v) = uv$ is a copula. Then $c_u(v) = \frac{\partial C(u,v)}{\partial u} = v$, so the inverse is $c_u^{-1}(v) = v$. Hence, in this case, to simulate from Π, we simply simulate u and v independently from $U[0,1]$

This makes sense since we know from Corollary 10 that U and V has copula Π if and only if they are independent.

We have now shown how to simulate X from a distribution F, and (U,V) from a copula C. We combine these two methods to show how to simulate (X,Y) from the bivariate distribution function H.

1. Simulate (U,V) from C.
2. Set $X = F^{-1}(U)$ and $Y = G^{-1}(V)$.
3. Now (X,Y) is a simulation from the bivariate distribution H.

Chapter 9

Conclusion

In this thesis we have seen that there is a link between the bivariate distribution, its marginal distributions and the dependence structure between the marginals. Said in another way, the copula represents the dependence structure between the marginals in a multivariate distribution. Sklar's theorem describes this connection very nicely. In applications it is not hard to imagine that if you want to make a multivariate distribution model, it can be useful to split the problem into two parts, namely choosing the marginal distribution first and the dependence structure afterwards. The copula makes this strategy easy.

We estimated the parameter ρ in the Gaussian copula with all the three different estimation methods MLE, IMFE, and non parametric. The MLE and IMFE method gave the same estimation of ρ in the binormal case, and we calculated the likelihood function of ρ to show why this is the case. There is a bigger standard deviation in the nonparametric method compared to the parametric method, however the standard deviation becomes similar when N gets larger. This makes sense since the model assumptions is not as important for the estimation when we have a lot of data.

We have shown how the copula is invariant under increasing transformations. We have also demonstrated how you can simulate from a copula by using a similar method to the inverse transform sampling method, and how you can combine this method with the same inverse transform sampling method to be able to simulate from a bivariate distribution.

Appendix

```
#code for Figure 2.1

library(plot3D)

x=seq(0,1,0.1)#x vector
y=seq(0,1,0.1)#y vector
#making a function f(x,y)
zvector=function(X,Y){
  n=length(x)
  z=matrix(0,n,n) #making a grid with all the x and y values
  #making the z values for the graph
  for(i in 1:n){
    for(j in 1:n){
      if(x[i]+y[j]<=4/3 & x[i]+y[j]>=2/3){
        z[i,j]= min(x[i],y[j],1/3,x[i]+y[j]-2/3)
      }
      else{
        z[i,j]=max(x[i]+y[j]-1,0)
      }
    }
  }
  return(z)
}
z=round(zvector(x,y),2)
```

```
#The plot
persp(x,y,z,theta=25,phi = 30,xlab = "u",ylab = "v",zlab = "Q(u,v)")


#MLE,IMFE estimation and the true MLE value
library(MASS)#package for mvrnorm
library(copula) #package for copula functions
library(pracma)#package for the modulo function (not very important)

#Parametric estimation MLE
#function that that calculates B number of estimaes of rho and the other estimates
#where it is each estimate is based on N samples
MLEestimating=function(B,N){
  datamatrix<-matrix(data = NA,ncol = 10,nrow = 5)
  for (j in 1:5) {
    rho=-0.2+0.2*j #rho=c(0,0.2,0.4,0.6,0.8)

    set.seed(123)
    rhoest<-rep(0,B) #0 vector
    mu1est<-rep(0,B)
    mu2est<-rep(0,B)
    si1est<-rep(0,B)
    si2est<-rep(0,B)

    for (i in 1:B) {
      mu1=0
      mu2=0
      s1=1
      s2=1

      mu <- c(mu1,mu2) #Mean
      sigma <- matrix(c(s1^2, s1*s2*rho, s1*s2*rho, s2^2),2) #Covariance matrix
      bvs <- mvrnorm(N, mu = mu, Sigma = sigma ) #bivarate samples

      #The liklyhood function
      MLErho2<- function(argu){
        rho<-argu[1]; MU1<-argu[2]; MU2<-argu[3]; SI1<-argu[4]; SI2<-argu[5]

        u<-matrix(c(pnorm(bvs[,1],mean=MU1,sd=SI1),
```

```
                  pnorm(bvs[,2],mean=MU2,sd=SI2)),ncol = 2 , byrow = F)
      qnormu<-qnorm(u)
      MLEvalue<- -N*log(2*pi)-.5*N*log(1-rho^2)-
        sum(qnormu[,1]^2-2*rho*qnormu[,1]*qnormu[,2]+qnormu[,2]^2)/(2*(1-rho^2))-
        sum(log(dnorm(qnormu[,1])))-sum(log(dnorm(qnormu[,2])))+
    sum(log(dnorm(bvs[,1],mean=MU1,sd=SI1))+log(dnorm(bvs[,2],mean=MU2,sd=SI2)))
      return(MLEvalue)
    }
    #Finding the MLE of rho and all the other estimates
    opt<-optim((c(0,0,0,1,1)),MLErho2, method = c("BFGS"),
               control=c(fnscale=-1,reltol=1e-10))
    #putting the estimates in a vector
    rhoest[i]<-opt$par[1]
    mu1est[i]<-opt$par[2]
    mu2est[i]<-opt$par[3]
    si1est[i]<-opt$par[4]
    si2est[i]<-opt$par[5]
  }
  #putting the mean and sd of the estimates in another matrix
  datamatrix[j,]<-c(mean(rhoest),sd(rhoest),mean(mu1est),sd(mu1est),mean(mu2est),
                    sd(mu2est), mean(si1est),sd(si1est),mean(si2est),sd(si2est))
}
#putting col and row names on the matrix
rownames(datamatrix)<-c("0","0.2","0.4","0.6","0.8")
colnames(datamatrix)<-c("meanrho","sdrho","meanmu1","sdmu1","meanmu2","sdmu2",
                        "meansi1","sdsi1","meansi2","sdsi2")
return(t(datamatrix)) #transposing the matrix
}

#Parametric estimation IFME
IFMEestimation=function(B,N){
  datamatrix<-matrix(data = NA,ncol = 10,nrow = 5)
  for (j in 1:5) {
    rho=-0.2+0.2*j #rho=c(0,0.2,0.4,0.6,0.8)

    set.seed(123)
    rhoest<-rep(0,B) #0 vector
    mu1est<-rep(0,B)
```

```
mu2est<-rep(0,B)
si1est<-rep(0,B)
si2est<-rep(0,B)

for (i in 1:B) {
  mu1=0
  mu2=0
  s1=1
  s2=1
  mu <- c(mu1,mu2) #Mean
  sigma <- matrix(c(s1^2, s1*s2*rho, s1*s2*rho, s2^2),2) #Covariance matrix
  bvs <- mvrnorm(N, mu = mu, Sigma = sigma ) #bivarate samples
  #calculating the famous MLE of the normal distribution
  #which is our estimate of the mean and sd of mu1, mu2 ,si1 and si2
  MLEmean<-c(mean(bvs[,1]),mean(bvs[,2]))
  MLEsd<-sqrt((N-1)/N)*c(sd(bvs[,1]),sd(bvs[,2]))

  #using the famous MLE of normal distribution to calculate u
  u<-matrix(c(pnorm(bvs[,1],mean=MLEmean[1],sd=MLEsd[1]),
              pnorm(bvs[,2],mean=MLEmean[2],sd=MLEsd[2])),ncol = 2 , byrow = F)

  qnormu<-qnorm(u)

  #function that that calculates IFME
  MLErho<- function(rho){
    MLEvalue<- -N*log(2*pi)-.5*N*log(1-rho^2)-
      sum(qnormu[,1]^2+qnormu[,2]^2-2*rho*qnormu[,1]*qnormu[,2])/(2*(1-rho^2))-
      sum(log(dnorm(qnormu[,1]))+log(dnorm(qnormu[,2])))
    return(MLEvalue)
  }
  #estimating rho
  opt<-optimize(MLErho,interval = c(-1,1),maximum = T,tol=1e-10)
  #putting the estimates in a vector
  rhoest[i]<-opt$maximum
  mu1est[i]<-MLEmean[1]
  mu2est[i]<-MLEmean[2]
  si1est[i]<-MLEsd[1]
  si2est[i]<-MLEsd[2]
```

```
    }
    #putting the mean and sd of the estimates in another matrix
    datamatrix[j,]<-c(mean(rhoest),sd(rhoest),mean(mu1est),sd(mu1est),mean(mu2est),
                      sd(mu2est),mean(si1est),sd(si1est),mean(si2est),sd(si2est))
  }
  #putting col and row names on the matrix
  rownames(datamatrix)<-c("0","0.2","0.4","0.6","0.8")
  colnames(datamatrix)<-c("meanrho","sdrho","meanmu1","sdmu1","meanmu2","sdmu2",
                          "meansi1","sdsi1","meansi2","sdsi2")
  return(t(datamatrix)) #transposing the matrix for best visibility
}

#True MLE
MLEfasit=function(B,N){
  datamatrix<-matrix(data = NA,ncol = 10,nrow = 5)
  for (j in 1:5) {
    rho=-0.2+0.2*j #rho=c(0,0.2,0.4,0.6,0.8)

    set.seed(123)
    rhoest<-c(1:B)*0 #0 vector
    mu1est<-c(1:B)*0
    mu2est<-c(1:B)*0
    si1est<-c(1:B)*0
    si2est<-c(1:B)*0

    for(i in 1:B){
      mu1=0
      mu2=0
      s1=1
      s2=1
      mu <- c(mu1,mu2) #Mean
      sigma <- matrix(c(s1^2, s1*s2*rho, s1*s2*rho, s2^2),2) #Covariance matrix
      bvs <- mvrnorm(N, mu = mu, Sigma = sigma ) #bivarate samples

      #Putting the MLEfasit estimates in a vector
      rhoest[i]<-cor(bvs[,1],bvs[,2])
      mu1est[i]<-mean(bvs[,1])
      mu2est[i]<-mean(bvs[,2])
```

```
      si1est[i]<-sqrt((N-1)/N)*sd(bvs[,1])
      si2est[i]<-sqrt((N-1)/N)*sd(bvs[,2])
    }
    #putting the mean and sd of the estimates in another matrix
    datamatrix[j,]<-c(mean(rhoest),sd(rhoest),mean(mu1est),sd(mu1est),mean(mu2est),
                      sd(mu2est),mean(si1est),sd(si1est),mean(si2est),sd(si2est))
  }
  #putting col and row names on the matrix
  rownames(datamatrix)<-c("0","0.2","0.4","0.6","0.8")
  colnames(datamatrix)<-c("meanrho","sdrho","meanmu1","sdmu1","meanmu2","sdmu2",
                          "meansi1","sdsi1", "meansi2","sdsi2")
  return(t(datamatrix)) #transposing the matrix
}

#Estimating with the MLE method,
#1000 datasets with N=10 result rounded to 4 digits
round(MLEestimating(100,1000),4)
#Estimating with the IFME method,
#1000 datasets with N=10 result rounded to 4 digits
round(IFMEestimation(100,1000),4)
#True MLE of 100 datasets with N=1000 result rounded to 4 digits
round(MLEfasit(100,1000),4)




#Noneparametric estimation MPLE
library(MASS)#package for mvrnorm
library(copula)#package for copula functions
library(pracma)#package for the modulo function (not very important)

MPLEestimation=function(B,N){
  datamatrix<-matrix(data = NA,ncol = 2,nrow = 5)
  for (j in 1:5) {
    rho=-0.2+0.2*j #rho=c(0,0.2,0.4,0.6,0.8)

    set.seed(123)
    rhoest<-c(1:B)*0 #0 vector
```

```
    for (i in 1:B) {
      mu1=0
      mu2=0
      s1=1
      s2=1
      mu <- c(mu1,mu2) #Mean
      sigma <- matrix(c(s1^2, s1*s2*rho, s1*s2*rho, s2^2),2) #Covariance matrix
      bvs <- mvrnorm(N, mu = mu, Sigma = sigma ) #bivarate samples
      #nonparametric u
      u<- matrix(c(rank(c(bvs[,1])),rank(c(bvs[,2]))),ncol=2,byrow = F)/(N+1)

      qnormu<-qnorm(u)

      #function that calculates the IFME depending on rho
      MLErho<- function(rho){
        #MLEvalue<--N*log(sqrt(1-rho^2))-sum(bvs^2)/(2*(1-rho^2))+
        #  2*rho*sum(bvs[,1]*bvs[,2])/(2*(1-rho^2))-sum(dnorm(bvs,log = T))
        MLEvalue<- -N*log(2*pi)-.5*N*log(1-rho^2)-
          sum(qnormu[,1]^2+qnormu[,2]^2-2*rho*qnormu[,1]*qnormu[,2])/(2*(1-rho^2))-
          sum(log(dnorm(qnormu[,1]))+log(dnorm(qnormu[,2])))
        return(MLEvalue)
      }
      #calculating the optimal rho
      opt<-optimize(MLErho,interval = c(-1,1),maximum = T,tol=1e-10)
      rhoest[i]<-opt$maximum #putting the estimates in a vector
    }
    #putting the mean and sd of the rho estimates in a matrix
    datamatrix[j,]<-c(mean(rhoest),sd(rhoest))
  }
  #putting col and row names on the matrix
  rownames(datamatrix)<-c("0","0.2","0.4","0.6","0.8")
  colnames(datamatrix)<-c("meanrho","sdrho")
  return(t(datamatrix)) #transposing the matrix
}
#Estimating with the MPLE method,
#100 datasets with N=1000 result rounded to 4 digits
round(MPLEestimation(100,1000),4)
```

```r
#different way of writing the code
cc<-normalCopula(dim = 2) #Gaussian copula with dimension 2 with no rho specified
#fits the best gaussian copula with nonparametrick marginals
MPLE<-fitCopula(normalCopula(dim = 2),data = pobs(bvs),method = "mpl")
summary(MPLE)
```

Bibliography

[1] Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T. 2000. Copulas for Finance - A Reading Guide and Some Applications. available at SSRN 1032533.

[2] Nelsen, R. 2006. An Introduction to Copulas

[3] Hofert, M., Kojadinovic, I., Mächler, M., Yan, J. 2018. Elements of Copula modeling with r

[4] Li, David X. 2000. A Copula Function Approach *The Journal of Fixed Income Spring*, 9 (4) 43-54

[5] Watts, S. 2016. A Recipe for Disaster or Cooking the Books?

[6] Salmon, F., 2012. The formula that killed wall street. *Significance* 9, 16-20.

NTNU

Norwegian University of
Science and Technology