

Bachelor Thesis



**Czech
Technical
University
in Prague**

F3

**Faculty of Electrical Engineering
Department of Cybernetics**

Modelling Dependence of Random Variables by Copulas

**Principles of random variables modelling dependence
by copulas**

Daria Kuznetsova

Supervisor: prof. Ing. Mirko Navara, DrSc.

Study program: Open Informatics

Specialisation: Artificial Intelligence and Computer Science

May 2024

I. Personal and study details

Student's name: **Kuznetsova Daria** Personal ID number: **510863**
Faculty / Institute: **Faculty of Electrical Engineering**
Department / Institute: **Department of Cybernetics**
Study program: **Open Informatics**
Specialisation: **Artificial Intelligence and Computer Science**

II. Bachelor's thesis details

Bachelor's thesis title in English:

Modelling Dependence of Random Variables by Copulas

Bachelor's thesis title in Czech:

Modelování závislosti náhodných velí in pomocí kopulí

Guidelines:

Study the principles of modelling dependence of random variables by copulas. Put emphasis on copulas describing mixtures of independent distributions. When a random vector has a conditionally independent distribution originated by a mixture of independent components, what does it mean for the corresponding copula? What are the consequences for the joint extreme values distribution?

Discuss the consequences for modelling of financial risks by copulas.

Bibliography / sources:

[1] R. B. Nelsen: An Introduction to Copulas, 2nd edition, volume 139 of Lecture Notes in Statistics. Springer, New York, 2006.

[2] L. Wasserman: All of Statistics. A Concise Course in Statistical Inference. Springer, 2004.

Name and workplace of bachelor's thesis supervisor:

prof. Ing. Mirko Navara, DrSc. Machine Learning FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **06.02.2024** Deadline for bachelor thesis submission: **24.05.2024**

Assignment valid until: **21.09.2025**

prof. Ing. Mirko Navara, DrSc.
Supervisor's signature

prof. Dr. Ing. Jan Kybic
Head of department's signature

prof. Mgr. Petr Páta, Ph.D.
Dean's signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

Date of assignment receipt

Student's signature

Acknowledgements

I would like to thank my family, who always supports me in all my endeavors and stands by my side.

I would also like to thank my supervisor, Mirko Navara, for his wise advice and ideas for the development of this work.

Declaration

I declare that I prepared the presented work independently and that I listed all the information sources used in accordance with the Methodical Instruction on the observance of ethical principles in the preparation of university final theses.

I declare that I used AI tools for exploring available copula-related functions in public MATLAB libraries. The code used in this work in MATLAB was written independently by me. Also it was used to assist with translating specific sentences from my native language to English and to help rephrase my work in a style more suitable for academic work.

Prague, 24 May 2024

Daria Kuznetsova

Abstract

This thesis deals with the principles of modelling dependence of random variables by copulas. The study of copulas and their applications in statistics is a rather modern phenomenon. While mixtures of copulas are well described, little is known about copulas describing conditionally independent distributions. In this thesis, we investigate how the properties of copulas are restricted by the assumption that they describe joint distributions of mixtures of two (and more) independent distributions.

Considering the lack of available literature sources mentioning formulas for the probability density functions of copulas and some found errors in already published sources, we derive densities of copulas by taking the second mixed derivatives of the copulas (assuming they are absolutely continuous).

The densities of various types of copulas near the extremes of their domain and their limits are studied. Conclusions regarding the applicability of such copulas for financial risk modeling are discussed and presented. Also there are outlined potential directions of possible future research and development based on the obtained results.

Keywords: Copula, random variable, probability distribution, quantile function, singular value decomposition, copula density

Supervisor: prof. Ing. Mirko Navara, DrSc.

Abstrakt

Tato práce se zabývá principy modelování závislosti mezi náhodnými veličinami pomocí kopulí. Studium kopulí a jejich aplikace ve statistice je poměrně moderním jevem. Zatímco směsi kopulí jsou dobře popsány, málo se ví o kopulích popisujících podmíněně nezávislá rozdělení. V této práci zkoumáme, jak jsou vlastnosti kopulí omezeny předpokladem, že popisují sdružené rozdělení směsi dvou (a více) nezávislých rozdělení.

S ohledem na nedostatek dostupných literárních zdrojů uvádějících vzorce pro hustoty pravděpodobnosti kopulí a některé nalezené chyby v již publikovaných zdrojích jsou zde odvozeny hustoty kopulí pomocí druhých smíšených derivací kopulí (za předpokladu, že jsou absolutně spojitě).

Jsou studovány hodnoty hustoty různých typů kopulí v blízkosti extrémů jejich definičního oboru a jejich limity. Jsou diskutovány a prezentovány závěry týkající se použitelnosti takových kopulí pro modelování finančního rizika. Také jsou naznačeny potenciální cesty pro další výzkum a vývoj na základě získaných výsledků.

Klíčová slova: Kopule, náhodná veličina, rozdělení pravděpodobnosti, kvantilová funkce, singulární rozklad, hustota kopule

Překlad názvu: Modelování závislosti náhodných veličin pomocí kopulí — Principy modelování závislosti náhodných veličin pomocí kopulí

Contents

1 Introduction	1
1.1 Foundations of copulas	2
1.2 Use of copulas in economics	4
1.3 The 2008 mortgage crisis: applying Gaussian copulas in economics	6
2 Copulas critical points and analysis of mathematical models	9
2.1 Derivation of formulas for bivariate copulas' probability densities	9
2.2 Visualization of copulas	14
2.3 Computation of limits of copula in extremes of domain	17
3 Identifying the number of random variables included in the mixture	27
3.1 Step 1. Sampling random variables from normal distributions	29
3.2 Step 2. Creating independent random vectors	30
3.3 Step 3. Creating a new mixed variable.	30
3.4 Step 4. Corresponding samples from uniform distributions	31
3.5 Step 5. Construct the copula from two uniformly distributed variables	33
3.6 Step 6. Discretizing by SVD and essential rank	34
4 Conclusions	39
Bibliography	41
A MATLAB codes	43

Figures

1.1 3D graphs of copulas [1].	3
1.2 Contour diagrams of copulas [1]. .	3
1.3 Stock portfolios [12].	4
1.4 Bank loans [12].	5
2.1 Gaussian copula density, $\rho = 0$. .	15
2.2 Gaussian copula density, $\rho = 0.25$.	15
2.3 Gaussian copula density, $\rho = 0.75$.	15
2.4 Student- t copula density, $\nu = 3$.	16
2.5 Student- t copula density, $\nu = 60$.	16
2.6 Clayton copula density, $\theta = 1$. . .	17
2.7 Frank copula density, $\theta = 2.5$. . .	17
2.8 Zoomed-in Gaussian, Student- t copula density plots near $(0,0)^1$. . .	20
2.9 Zoomed-in Archimedean copula density plots near $(0,0)^2$	22
2.10 Zoomed-in Gaussian, Student- t copula density plots near $(1,1)^3$. . .	23
2.11 Zoomed-in Archimedean copula density plots near $(1,1)^4$	24
3.1 Histogram of simulated X_1	30
3.2 RVICs Z_1, Z_2	31
3.3 Simulated CDFs.	32
3.4 Quantile functions of simulated random variables.	32
3.5 Copula contours for simulated Z .	33
3.6 Copula density for simulated Z .	34
3.7 Graph of 3 simulated random vectors.	36

Tables

2.1 Bivariate copulas.	9
2.2 Formulas for Gaussian copulas. .	12
2.3 Formulas for Student- t copula. .	12
2.4 Probability densities of copulas.	18
3.1 Table of notations for Chapter 3.	29



Chapter 1

Introduction

In the modern world, research and practical applications in the fields of statistics and probability are increasingly confronted with the growing need for precise and flexible modeling of dependencies among random variables. These dependencies play a pivotal role in forecasting, decision-making, and risk management across various domains, including finance, medicine, climatology, and many others.

However, there is a crucial aspect of dependencies that requires special attention: conditional independence. Understanding this concept and its significance constitutes a fundamental aspect in the field of modeling dependencies among random variables. Conditional independence entails that, under certain conditions, variables can be considered independent despite the presence of common factors or interconnections. This concept becomes particularly relevant when analyzing data and developing models where it is necessary to account for the context and conditions under which the data were collected.

In this thesis, we will study the principles of modelling dependence of random variables by copulas and endeavor to explore and deepen our understanding of conditional independence and its role in modeling dependencies among random variables. We will examine both the theoretical and practical aspects of this concept and demonstrate how it can be applied in real-world scenarios, especially within the context of finance and the prediction of extreme events. This work aims to underscore the importance of conditional independence in modeling the dependencies of random variables and its impact on practical applications across diverse domains.

We will concentrate on the bivariate normal distribution because it is the most important case in practice (applications in quantitative finance include pricing of options and estimation of asset correlations; the impressive list by Balakrishnan and Lai (2009) [10] mentions applications in agriculture, biology, engineering, economics and finance, the environment, genetics, medicine, psychology, quality control, reliability and survival analysis, sociology, physical

sciences and technology) [14]. We will give a view on its limiting values close to extreme points.

Firstly, in Chapter 1 we will study the existing terms, properties and limitations for copulas, and also analyze the relevance and problematics of this issue in economics. Further, in Chapter 2, graphs of various types of copulas will be presented, as well as formulas for finding the limits of the probability density of these types of copulas in the proximity of their extremes will be derived. These limits will also be calculated. In Chapter 3, we tried to answer the question whether mathematical analysis, statistical tools, and optimization can determine the number of random vectors with independent components (RVICs) constituting a copula of the mixture.

1.1 Foundations of copulas

Fundamental principles of copulas have been outlined by R. B. Nelsen [16]. We will delve into the core principles he has introduced, including the definition of copulas, their properties, and their role in capturing dependence structures.

So what are copulas and what is it for? In general terms copulas can be described as functions that join or “couple” multivariate distribution functions to their one-dimensional marginal distribution functions which are uniform on the interval $(0, 1)$. For example, a 2D-copula (two-dimensional copula) is a mathematical concept used to model the dependence between two random variables. It describes the structure of dependence between these two variables, independently of their specific marginal distributions. In other words, a 2D-copula focuses solely on the relationship between the variables, disregarding their individual distributions. Thus, consider arbitrary continuous marginal distribution functions F_1, \dots, F_n . Then, we can define a multivariate distribution function using the copula such that $F(y_1, \dots, y_n) = C(F_1(y_1), \dots, F_n(y_n))$. Here, F is a multivariate distribution function in this equation. In 1959, Sklar [19] showed that any continuous multivariate distribution function F can be written in the form of this equation—using a copula representation. In the subsequent discussion, we will delve into the theoretical aspects using illustrative examples of copulas for two random variables. To make this more tangible, let’s consider two random variables, denoted as X and Y , which are transformed to uniformly distributed random variables $S = F_X(X)$ and $T = F_Y(Y)$.

We form a copula as $C(s, t) = F(F_X^{-1}(s), F_Y^{-1}(t))$, where the data being located in the copula are simulated and predefined. Thus any copula is a function with the following properties [5]:

1. $\text{Dom } C = S_1 \times S_2$, where S_1 and S_2 are subsets of $I = [0, 1]$ containing 0 and 1.

2. For every $s \in S_1$ and every $t \in S_2$:

$$\begin{aligned} C(s, 1) &= s \\ C(1, t) &= t \\ C(s, 0) &= C(0, t) = 0 \end{aligned}$$

3. For any values $a, b, c, d \in I$ such that $a \leq b$ and $c \leq d$:

$$C(b, d) - C(b, c) - C(a, d) + C(a, c) = P(a < x \leq b, c < y \leq d) \geq 0.$$

Combining the above equations yields the copula (Fréchet-Hoeffding) bounds:

$$W(s, t) := \max(s + t - 1, 0) \leq C(s, t) \leq \min(s, t) =: M(s, t).$$

Consequently, C is uniformly continuous within its defined domain.

Based on these definitions and formulas above, let's explore how copula plots look like. The graph of any copula is a continuous surface within the unit cube I^3 whose boundary is the skew quadrilateral with vertices $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 1)$ and $(0, 1, 0)$. Also the useful way to present the plot of a copula is with a contour diagram – the graph of level sets in I^2 given by $C(s, t) = \text{const}$. In Figure 1.1 there are presented the graphs of the copulas M and W , as well as the graph of Π , a portion of the hyperbolic paraboloid $\Pi(s, t) = st$. In Figure 1.2 we present the contour diagrams of these mentioned copulas.

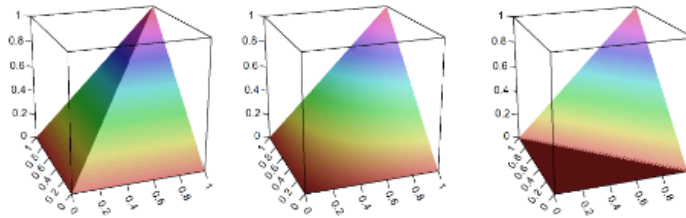


Figure 1.1: 3D graphs of copulas [1].

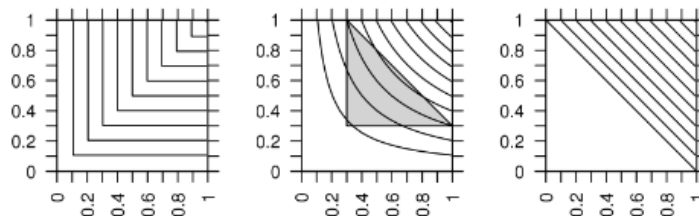


Figure 1.2: Contour diagrams of copulas [1]. (The light grey triangle represents the Fréchet-Hoeffding bounds for $x = y = 0.3$.)

1.2 Use of copulas in economics

Why are copulas useful in many fields of life and why exactly them? The following Figures 1.3 and 1.4 can help us to answer this question.

Dependencies and risk: stock portfolios

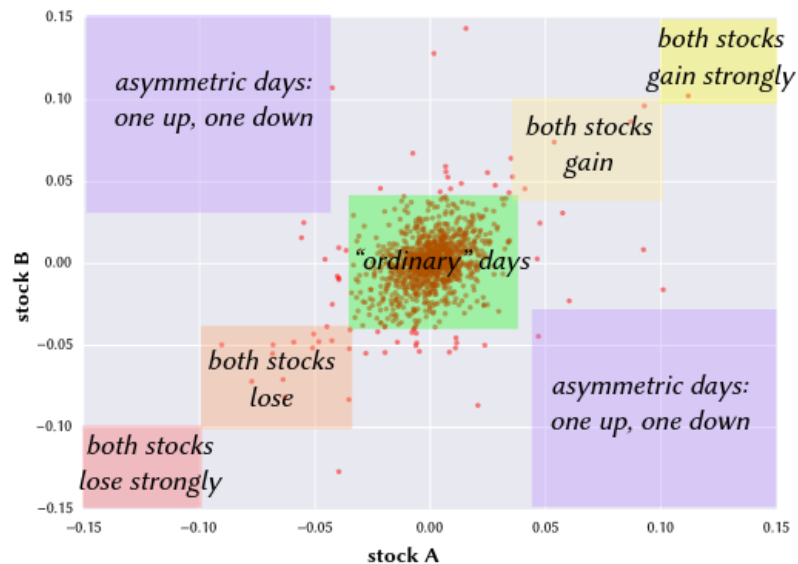


Figure 1.3: Stock portfolios [12].

The graph in Figure 1.3 shows the relationship between two stocks, labeled as stock A and stock B. The axes of the graph represent the changes in prices of these stocks on a particular day.

The graph has different areas, painted in their special colour:

- **Green area** (“ordinary days”): the most common situation where changes in the prices of both stocks are not very strong.
- **Yellow area** (“both stocks gain strongly” and “both stocks gain”): days when both stocks show growth. A lighter shade of yellow corresponds to moderate growth, while a darker shade indicates strong growth.
- **Red area** (“both stocks lose” and “both stocks lose strongly”): days when both stocks lose in value. A darker shade of red indicates larger losses.
- **Purple areas** (“asymmetric days: one up, one down”): days when one stock rises in price and the other falls.

Due to the fact that copulas are mathematical functions used to describe the relationship or dependency between random variables, their application

is particularly useful in finance for modeling and analyzing dependencies between different financial assets, such as stocks. This is related to several advantages of copulas:

- **Flexibility in modeling dependencies:** Copulas allow modeling various types of dependencies, including asymmetric and nonlinear relationships, which are often encountered between financial assets. On the graph in Figure 1.3 above, this is manifested in the form of asymmetric days, when one stock rises and the other falls.

- **Modeling extreme events:** Copulas can effectively model the joint behavior of assets in extreme market conditions, such as crises. This helps in analyzing scenarios where stocks either simultaneously drop sharply or rise sharply (as shown by the red and yellow areas on the graph).

- **Improving risk understanding:** Using copulas allows for a better assessment of portfolio risk, as they provide information about the likelihood of joint price movements of stocks. This is important for determining the likelihood of large losses or gains.

- **Portfolio diversification:** Understanding the dependencies described by copulas helps to identify which stocks in the portfolio behave independently of each other, which is critically important for a diversification strategy.

Using copulas to analyze data presented on such graphs can significantly improve investment management and risk strategies, providing a deeper and more accurate understanding of how stocks may react in different market scenarios.

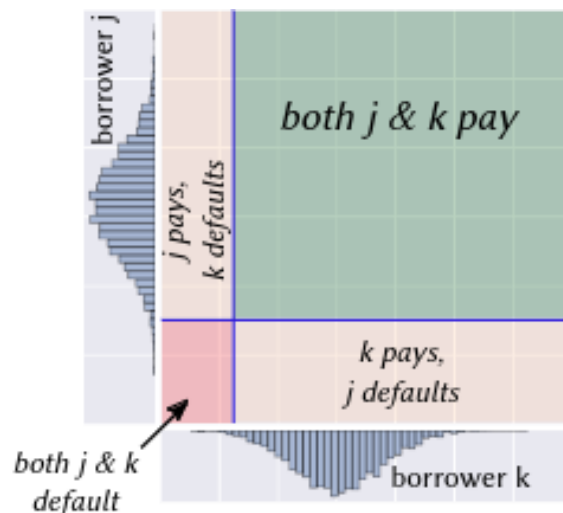


Figure 1.4: Bank loans [12].

The graph in Figure 1.4 is used to show how two borrowers, labeled as borrower j and borrower k, behave in terms of loan repayment. The graph is divided into different sections based on the outcomes of their loan repayments.

Chapter 2

Copulas critical points and analysis of mathematical models

2.1 Derivation of formulas for bivariate copulas' probability densities

Table 2.1 contains formulas defining the values of the main copula types, which will be studied in this thesis.

Copula type	$C(u, v)$	Coefficient boundaries
Gaussian	$\Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$	$-1 \leq \rho \leq 1$
Student- t	$t_{\rho, \nu}(t_\nu^{-1}(u), t_\nu^{-1}(v))$	$-1 \leq \rho \leq 1, \nu \in \mathbb{N}$
Clayton	$(\max(u^{-\theta} + v^{-\theta} - 1, 0))^{-\frac{1}{\theta}}$	$-1 \leq \theta < \infty, \theta \neq 0$
Frank	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$	$-\infty < \theta < \infty, \theta \neq 0$

Table 2.1: Bivariate copulas.([21] and [16])

Here:

- Φ^{-1} is the quantile function of the standard normal distribution;
- Φ_ρ is the joint cumulative distribution function of the two-dimensional normal distribution with zero mean vector and correlation matrix

$$R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix};$$

- t_ν^{-1} is the quantile function of Student- t distribution with ν degrees of freedom;

- $t_{\rho,\nu}$ is the joint cumulative distribution function of the two-dimensional Student- t distribution with correlation coefficient ρ . The covariance matrix for multivariate t -distribution differs from the Gaussian correlation matrix R [6], especially:

$$\text{Cov}(\mathbf{X}) = \frac{\nu}{\nu - 2} R$$

for $\nu > 2$.

Nelsen [16] has conditions of strictness for Clayton copula $\theta \geq 0$, where for $\theta = 0$: $C_0 = \Pi(u, v) = uv$ is the product copula.

According to [7], there is another notation for Frank copulas, namely: the family $(T_\lambda^{\mathbf{F}})_{\lambda \in [0, \infty]}$ of *Frank t -norms* (copulas) is given by:

$$T_\lambda^{\mathbf{F}}(x, y) = \begin{cases} \min(x, y) & \text{if } \lambda = 0, \\ x \cdot y & \text{if } \lambda = 1, \\ \max(x + y - 1, 0) & \text{if } \lambda = \infty, \\ \log_\lambda \left(1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1} \right) & \text{otherwise.} \end{cases}$$

The parameter λ is defined here as $\lambda = e^{-\theta}$.

The Gaussian copula is dependent solely on the correlation coefficient ρ between random variables.

The Student- t copula, like the Gaussian copula, serves as a tool to model dependency between random variables. However, it introduces an additional parameter known as the number of degrees of freedom. So in the context of the Student- t copula:

- **Correlation Coefficient:** This parameter, similarly to the Gaussian copula, determines the strength and direction of the linear relationship between random variables. It quantifies the degree to which the variables move together.
- **Degrees of Freedom:** The number of degrees of freedom (often denoted as ν) determines the thickness of the tails of the distribution. Higher values of degrees of freedom result in thinner tails, approaching a Gaussian distribution, while lower values lead to fatter tails, implying greater probability of extreme events.

Some Archimedean copula families, especially Clayton and Frank copulas, are characterized by a single parameter (θ). They have different boundaries of this parameter (see Table 2.1).

Considering the lack of available literature sources mentioning formulas for the probability densities of copulas, and some found errors in them (as for example in [9]), it was decided to derive such formulas.

In this section, we will derive some formulas for computing the density of key copula types (Gaussian, Student- t , Frank). Since we need to find the formula for the probability density of the copula, and for two variables it will be a volumetric figure in space, we will need to find the second mixed derivative of the cumulative distribution function (assuming the copula is absolutely continuous).

In this chapter, the following notations and formulas will be used:

X, Y	Random variables, which will be evaluated at points $x, y \in (-\infty, +\infty)$
F_X, F_Y	Distribution functions of X, Y
$S = F_X(X),$ $T = F_Y(Y)$	Transformed random variables, which will be evaluated at points $s = F_X(x), t = F_Y(y) \in (0, 1)$
$f_X(x) = F'_X(x) = \frac{\partial}{\partial x} \overbrace{F_X(x)}^s = \frac{\partial s}{\partial x}$	Density of random variable X
$f_Y(y) = F'_Y(y) = \frac{\partial}{\partial y} \overbrace{F_Y(y)}^t = \frac{\partial t}{\partial y}$	Density of random variable Y
$F_{X,Y}$	Joint distribution function of (X, Y)
$f_{X,Y}(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F_{X,Y}(x, y)$	Joint density of a random vector (X, Y)
$C(s, t) = F_{X,Y}(F_X^{-1}(s), F_Y^{-1}(t))$	The copula of (X, Y) (the joint distribution function of (S, T))
$c(s, t) = \frac{\partial}{\partial s} \frac{\partial}{\partial t} C(s, t)$	The copula density of (X, Y)

Using the notations above, we can obtain the copula density generally as:

$$\begin{aligned}
 c(s, t) &= \frac{\partial}{\partial s} \frac{\partial}{\partial t} C(s, t) = \frac{\partial}{\partial s} \frac{\partial}{\partial t} F_{X,Y}(F_X^{-1}(s), F_Y^{-1}(t)) \\
 &= \frac{\partial}{\partial t} \left(\frac{\partial}{\partial s} F_{X,Y}(F_X^{-1}(s), F_Y^{-1}(t)) \cdot \underbrace{\frac{\partial}{\partial s} \overbrace{F_X^{-1}(s)}^x}_{\frac{\partial x}{\partial s}} \right) \\
 &= \left(\frac{\partial}{\partial s} \frac{\partial}{\partial t} F_{X,Y}(F_X^{-1}(s), F_Y^{-1}(t)) \right) \cdot \underbrace{\frac{\partial x}{\partial s}}_{\frac{1}{f_X(x)}} \cdot \underbrace{\frac{\partial y}{\partial t}}_{\frac{1}{f_Y(y)}} \\
 &= \frac{1}{f_X(F_X^{-1}(s))} \cdot \frac{1}{f_Y(F_Y^{-1}(t))} \cdot f_{X,Y}(F_X^{-1}(s), F_Y^{-1}(t)) \quad (2.1)
 \end{aligned}$$

Abbreviation	Formula
$f_X(x)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$
$f_{X,Y}(x, y)$	$\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[x^2-2\rho xy+y^2]}$

Table 2.2: Formulas for Gaussian copulas ([15] and [9]).

a) Probability density of Gaussian copulas

Substituting known formulas, listed in Table 2.2, into (2.1), we obtain an expression for the probability density c_ρ of the Gaussian bivariate copula:

$$\begin{aligned}
 c_\rho(s, t) &= c_\rho(F_X(x), F_Y(y)) \\
 &= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}[\overbrace{F_X^{-1}(s)^2}^{x^2} - 2\rho \overbrace{F_X^{-1}(s)}^x \overbrace{F_Y^{-1}(t)}^y + \overbrace{F_Y^{-1}(t)^2}^{y^2}]\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{F_X^{-1}(s)^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{F_Y^{-1}(t)^2}{2}\right)} \\
 &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{1}{1-\rho^2} \cdot x^2 - \frac{2\rho}{1-\rho^2} \cdot xy + \frac{1}{1-\rho^2} \cdot y^2 - x^2 - y^2\right)\right) \\
 &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho}{2(1-\rho^2)}(\rho x^2 - 2xy + \rho y^2)\right)
 \end{aligned}$$

The resulting formula corresponds to the expression for the normal copula density in [21].

b) Probability density of Student- t copulas

Let's conduct similar calculations for the density of the Student- t copula for two variables. The formulas to be used are listed in Table 2.3).

Abbreviation	Formula
$f_X(x)$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
$f_{X,Y}(x, y)$	$\frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi\nu\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}$

Table 2.3: Formulas for Student- t copula ([15] and [13]).

Here:

- ν is the number of degrees of freedom,

■ Γ is the Gamma function.

However, in [9], comprehensive presentation describing copulas of Institute of Mathematics and Statistics, University of São Paulo, Brazil, there was found a mistake in the joint density formula:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)^2}\right)^{-\frac{\nu+2}{2}}$$

In this formula, the first multiplier in the denominator is missing the coefficient with the Gamma function, which depends on the degrees of freedom, a characteristic feature of the t -distribution; instead, the coefficient $2 \cdot \pi$ is indicated there. Also, in the denominator of the second multiplier, $(1-\rho^2)$ is squared, whereas in other studied sources of Student- t copula ([13], [25] and [3]), this squaring is not present. For further calculations, the formula from source [13] will be used because it is mentioned there explicitly.

The corresponding Student- t copula density may be calculated by (2.1), similarly to the density of the Gaussian copula:

$$\begin{aligned} c_{\rho,\nu}(s, t) = c_{\rho,\nu}(F_X(x), F_Y(y)) &= \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)^2}\right)^{-\frac{\nu+2}{2}} \\ &\quad \frac{\Gamma(\frac{\nu+1}{2})^2}{\pi\nu\Gamma(\frac{\nu}{2})^2} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ &= \frac{1}{\sqrt{1-\rho^2}} \cdot \frac{\Gamma(\frac{\nu}{2})\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{1+\nu}{2})^2} \cdot \frac{\left(\left(1 + \frac{x^2}{\nu}\right)\left(1 + \frac{y^2}{\nu}\right)\right)^{\frac{\nu+1}{2}}}{\left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)^2}\right)^{\frac{\nu+2}{2}}} \end{aligned}$$

The resulting formula corresponds to the bivariate Student- t copula in [25].

c) Probability density of Frank copulas

The formula for the density of the Frank copula is specified in many sources; however, we will verify the accuracy of the result through calculations.

Similar to the Gaussian and Student- t copulas, the probability density of a copula can be obtained by taking the mixed second derivative of the copula. Firstly, we take the derivative with respect to the first variable, considering the second as a constant. Then we do the same, but with the second variable. The calculations are presented below:

$$\begin{aligned} \frac{\partial C(u, v)}{\partial u} &= -\frac{1}{\theta} \cdot \frac{(e^{-\theta} - 1)}{(e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1} \cdot \frac{(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \cdot e^{-\theta u} \cdot (-\theta) \\ &= \frac{(e^{-\theta v} - 1) \cdot e^{-\theta u}}{(e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1} \end{aligned}$$

Let's denote the long expression $k = (e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1$, then we obtain:

$$\begin{aligned} \frac{\partial^2 C(u, v)}{\partial u \partial v} &= \frac{e^{-\theta u} e^{-\theta v} \cdot (-\theta) \cdot k - (e^{-\theta u} - 1) e^{-\theta v} \cdot (-\theta) (e^{-\theta v} - 1) e^{-\theta u}}{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)^2} \\ &= \frac{(-\theta) e^{-\theta u} e^{-\theta v} (e^{-\theta} - 1)}{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)^2} = \frac{(-\theta) e^{-\theta(u+v)} (e^{-\theta} - 1)}{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)^2} \end{aligned}$$

When specific numbers were substituted into the resulting formula, the results matched the numbers on the Frank copula graphs in MATLAB. In [22], the following formula was utilized, but with slightly different notation:

$$c(u, v) = \frac{-\theta \cdot g(1)(1 + g(u + v))}{(g(u)g(v) + g(1))^2},$$

where the function g is defined by: $g(x) = e^{-\theta x} - 1$.

2.2 Visualization of copulas

In this section, we delve into the visualization of copulas, essential tools for modeling dependency structures. We start by exploring various types of copulas through graphical representations, emphasizing an examination of how copulas behave near their extremal values. Additionally, we provide an overview of utilizing programming software, particularly MATLAB libraries and functions [17] – `copulacdf` and `copulapdf` – for constructing copula plots. Through visualizations, we aim to elucidate the distinctive characteristics and behavior of different copula families, providing insights into their applications in diverse fields such as finance, statistics, and risk management.

To accurately represent the values on the axes representing random variables, a small margin was truncated from both sides of 0 and 1 (as it is hard to compute limits at the boundary points for graphs). The parameter was set to a value of $a = 0.0001$. Thus, the interval for depiction on all subsequent plots in this work along the axes of random variables will be $(a, 1 - a)$.

Let's start by examining Gaussian copulas for independent random variables, dependent random variables with a correlation coefficient of 0.25, and fully dependent random variables (Figures 2.1–2.3). Each figure represents a copula over the entire domain and on the interval close to the origin of coordinates $(a, 0.1)$.

The Gaussian copula, dependent solely on the correlation coefficient, visually portrays the relationship between random variables in a multivariate normal distribution.

In the case of zero correlation (Figure 2.1), the copula density is a flat surface. This occurs because there is no linear relationship between the random variables. Each variable behaves independently of the other, resulting in a copula that is spread uniformly across the entire space.

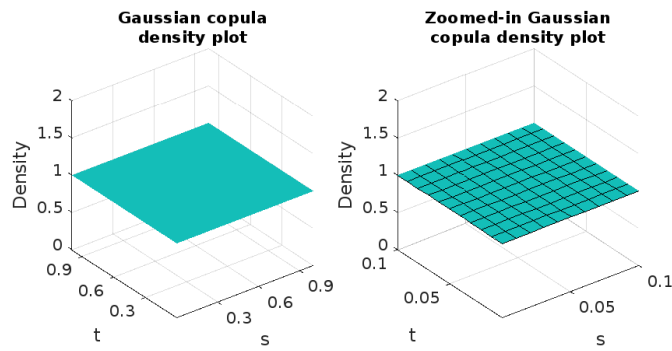


Figure 2.1: Gaussian copula density, $\rho = 0$.

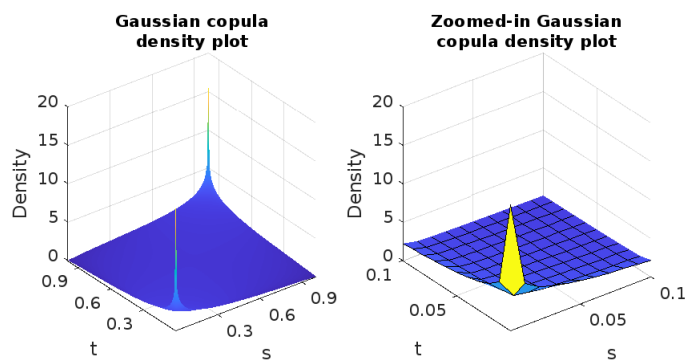


Figure 2.2: Gaussian copula density, $\rho = 0.25$.

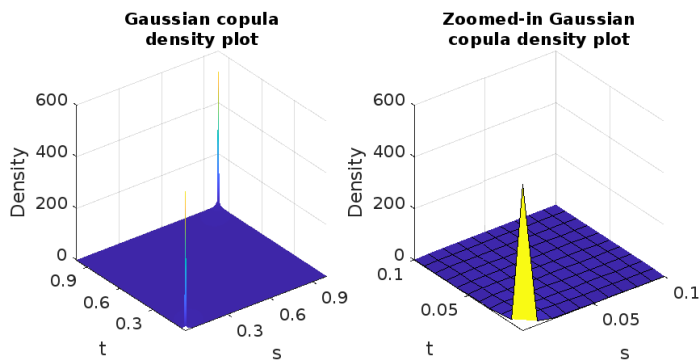


Figure 2.3: Gaussian copula density, $\rho = 0.75$.

As the correlation coefficient ρ increases to 0.25, the copula density begins to elongate along the diagonal axis. This elongation indicates a weak positive linear relationship between the variables. Although the dependence is present, it is not strong enough to significantly alter the shape of the copula.

With a correlation coefficient of 0.75, the copula density exhibits pronounced elongation along the diagonal axis. This elongation signifies a strong positive linear relationship between the variables.

Due to the fact that the values of copula densities tended to $+\infty$, some graphs were plotted with a constraint on the z -axis in the interval $[0, 30]$.

Let's now examine how the graphs of the Student- t copula look like.

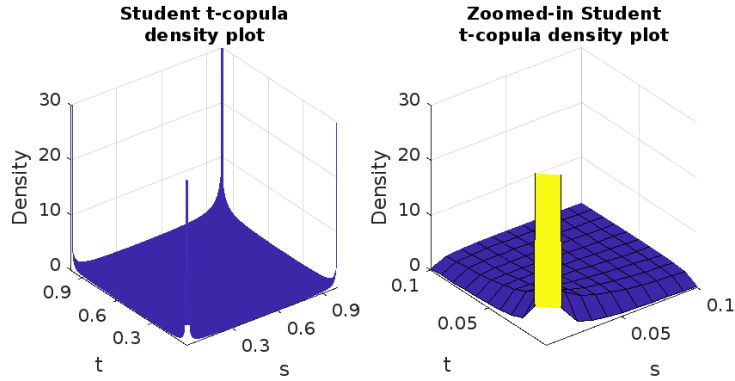


Figure 2.4: Student- t copula density, $\rho = 0.25$, $\nu = 3$.

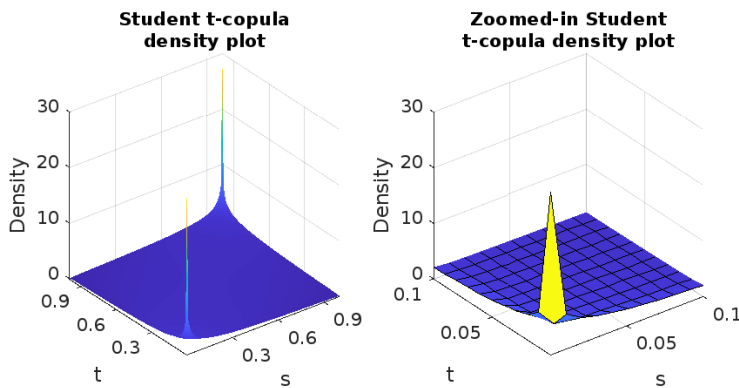


Figure 2.5: Student- t copula density, $\rho = 0.25$, $\nu = 60$.

In Figures 2.4 and 2.5, we observe the Student- t copula with 3 degrees of freedom and 60 degrees of freedom, respectively. The Student- t copula with 3 degrees of freedom exhibits heavier tails and greater tail dependence compared to the one with 60 degrees of freedom. As the degrees of freedom increase, the copula approaches the Gaussian copula depicted in Figure 2.3: with the same parameter ρ , when $\nu \rightarrow +\infty$. This convergence is indicative of the behavior of the Student- t copula toward a Gaussian copula as the tails become thinner with higher degrees of freedom.

Now it's time to explore the graphical representations of the Archimedean copula family, especially Clayton and Frank copulas.

As observed from Figures 2.6–2.7, as the coordinates approach the extreme point $(0, 0)$, the Frank copula tends towards a constant, whereas the Clayton copula is unbounded.

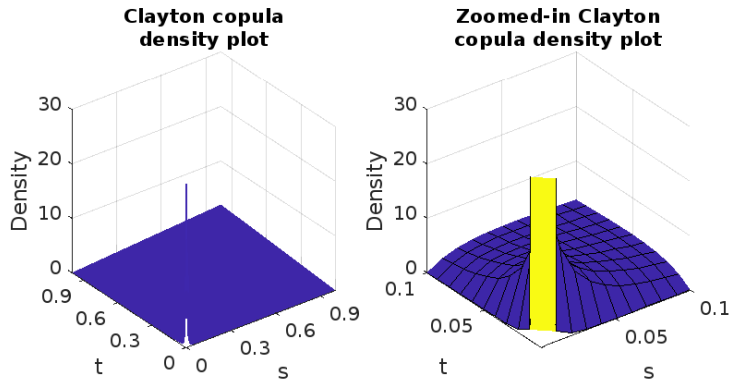


Figure 2.6: Clayton copula density, $\theta = 1$.

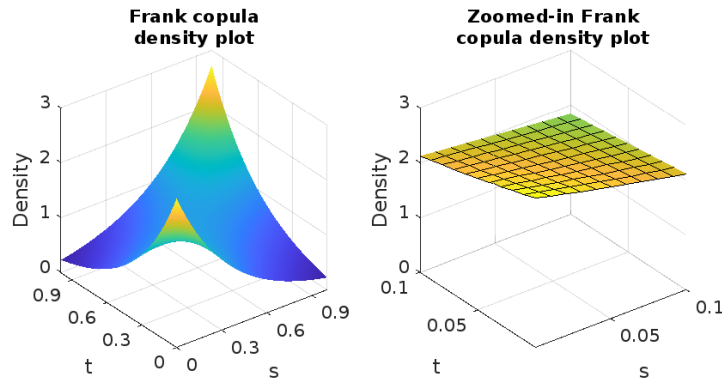


Figure 2.7: Frank copula density, $\theta = 2.5$.

There are obvious differences in the copula functions. Gaussian copula, Student- t copula and Frank copula have symmetric tails and can capture the symmetric tail correlations between random variables; Student- t copula has thicker tails, it is more sensitive to tail correlation changes between random variables and can capture symmetric tail correlations between random variables better. Frank copula has a simple structure and strong adaptability, and can meet the needs of most fields of application. Clayton copula is suitable for applications with asymmetric tails, lower tail correlation, and upper tail asymptotic independent two-dimensional random vectors.

2.3 Computation of limits of copula in extremes of domain

Except for computing the limits of copula extreme values, we should check the continuity of the function around the extreme points. We use the definition of continuity:

Definition 2.1. [23] A function f defined in the neighborhood of point a is continuous at point a if and only if $\lim_{x \rightarrow a} f(x) = f(a)$.

There are presented the final formulas for calculating the probability densities of studied copulas in Table 2.4.

Copula type	Probability density formula $c(u, v)$
Gaussian	$\frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho}{2(1-\rho^2)}(\rho x^2 - 2xy + \rho y^2)\right)$
Student- t	$\frac{1}{\sqrt{1-\rho^2}} \cdot \frac{\Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{1+\nu}{2}\right)^2} \cdot \frac{\left(\left(1 + \frac{x^2}{\nu}\right) \left(1 + \frac{y^2}{\nu}\right)\right)^{\frac{\nu+1}{2}}}{\left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}$
Clayton	$(\theta + 1)(uv)^{-\theta-1}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-2}, \quad \theta > 0$
Frank	$\frac{(-\theta) \cdot e^{-\theta(u+v)}(e^{-\theta} - 1)}{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)^2}$

Table 2.4: Probability densities of copulas [22].

Here:

- ρ is the correlation coefficient;
- ν is the number of degrees of freedom;
- Γ is the Gamma function.

Now let's examine the copula density values around the extreme point $(0, 0)$.

Definition 2.2. [15] Continuous random variables have a continuous distribution function. A random variable X is absolutely continuous if there exists a non-negative function $f_X : \mathbb{R} \rightarrow [0, \infty)$ (the density of random variable X) such that

$$F_X(t) = \int_{-\infty}^t f_X(u) du.$$

The density then satisfies

$$\int_{-\infty}^{\infty} f_X(u) du = 1.$$

Remark 2.3. Strictly speaking, densities of random vectors are not defined uniquely, only up to a set of measure zero. Each random vector has at most one continuous density.

All the formulas in Table 2.4 define densities continuous on the interior of the unit square, $(0, 1)^2$. We investigate these uniquely defined densities and we ask whether they have (possibly improper/infinite) limits at the border of the unit square, in particular in the origin $(0, 0)$.

To create visualizations and perform additional analyses, there were used MATLAB libraries. For a more precise reflection of copula values over the interval $(0, 0.1]$, 2D graphs were constructed. In these graphs, the diagonal line $s = t$ was taken as the horizontal axis, while the vertical axis directly reflects the copula values. Copulas were obtained for variables with a correlation coefficient of 0.25 (and corresponding coefficients θ for Archimedean copulas). The results are presented in Figures 2.8–2.9.

For Gaussian copula density we have:

$$\begin{aligned} & \lim_{(s,t) \rightarrow (0,0)} c(s, t) \\ &= \lim_{(x,y) \rightarrow (-\infty, -\infty)} \frac{1}{\sqrt{1-\rho^2}} \exp\left(\underbrace{-\frac{1}{2(1-\rho^2)}}_{<0} \cdot \underbrace{\rho^2}_{>0} \cdot \left(\underbrace{(x^2 + y^2)}_{\rightarrow +\infty} - \underbrace{\frac{2xy}{\rho}}_{\rightarrow +\infty}\right)\right) \end{aligned}$$

The value of the Gaussian probability density depends on the specific value of the correlation coefficient ρ between random variables:

- For $\rho < 0$:

$$-\frac{1}{2(1-\rho^2)} \cdot \rho^2 < 0; \quad (x^2 + y^2) \rightarrow +\infty; \quad \frac{-2xy}{\rho} \rightarrow +\infty, \text{ so}$$

$$\lim_{(x,y) \rightarrow (-\infty, -\infty)} c(F_X(x), F_Y(y)) = e^{-\infty} = 0.$$

- For $\rho = 0$: $\lim_{(x,y) \rightarrow (-\infty, -\infty)} c(F_X(x), F_Y(y)) = \frac{1}{\sqrt{1-0}} \cdot e^0 = 1.$

- For $\rho > 0$: For $y = kx$, $k \geq 0$:

In this case, the exponent's degree takes the form:

$$\lim_{x \rightarrow -\infty} \left(x^2 + k^2 \cdot x^2 - \frac{2kx^2}{\rho} \right) = \lim_{x \rightarrow -\infty} \left(x^2 \cdot \left(1 + k^2 - \frac{2}{\rho}k \right) \right)$$

Now let's solve the equation for which cases the expression $\left(1 + k^2 - \frac{2}{\rho}k \right)$ equals 0. This is a quadratic equation, and its solutions are given by the quadratic formula:

$$k_{1,2} = \frac{\frac{2}{\rho} \pm \sqrt{\frac{4}{\rho^2} - 4}}{2} = \frac{\frac{2}{\rho} \pm \sqrt{\frac{4 \cdot (1-\rho^2)}{\rho^2}}}{2} = \frac{\frac{2}{\rho} \pm \frac{2}{\rho} \cdot \sqrt{1-\rho^2}}{2} = \frac{1 \pm \sqrt{1-\rho^2}}{\rho}$$

After denoting $k_1 = \frac{1-\sqrt{1-\rho^2}}{\rho}$ and $k_2 = \frac{1+\sqrt{1-\rho^2}}{\rho}$, we have:

$$\left\{ \begin{array}{l}
 0 \leq k < k_1 : \lim_{x \rightarrow -\infty} x^2 \cdot \left(1 + k^2 - \frac{2}{\rho} k \right) = +\infty, \text{ then} \\
 \lim_{(s,t) \rightarrow (0,0)} c(s,t) = \lim_{x \rightarrow -\infty} \frac{1}{\sqrt{1-\rho^2}} \cdot e^x = 0 \\
 \\
 k = \frac{1-\sqrt{1-\rho^2}}{\rho} : \lim_{x \rightarrow -\infty} x^2 \cdot \left(1 + k^2 - \frac{2}{\rho} k \right) = 0, \text{ then} \\
 \lim_{(s,t) \rightarrow (0,0)} c(s,t) = \frac{1}{\sqrt{1-\rho^2}} \cdot e^0 = \frac{1}{\sqrt{1-\rho^2}} \\
 \\
 k_1 < k < k_2 : \lim_{x \rightarrow -\infty} x^2 \cdot \left(1 + k^2 - \frac{2}{\rho} k \right) = -\infty, \text{ then} \\
 \lim_{(s,t) \rightarrow (0,0)} c(s,t) = \lim_{x \rightarrow -\infty} \frac{1}{\sqrt{1-\rho^2}} \cdot e^{-x} = +\infty \\
 \\
 k = \frac{1+\sqrt{1-\rho^2}}{\rho} : \lim_{x \rightarrow -\infty} x^2 \cdot \left(1 + k^2 - \frac{2}{\rho} k \right) = 0, \text{ then} \\
 \lim_{(s,t) \rightarrow (0,0)} c(s,t) = \frac{1}{\sqrt{1-\rho^2}} \cdot e^0 = \frac{1}{\sqrt{1-\rho^2}} \\
 \\
 k_2 < k \leq 1 : \lim_{x \rightarrow -\infty} x^2 \cdot \left(1 + k^2 - \frac{2}{\rho} k \right) = +\infty, \text{ then} \\
 \lim_{(s,t) \rightarrow (0,0)} c(s,t) = \lim_{x \rightarrow -\infty} \frac{1}{\sqrt{1-\rho^2}} \cdot e^x = 0
 \end{array} \right.$$

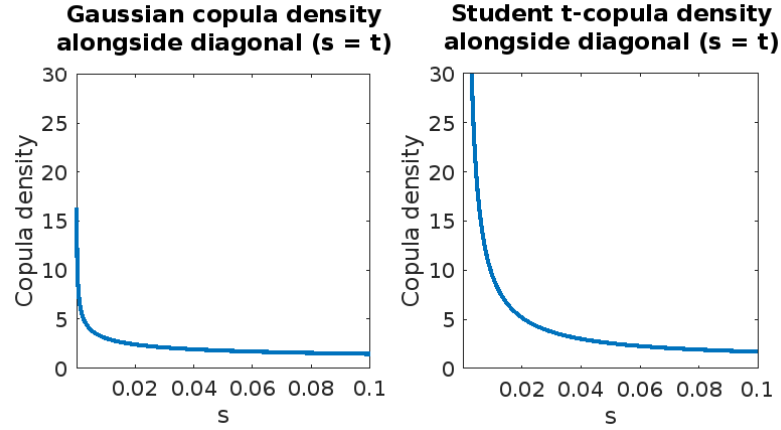


Figure 2.8: Zoomed-in Gaussian, Student- t copula density plots near $(0,0)$ ¹.

¹Parameters for plots were chosen as $\rho = 0.25$, $\nu = 3$.

Based on this description, the following conclusions can be drawn regarding the limits of Gaussian copula's probability density: for the case where $y = kx$, the limit is $+\infty$ for k in some interval containing 1, and the limit is 0 outside this interval. At the bounds of the interval, the limit is $\frac{1}{\sqrt{1-\rho^2}}$.

Hence, for $\rho > 0$, we can see that the Gaussian copula around the point $(0, 0)$ doesn't have a limit because its directional limits differ along different curves, spanning the range $(0, +\infty)$ on an arbitrarily small neighborhood.

For Student- t copula density we have:

$$\begin{aligned} & \lim_{(s,t) \rightarrow (0,0)} c(s,t) \\ &= \lim_{(x,y) \rightarrow (-\infty, -\infty)} \underbrace{\frac{1}{\sqrt{1-\rho^2}}}_{const \ c_1 > 0} \cdot \frac{\Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})^2} \cdot \frac{\left(\left(1 + \frac{x^2}{\nu}\right) \left(1 + \frac{y^2}{\nu}\right)\right)^{\frac{\nu+1}{2}}}{\left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \end{aligned}$$

The expression

$$\frac{\left(\left(1 + \frac{x^2}{\nu}\right) \left(1 + \frac{y^2}{\nu}\right)\right)^{\frac{\nu+1}{2}}}{\left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} = \sqrt{\frac{\left(\left(1 + \frac{x^2}{\nu}\right) \left(1 + \frac{y^2}{\nu}\right)\right)^{\nu+1}}{\left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{\nu+2}}}$$

is the square root of a rational function in variables x and y . We're taking the square root, its argument should be (and is) positive. We will continue working with the expression under the square root. Moreover, as $(x, y) \rightarrow (-\infty, -\infty)$, we neglect insignificant terms, as the value will prevail and carry greater weight for the terms with variables x, y .

To compute the limit, we use substitutions $x = -uv$, $y = -\frac{u}{v}$, where $u, v > 0$, $u \rightarrow +\infty$, v arbitrary.

$$\begin{aligned} \lim_{(x,y) \rightarrow (-\infty, -\infty)} \frac{(x^2 y^2)^{\nu+1}}{(x^2 - 2\rho xy + y^2)^{\nu+2}} &= \lim_{u \rightarrow +\infty} \frac{\left(u^2 v^2 \frac{u^2}{v^2}\right)^{\nu+1}}{\left(u^2 v^2 - 2\rho uv \frac{u}{v} + \frac{u^2}{v^2}\right)^{\nu+2}} \\ &= \lim_{u \rightarrow +\infty} \frac{u^{4(\nu+1)}}{u^{2(\nu+2)} \left(v^2 - 2\rho + \frac{1}{v^2}\right)^{\nu+2}} \\ &= \lim_{u \rightarrow +\infty} \frac{u^{2\nu}}{\left(v^2 - 2\rho + \frac{1}{v^2}\right)^{\nu+2}} \end{aligned}$$

As we can see, the numerator and the denominator depend on different variables, which are not connected to each other in any way. That's why

it's impossible to precisely determine the value of the limit for this part of an expression. The numerator can be arbitrarily large, so the values are not bounded from above. Also the denominator can be chosen arbitrarily large, allowing to achieve (for u fixed) an arbitrarily small value. In any neighborhood of $(0, 0)$, the values of the copula density span the whole interval $(0, \infty)$. Therefore, the limit of the copula density at the point $(0, 0)$ doesn't exist.

The solutions for the limits coincided with the plots on the graphs simulated in MATLAB (Figure 2.8). For a positive correlation coefficient $\rho = 0.25$ and movement along the diagonal $s = t$ to $(0, 0)$, the limit of the Gaussian copula density tends to $+\infty$. The limiting value of Student- t copula density is higher compared to the Gaussian copula due to the presence of additional positive coefficients (product of Gamma functions), which increase the weight of values as the number of the degrees of freedom grows.

Now let's examine the limits of Archimedean copulas near the origin $(0, 0)$.

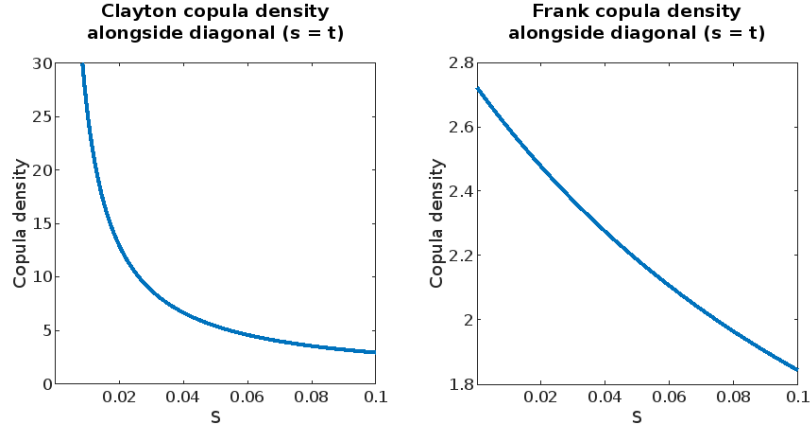


Figure 2.9: Zoomed-in Archimedean copula density plots near $(0, 0)^2$.

Clayton copulas for $\theta < 0$ are nilpotent [7], therefore equal to 0 in a neighborhood of $(0, 0)$ and so with zero density there, too. For $\theta > 0$, we consider the special case $u = v^k$, $0 < k < 1$. We analyze the limit

$$\lim_{v \rightarrow 0} c(v^k, v) = \lim_{v \rightarrow 0} (\theta + 1) v^{(k+1)(-\theta-1)} \underbrace{(v^{-k\theta} + v^{-\theta} - 1)}^{-\frac{1}{\theta}-2}.$$

In the limit, the sum $v^{-k\theta} + v^{-\theta} - 1$ may be replaced by its highest order term, which is $v^{-\theta}$,

$$\lim_{v \rightarrow 0} c(v^k, v) = \lim_{v \rightarrow 0} (\theta + 1) v^{(k+1)(-\theta-1) + (-\theta)(-\frac{1}{\theta}-2)} = \lim_{v \rightarrow 0} (\theta + 1) v^{\theta - (1+\theta)k}.$$

The parameter θ is fixed, while $k \in (0, 1)$ may be chosen arbitrarily. Thus the exponent $\theta - (1 + \theta)k$ can have any sign, making it possible to get results from 0 to ∞ . The limit $\lim_{(u,v) \rightarrow (0,0)} c(u, v)$ does not exist.

²Parameters for plots were chosen as Clayton $\theta = 1$, Frank $\theta = 2.5$.

For Frank copula density we have:

$$\lim_{(u,v) \rightarrow (0,0)} \frac{\overbrace{(-\theta) e^{-\theta(u+v)} (e^{-\theta} - 1)}^{\rightarrow 1}}{\underbrace{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)}_{\rightarrow 0}} = \frac{-\theta}{e^{-\theta} - 1}$$

This is always a positive finite number.

The solutions for the limits coincided with the plots on the graphs simulated in MATLAB. While moving along the diagonal $s = t$ to $(0, 0)$, the limit of the Clayton copula probability density tends to $+\infty$. The limiting value of Frank copula density has exact value, which fully corresponds to the data obtained in Figure 2.9 above.

The exact value of the copula density can be very useful in forecasting various events. For example, in an article [2] studying Value at Risk (VaR) – a method to estimate the worst risk of an investment, calculations were conducted using the Frank copula. Previous research on VaR by other researchers found that VaR value was reliable only for data fulfilling the normality assumption. Therefore, it was necessary to estimate VaR without ignoring the presence of heteroscedasticity and unfulfilled residual normality of the joint distribution model. This research aimed to measure VaR using the Frank Copula with stock return data from 2014 to 2018. They were able to predict the risk obtained from the calculation of VaR with the Frank copula and Monte Carlo simulation, and maximum loss in one-day prediction at high levels of confidence.

Now let's examine the copula density values around the extreme point $(1, 1)$.

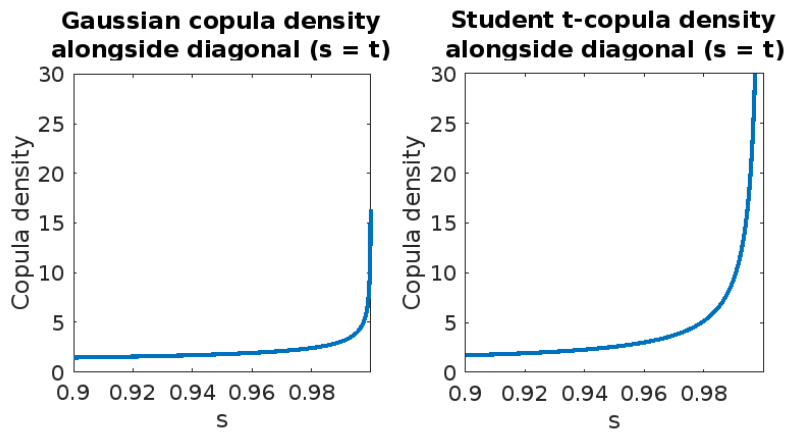


Figure 2.10: Zoomed-in Gaussian, Student- t copula density plots near $(1,1)$ ³.

For Gaussian and Student- t copula probability densities we have the same situation, as we had for the extreme point $(0, 0)$ (Figure 2.10).

³Parameters for plots were chosen as $\rho = 0.25$, $\nu = 3$.

As was analyzed in [11], Student- t copula can always be mistaken for a Gaussian copula if its number of degrees of freedom is sufficiently large. Then, depending on the correlation coefficient, the Student- t copula can predict a non-negligible tail dependence which is completely missed by the Gaussian copula assumption. In other words, the Gaussian copula predicts no tail dependencies and therefore does not account for extreme events that may occur simultaneously but nevertheless too rarely to modify the test statistics. To quantify the probability for neglecting such events, in this research have been investigated the situations when one is unable to distinguish between the Gaussian and Student- t copulas for a given number of degrees of freedom. It led to the conclusion that it may be very dangerous to embrace blindly the Gaussian copula hypothesis when the correlation coefficient between the pair of assets is too high as the tail dependence neglected by the Gaussian copula can be as large as 0.6.

Now let's examine the limits of Archimedean copulas near the point $(1, 1)$ because they will be different.

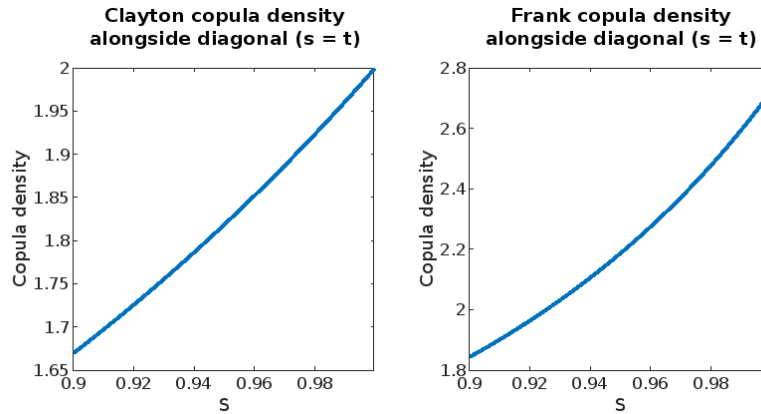


Figure 2.11: Zoomed-in Archimedean copula density plots near $(1, 1)$ ⁴.

For Clayton copula density we have:

$$\lim_{(u,v) \rightarrow (1,1)} \frac{\theta + 1}{u^{\theta+1} \cdot v^{\theta+1} \left(\frac{1}{u^\theta} + \frac{1}{v^\theta} - 1 \right)^{\frac{1}{\theta} + 2}} = \frac{\theta + 1}{1} = \theta + 1$$

If in the case of the point $(0, 0)$ the Clayton copula density doesn't have a limit, then at the point $(1, 1)$ it assumes a precise number. Therefore, it can be convenient to use the Clayton copula to forecast the behavior of functions at this point, as it has an exact value here.

⁴Parameters for plots were chosen as Clayton $\theta = 1$, Frank $\theta = 2.5$.

For Frank copula density we have:

$$\begin{aligned} \lim_{(u,v) \rightarrow (1,1)} \frac{(-\theta) \overbrace{e^{-\theta(u+v)}}^{e^{-2\theta}} (e^{-\theta} - 1)}{\underbrace{((e^{-\theta u} - 1)(e^{-\theta v} - 1) + e^{-\theta} - 1)}_{(e^{-\theta} - 1)^2}} &= \frac{(-\theta)e^{-2\theta}(e^{-\theta} - 1)}{(e^{-\theta}(e^{-\theta} - 1))^2} \\ &= \frac{(-\theta)e^{-2\theta}(e^{-\theta} - 1)}{e^{-2\theta}(e^{-\theta} - 1)^2} = \frac{-\theta}{e^{-\theta} - 1} \end{aligned}$$

The solutions for the limits coincided with the plots on the graphs simulated in MATLAB (Figure 2.11). In the direction along the diagonal $s = t$ to $(1, 1)$, the limits of the Clayton and Frank copula probability densities tend to some exact value.

In this chapter, we analyzed and calculated the limits of probability density for the main types of bivariate copulas. We aimed to derive the limits of copula density near the extremes of the domain. Our focus was on a conditionally independent distribution created by a mixture of independent components. We wanted to detect the limit of its probability density in extreme points. However, we did not achieve this. Even for one random vector with independent components, it usually has not a meaningful solution (the density often tends to $+\infty$ and its limit need not exist).

Chapter 3

Identifying the number of random variables included in the mixture

Many random variables can be expressed as mixtures of other random variables. The use of copulas provides a framework for probabilistic understanding the involvement of multiple random variables in such mixtures. We will create a definition which will be used throughout this chapter.

Definition 3.1 (Random Vector with Independent Components - RVIC). In 2D, a random vector with independent components (RVIC) can be defined as the formation of 2 independent random variables.

Recall that random variables X, Y are independent if for all intervals I, J the events $X \in I, Y \in J$ are independent, i.e.,

$$P(X \in I, Y \in J) = P(X \in I) \cdot P(Y \in J)$$

It is sufficient to consider intervals of the form $(-\infty, t]$, i.e.,

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) = F_X(x) \cdot F_Y(y)$$

for all $x, y \in \mathbb{R}$ [15], where F_X, F_Y are cumulative distribution functions of X, Y .

The RVIC from these random variables is $Z = (X, Y)$ with the joint cumulative distribution function

$$F_Z(x, y) = F_X(x) \cdot F_Y(y).$$

Let's consider RVICs $Z_1 = (X_1, Y_1)$ and $Z_2 = (X_2, Y_2)$. Let $Z = (X, Y)$ be a mixture of Z_1 and Z_2 with coefficients $c, 1 - c$.

Now let's define the differences between mixture of copulas and copula of a mixture:

1. **Mixture of copulas.** This involves combining multiple copulas to create a new copula. Each copula in the mixture can model different types of dependencies. The overall copula is a convex combination of individual copulas.

Abbreviation	Meaning
mu	mean
sigma	standard deviation
X_1, Y_1, X_2, Y_2	1D random variables [1 * 2450] and [1 * 1050]
Z_1	RVIC with coordinates (X ₁ , Y ₁)
Z_2	RVIC with coordinates (X ₂ , Y ₂)
num_points_Z1, num_points_Z2	numbers of points belonging to the mixture (2450 and 1050)
X	all the points from X ₁ and X ₂ for the mixture
Y	all the points from Y ₁ and Y ₂ for the mixture
Z	mixture with coordinates (X, Y)
sorted_X	a sorted array of X in ascending order to calculate distribution and quantile functions
ecdf_values_X	cumulative distribution function values of X
copula_values	copula values matrix [3500 * 3500]
n	number of dividing intervals
intervals	uniform distribution of n intervals on [0, 1]
points	indices for selecting copula values from the matrix
A	matrix (n + 1, n + 1) with copula values for selected points

Table 3.1: Table of notations for Chapter 3.

3.1 Step 1. Sampling random variables from normal distributions

Every random variable with a normal distribution was sampled using MATLAB. The means and the standard deviations have been set to different values. The code below describes this.

In Figure 3.1 we can observe the histogram of random variable X_1 . This plot resembles classical Gaussian curve. The other variables were sampled similarly, with different means and standard deviations.

```
% Generating a random variable X_1 with normal distribution
n_samples = 3500;
mu = 10;
sigma = 2;
X_1 = mu + sigma * randn(n_samples, 1);

% plot the graph
figure;
histogram(X_1, 50); % Histogram with 50 bins
title('Random Variable $X_1$ from Normal Distribution',
      'Interpreter', 'latex');
```

```
xlabel('Value $X_1$', 'Interpreter', 'latex');
ylabel('Frequency', 'Interpreter', 'latex');
```

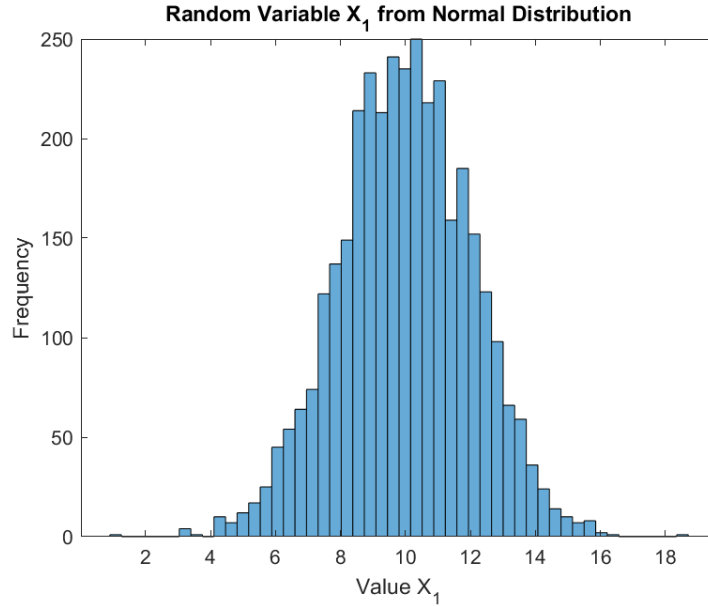


Figure 3.1: Histogram of simulated random variable X_1 .

3.2 Step 2. Creating independent random vectors

Then there were formed two RVICs Z_1, Z_2 from these simulated random variables: $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2)$.

Figure 3.2 shows the location of the simulated points in 2D. (Only 100 points are drawn from each component to keep the diagram uncluttered.)

3.3 Step 3. Creating a new mixed variable

We denote by Z the mixture of RVICs Z_1 and Z_2 with coefficient c [15]. It has a probability measure given by:

$$P_Z = c \cdot P_{Z_1} + (1 - c) \cdot P_{Z_2}$$

and a cumulative distribution function given by:

$$F_Z = c \cdot F_{Z_1} + (1 - c) \cdot F_{Z_2}$$

where

- P_Z is the probability measure of the mixture Z ,

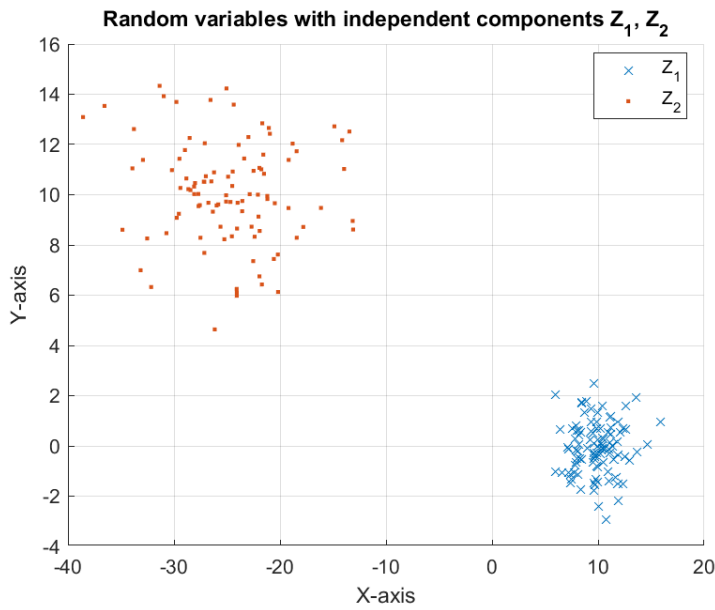


Figure 3.2: RVICs Z_1, Z_2 .

- F_Z is the cumulative distribution function of the mixture Z ,
- P_{Z_1} and F_{Z_1} are the probability measure and cumulative distribution function of Z_1 , respectively,
- P_{Z_2} and F_{Z_2} are the probability measure and cumulative distribution function of Z_2 , respectively,
- c is the mixing coefficient.

We would like the new mixed variable to have 3500 points, so we constructed it from 2450 points from RVIC Z_1 and 1050 points from RVIC Z_2 , so 3500 in total (with $c = 0.7$).

Here is the MATLAB code for the processes above:

```
num_points_Z1 = 0.7 * n_samples;
num_points_Z2 = 0.3 * n_samples;

X = [X1; X2];
Y = [Y1; Y2];
Z_mix = [X Y];
```

3.4 Step 4. Corresponding samples from uniform distributions on the interval $[0, 1]$

To find (samples of) uniform distribution corresponding to the simulated data, the following calculations were carried out.

The quantile function, also known as the percent-point function or inverse cumulative distribution function, can be calculated using the cumulative distribution function (CDF) of the normal distribution.

After transforming the first coordinate X of the mixture Z , we obtain a new random variable $S = F_X(X)$, and similarly for the second coordinate, $T = F_Y(Y)$ [15]. Now we have two uniformly distributed variables which will be further used in the mixture.

Here is the MATLAB code for the processes above:

```
sorted_X = sort(X);
for i = 1:n_samples
    ecdf_valuesX(i) = sum(X <= sorted_X(i)) / n_samples;
end
```

Similar calculations were performed for X_1, X_2, Y_1, Y_2 .

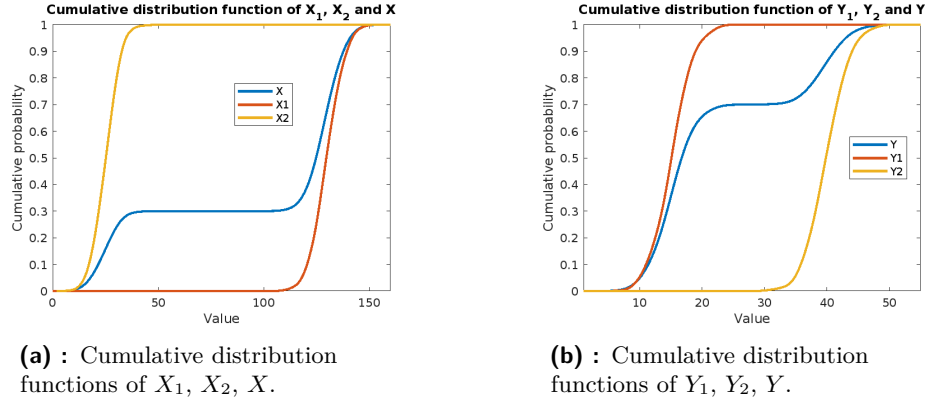


Figure 3.3: Cumulative distribution functions of simulated random variables.

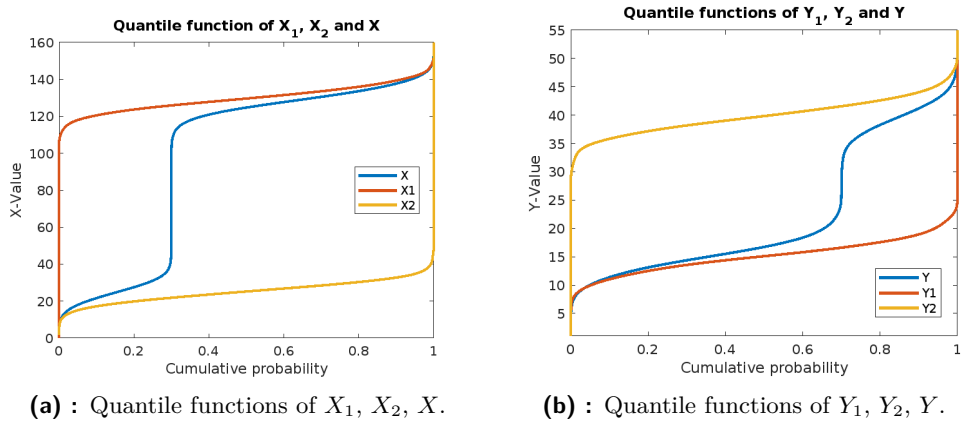


Figure 3.4: Quantile functions of simulated random variables.

In Figures 3.3–3.4 we can see the corresponding distributions or both of the transformed variables X and Y , which then will be used for copula construction.

3.5 Step 5. Construct the copula from two uniformly distributed variables

Now we are ready to construct an appropriate copula for our simulated data. The “copula values” will be a matrix $[3500 \times 3500]$, where for every point will be counted a cumulative probability of this point in 2D.

Here is the MATLAB code:

```
copula_values = zeros(n_samples);

for i = 1:n_samples
    for j = 1:n_samples
        count_points_in_quadrant =
            sum(X <= sorted_X(i) & Y <= sorted_Y(j));
        copula_values(i, j) = count_points_in_quadrant / n_samples;
    end
end
end
```

The graph of the resulting copula is drawn in Figure 3.5, its density in Figure 3.6. The latter is not sufficiently precise because it is a histogram of a discretized function. The density being small in some regions, the number of data points in the bins are small and subject to big fluctuations due to randomness of the simulation. A high number of points would improve it, but the computational complexity would have been high. The graph at least shows two heaps which originate from the two components of the mixture, rescaled to the unit interval.

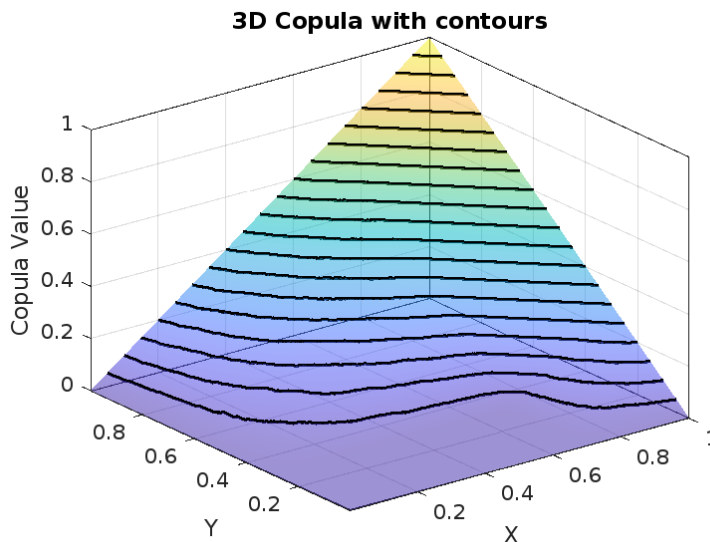


Figure 3.5: Copula CDF contours for simulated Z .

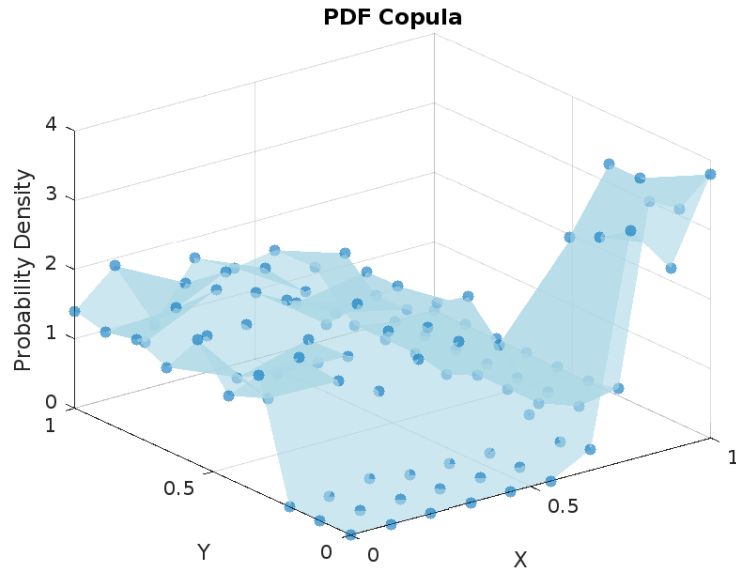


Figure 3.6: Copula density for simulated Z .

3.6 Step 6. Discretizing by SVD and essential rank

Next, let's envisage our continuous random variables in the form of discrete ones. Let's choose n partition points that will divide the data into $n + 1$ parts (n must be greater than the number of mixture components, i.e., $n \geq 2$). Create a matrix with copula values for these points.

Points were chosen using the equal interval method. For instance, if there were 3500 points in the simulated data, copula values were selected for points with indices 0, $3500/7$, $3500/7 \cdot 2$, $3500/7 \cdot 3$ and so on.

For each of the selected points (s, t) , the value of the copula is obtained using the formula:

$$C(s, t) = F_Z(F_X^{-1}(s), F_Y^{-1}(t))$$

where:

- F_Z is the cumulative distribution function of Z ,
- F_X^{-1}, F_Y^{-1} are the quantile functions of X, Y .

Finally, by discretization we construct a copula matrix from these calculated copula values, which will be used in subsequent calculations.

The code below describes the process of selecting discrete points and forming a matrix:

```
% Selecting points and making a matrix
n = 7;
```

```

intervals = linspace(0, 1, n+5);
points = zeros((n+1)^2, 2);
counter = 1;
for i = 5:n+5
    for j = 5:n+5
        points(counter, 1) = intervals(i);
        points(counter, 2) = intervals(j);
        counter = counter + 1;
    end
end

interp_values = interp2(ecdf_valuesX, ecdf_valuesY,
                        copula_values, points(:, 1), points(:, 2));

% Making matrix from points
A = reshape(interp_values, [n+1, n+1]);
disp(A);

```

The essential rank of a matrix is a concept related to singular value decomposition (SVD), which is a factorization of a matrix $A \in R^{m \times n}$ into the product of three matrices:

$$A = USV^T = \sum_{i=1}^p s_i u_i v_i^T,$$

where $p = \min(m, n)$, the matrix $S = \text{diag}(s_1 \dots s_p) \in R^{p \times p}$ is diagonal, and the matrices $U = [u_1 \dots u_p] \in R^{m \times p}$ and $V = [v_1 \dots v_p] \in R^{n \times p}$ have orthonormal columns [24].

The essential rank is the number of essential singular values in the diagonal matrix S – values, which are the biggest ones and as far from zero as possible.

The nearest lower-rank matrix B from SVD, with rank k , can be defined as:

$$B = US_k V^T = \sum_{i=1}^k s_i u_i v_i^T$$

This expression describes the construction of the nearest lower-rank matrix B from the SVD of matrix A , where only the first k singular values are retained. The matrix S_k is formed by retaining the top k singular values and zeroing out the remaining $n - k$ singular values.

There was made an SVD of matrix A in MATLAB:

```

A = copula_matrix;
rank_A = rank(A);
disp(rank_A); % 6

% Singular Value Decomposition

```

3. Identifying the number of random variables included in the mixture

```
[U, S, V] = svd(A);
disp(S);
```

Here you can see the resulting matrix S :

3.7879	0	0	0	0	0	0	0
0	0.3074	0	0	0	0	0	0
0	0	0.0053	0	0	0	0	0
0	0	0	0.0031	0	0	0	0
0	0	0	0	0.0014	0	0	0
0	0	0	0	0	0.0006	0	0
0	0	0	0	0	0	0.0000	0
0	0	0	0	0	0	0	0.0000

The rank of an original matrix A is 6, but the most significant singular values are the first two. That can be interpreted that the essential rank is equal to the number of components that entered the mixture. Thus, the essential rank of A is 2 (components Z_1, Z_2).

The same experiment was conducted for a mixture of more than two random vectors with independent components (RVICs).

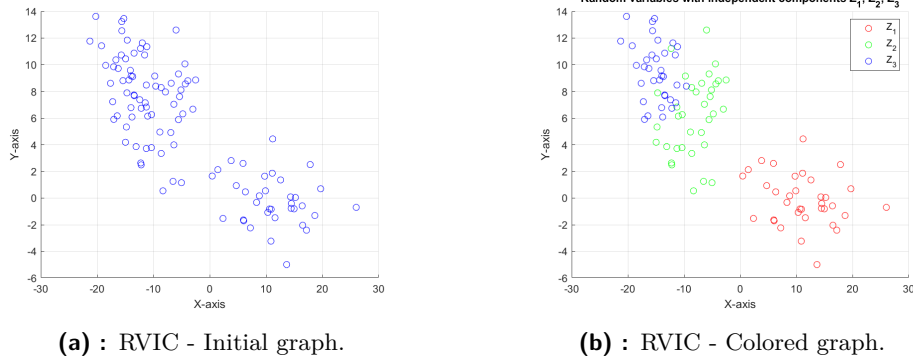


Figure 3.7: Graph of 3 simulated random vectors.

In Figure 3.7a, it is quite difficult to see exactly how many random vectors with independent components are depicted and contributing to the mixture, as their means and variances are close to each other.

The singular values matrix S looks like:

3.1056	0	0	0	0	0
0	0.3171	0	0	0	0
0	0	0.0211	0	0	0
0	0	0	0.0005	0	0
0	0	0	0	0.0001	0
0	0	0	0	0	0.0000



Chapter 4

Conclusions

In this thesis we studied the principles of modelling dependence of random variables by copulas. They are important for forecasting in real world in such fields as medicine, engineering, economics, agriculture, and so on. The thesis concentrated on bivariate normal distribution models.

We studied why the Gaussian copula [18], which was initially celebrated for its simplicity and effectiveness and enabled the modeling of complex financial risks, was leading to the creation of vast quantities of new securities. However, the method's limitations were ignored, and cracks began to appear in 2008 when the financial system faced unprecedented challenges, resulting in a global economic meltdown.

The linear nature of the Gaussian copula means that the copula fails to capture complex non-linear dependencies that may manifest in extreme situations. Particularly in economic phenomena prone to extreme fluctuations, such as financial crises, where non-linear and unstable deviations are crucial, the Gaussian copula may provide distorted and underestimated results [18].

Considering the lack of available literature sources mentioning formulas for the probability densities of copulas and existence of mistakes in it [9], it was decided to derive such formulas. There were derived some formulas for computing the density of key copula types (Gaussian, Student- t , Frank).

We analyzed, calculated and presented a probability density limiting values close to extreme points of some basic types of bivariate copulas such as Gaussian, Student- t and Archimedean family copulas (Clayton, Frank).

It was a mistake for economists to use the Gaussian copula to predict market behavior, since it need not have limits at the extremes. Its directional limits usually tend to $+\infty$ or 0. However, for example, Frank copula (near the extreme points of the domain) or Clayton copula (near the point $(1, 1)$) have exact limits that are expressed by formulas. Thus, they can be used to predict events that could lead to overheating of the economy following a potential crisis of any market, depending on parameters that are random variables [2].

We aimed to derive the limits of copula density near the extremes of the

domain. Our focus was on a conditionally independent distribution created by a mixture of independent components. We wanted to detect the limiting values there. However, we did not achieve this. Even for one random vector with independent components, a meaningful solution may not always exist. For example, Gaussian or Student- t copulas do not have limits near some extreme points.

During the process of analyzing copulas which are mixtures of some random vectors with independent components, we answered the question whether mathematical analysis, statistical tools and optimization can determine the number of such components constituting these copulas. We have proposed a methodology for identifying the number of components in a copula mixture. Through theoretical analysis and MATLAB implementation, we demonstrated the approach's applicability and provided insightful visualizations for it.

Therefore, having a copula describing a mixture of random vectors with independent components, we can predict how many such components formed that given copula. This tool can be useful in analyzing the dynamics of indicators at extreme points. From an economic perspective, this is an indicator for periods of economic overheating and signals a change in the behavior of economic agents and market dynamics.

Copulas are a complex topic, especially when working with more than two variables. Here there were computed and verified probability density functions for bivariate copulas. In real life, there are very few processes that depend only on 2 factors (causes), so it would be quite difficult to forecast their behavior without knowing the laws that govern their behavior as they approach extreme bounds.

Since the purpose of this work was to study the behavior of copulas and their applicability to economics, certain copulas (Frank, Clayton) could be used for these purposes, however, in the case of a large number of random variables, detailed analysis could be carried out more efficiently using computer technology (like MATLAB, R for graphs. Python, etc.).



Bibliography

- [1] B. Gräler. Introduction to Copulas. Electronic presentation, Institute for Geoinformatics, Muenster, February 2011. Online. [Accessed 11 November 2023]. Available from: http://www.graeler.org/copulaIntro/03_IntroToCopulas.pdf.
- [2] J. A. Handini, D. A. I. Maruddani, and D. Safitri. Frank copula on value at risk (VaR) of the construction of bivariate portfolio. Journal of Physics: Conference Series, 1217, May 2019.
- [3] F. Ielpo, C. Merhy, and G. Simon. Dealing with Risk Factors. Engineering Investment Process, Elsevier, 2017.
- [4] The MathWorks Inc. Copulas: Generate Correlated Samples. Online. [Accessed 15 May 2024]. Available from: <https://www.mathworks.com/help/stats/copulas-generate-correlated-samples.html>.
- [5] H. Joe. Dependence Modeling with Copulas. Monographs on Statistics and Applied Probability (Series). CRC Press, 2015.
- [6] B. M. G. Kibria and A. Joarder. A short review of multivariate t - distribution. Journal of Statistical Research ISSN, 40:256–422, 01 2006.
- [7] E.P. Klement, R. Mesiar, and E. Pap. Triangular Norms. Trends in Logic №8. Springer, 1 edition, 2000.
- [8] A. Kolesárová, R. Mesiar, J. Mordelová, and C. Sempi. Discrete copulas. IEEE Transactions on Fuzzy Systems, 14(5):698–705, 2006.
- [9] N. Kolev, L. Ferreira, and R. Aguilera. Dependence Analysis via Copulas. Electronic presentation, Institute of Mathematics and Statistics, University of São Paulo, Brazil, 2015. Online. [Accessed 15 March 2024]. Available from: https://edisciplinas.usp.br/pluginfile.php/7438154/mod_resource/content/1/Dependence%20Analysis%20via%20Copulas%20ICMC.pdf.
- [10] C. N. Lai and N. Balakrishnan. Continuous Bivariate Distributions. Springer, 2 edition, 2009.

- [11] Y. Malevergne and D. Sornette. Testing the Gaussian copula hypothesis for financial assets dependences. Quantitative Finance, 3:231–250, September 2003.
- [12] E. Marsden. Copula and multivariate dependencies. Online. [Accessed 12 May 2024]. Available from: <https://risk-engineering.org/static/PDF/slides-copula.pdf>.
- [13] P. G. F. McEwan. The GARCH-EVT-Copula model and simulation in scenario-based asset allocation. Faculty of Science, Nelson Mandela Metropolitan University, December 2015.
- [14] C. Meyer. The Bivariate Normal Copula. DZ BANK AG, Platz der Republik, D-60265 Frankfurt, 12 2009.
- [15] M. Navara. Probability and Mathematical Statistics (in Czech). Electronic textbook, CTU, Prague, 2007.
- [16] R. B. Nelsen. An Introduction to Copulas. Springer, 2006.
- [17] OpenAI. Chat GPT-3.5. Online. [Accessed 18 May 2024]. Available from: <https://www.openai.com/gpt-3/>.
- [18] F. Salmon. Recipe For Disaster: The Formula That Killed Wall Street. Wired Magazine, 8, February 2009.
- [19] A. Sklar. Fonctions de répartition à n dimensions et leurs marges, volume 8. Publ. Inst. Statist. Univ. Paris, 1959.
- [20] A. Staudt. Tail Risk, Systemic Risk and Copulas. Casualty Actuarial Society E-Forum, 2:1–23, nov 2010.
- [21] A. Sundaresan, P. K. Varshney, and N. S. V. Rao. Copula-Based Fusion of Correlated Decisions. IEEE Transactions on Aerospace and Electronic Systems, 47:454–471, January 2011.
- [22] DataCebo Team. Copulas: A Python library For Modeling Multivariate Distributions. Online. [Accessed 28 April 2024]. Available from: <https://github.com/sdv-dev/Copulas>.
- [23] J. Tkadlec. Differential and Integral Calculus of Functions of One Variable (in Czech). Electronic textbook, CTU, Prague, 2011.
- [24] T. Werner. Optimization (in Czech). Electronic textbook, CTU, Prague, 2023.
- [25] E. Zivot. Copulas. Electronic presentation, University of Washington, May 2013. Online. [Accessed 20 April 2024]. Available from: <https://faculty.washington.edu/ezivot/econ589/copulas.pdf>.

Appendix A

MATLAB codes

The MATLAB code for operations with 2 random vectors with independent components.

```
% Table of notations:
%
% n_samples - number of points in each random variable (3500)
% mu - mean
% sigma - deviation
% X_1, X_2, Y_1, Y_2 - 1D random variables [1 * 3500]
% Z_1 - 2D random variable with coordinates (X_1, Y_1)
% Z_2 - 2D random variable with coordinates (X_2, Y_2)
% num_points_Z1 / Z2 - number of points belonging to the mixture
% X1_chosen - chosen points from variable X_1.. etc.
% X - chosen points from X_1 and X_2 for the mixture
% Y - chosen points from Y_1 and Y_2 for the mixture
% Z_mix - mixture with coordinates (X, Y)
% sorted_X - sorted array X in ascending order for calculating
                the values of the distribution function
% ecdf_valuesX - values of the distribution function X in
                ascending order
% copula_values - copula matrix [3500 * 3500]

n_samples = 3500;
mu = 10; % mean
sigma = 2; % standard deviation
X_1 = mu + sigma * randn(n_samples, 1);

figure;
histogram(X_1, 50); % Histogram with 50 bins
title('Random Variable X_1 from Normal Distribution');
xlabel('Value X_1');
ylabel('Frequency');
```

```
saveas(gcf, 'x1_histogram.png');

mu = 0; % mean
sigma = 1; % standard deviation
Y_1 = mu + sigma * randn(n_samples, 1);

mu = -25; % mean
sigma = -5; % standard deviation
X_2 = mu + sigma * randn(n_samples, 1);

mu = 10; % mean
sigma = 2; % standard deviation
Y_2 = mu + sigma * randn(n_samples, 1);

figure;
scatter(X_1, Y_1, 'x', 'DisplayName', 'Z_1');
hold on;
scatter(X_2, Y_2, '.', 'DisplayName', 'Z_2');
hold on;
%scatter(X, Y, 'o', 'filled', 'DisplayName', 'Z');
%hold off;

title('Random variables with independent components Z_1, Z_2');
xlabel('X-axis');
ylabel('Y-axis');
legend('Location', 'best');
grid on;
saveas(gcf, 'original_rv.png');

% Mixture
num_points_Z1 = 0.7 * n_samples;
num_points_Z2 = 0.3 * n_samples;

indices_Z1 = randperm(n_samples, num_points_Z1);
indices_Z2 = randperm(n_samples, num_points_Z2);

X1_chosen = X_1(indices_Z1);
X2_chosen = X_2(indices_Z2);
Y1_chosen = Y_1(indices_Z1);
Y2_chosen = Y_2(indices_Z2);

X = [X1_chosen; X2_chosen];
Y = [Y1_chosen; Y2_chosen];

Z_mix = [X Y];
```

```

%Z_mix = [X1_chosen; Y1_chosen; X2_chosen; Y2_chosen];
disp(size(Z_mix));

figure;
scatter(X1_chosen, Y1_chosen, 'x', 'DisplayName', 'Chosen from Z_1');
hold on;
scatter(X2_chosen, Y2_chosen, '.', 'DisplayName', 'Chosen from Z_2');
hold on;

title('Chosen Random variables with independent components');
xlabel('X-axis');
ylabel('Y-axis');
legend('Location', 'best');
grid on;
saveas(gcf, 'chosen_rvic.png');

% Estimation of density to create a smooth CDF
[f_X, x_X] = ksdensity(X);
[f_X1, x_X1] = ksdensity(X1_chosen);
[f_X2, x_X2] = ksdensity(X2_chosen);

minX = min([min(x_X), min(x_X1), min(x_X2)]);
maxX = max([max(x_X), max(x_X1), max(x_X2)]);
x_X_0 = [minX, x_X, maxX];
x_X_01 = [minX, x_X1, maxX];
x_X_02 = [minX, x_X2, maxX];
f_X_0 = [0, cumsum(f_X) / sum(f_X), 1];
f_X_01 = [0, cumsum(f_X1) / sum(f_X1), 1];
f_X_02 = [0, cumsum(f_X2) / sum(f_X2), 1];

figure;
plot(x_X_0, f_X_0, 'LineWidth', 2);
hold on;
plot(x_X_01, f_X_01, 'LineWidth', 2);
hold on;
plot(x_X_02, f_X_02, 'LineWidth', 2);
title('Cumulative distribution function of X_1, X_2 and X');
xlabel('Value');
ylabel('Cumulative probability');
legend('X', 'X_1', 'X_2', 'Location', 'best');
xlim([-43, 20]);
saveas(gcf, 'cdf_X.png');

[f_Y, y_Y] = ksdensity(Y);
[f_Y1, y_Y1] = ksdensity(Y1_chosen);
[f_Y2, y_Y2] = ksdensity(Y2_chosen);

```

```
minY = min([min(y_Y), min(y_Y1), min(y_Y2)]);
maxY = max([max(y_Y), max(y_Y1), max(y_Y2)]);
y_Y_0 = [minY, y_Y, maxY];
y_Y_01 = [minY, y_Y1, maxY];
y_Y_02 = [minY, y_Y2, maxY];
f_Y_0 = [0, cumsum(f_Y) / sum(f_Y), 1];
f_Y_01 = [0, cumsum(f_Y1) / sum(f_Y1), 1];
f_Y_02 = [0, cumsum(f_Y2) / sum(f_Y2), 1];

figure;
plot(y_Y_0, f_Y_0, 'LineWidth', 2);
hold on;
plot(y_Y_01, f_Y_01, 'LineWidth', 2);
hold on;
plot(y_Y_02, f_Y_02, 'LineWidth', 2);
title('Cumulative distribution function of Y_1, Y_2 and Y');
xlabel('Value');
ylabel('Cumulative probability');
legend('Y', 'Y1', 'Y2', 'Location', 'best');
xlim([-4, 17]);
ylim([0, 1]);
saveas(gcf, 'cdf_Y.png');

% Quantile functions for X
figure;
plot(f_X_0, x_X_0, 'LineWidth', 2);
hold on;
plot(f_X_01, x_X_01, 'LineWidth', 2);
hold on;
plot(f_X_02, x_X_02, 'LineWidth', 2);
title('Quantile function of X_1, X_2 and X');
xlabel('Cumulative probability');
ylabel('X-Value');
legend('X', 'X_1', 'X_2', 'Location', 'best');
ylim([-43, 20]);
saveas(gcf, 'quantile_X.png');

% for Y
figure;
plot(f_Y_0, y_Y_0, 'LineWidth', 2);
hold on;
plot(f_Y_01, y_Y_01, 'LineWidth', 2);
hold on;
plot(f_Y_02, y_Y_02, 'LineWidth', 2);
```

```

title('Quantile functions of Y_1, Y_2 and Y');
xlabel('Cumulative probability');
ylabel('Y-Value');
legend('Y', 'Y_1', 'Y_2', 'Location', 'best');
ylim([-4, 17]);
xlim([0, 1]);
saveas(gcf, 'quantile_Y.png');

% Sort
sorted_X = sort(X);
sorted_X1 = sort(X1_chosen);
sorted_X2 = sort(X2_chosen);
sorted_Y = sort(Y);
sorted_Y1 = sort(Y1_chosen);
sorted_Y2 = sort(Y2_chosen);

% CDF by hands for 1000 samples
ecdf_valuesX = zeros(size(X));
ecdf_valuesX1 = zeros(size(X1_chosen));
ecdf_valuesX2 = zeros(size(X2_chosen));
ecdf_valuesY = zeros(size(Y));
ecdf_valuesY1 = zeros(size(Y1_chosen));
ecdf_valuesY2 = zeros(size(Y2_chosen));

for i = 1:n_samples
    ecdf_valuesX(i) = sum(X <= sorted_X(i)) / n_samples;
    ecdf_valuesY(i) = sum(Y <= sorted_Y(i)) / n_samples;
end
for i = 1:num_points_Z1
    ecdf_valuesX1(i) = sum(X1_chosen <= sorted_X1(i)) / num_points_Z1;
    ecdf_valuesY1(i) = sum(Y1_chosen <= sorted_Y1(i)) / num_points_Z1;
end
for i = 1:num_points_Z2
    ecdf_valuesX2(i) = sum(X2_chosen <= sorted_X2(i)) / num_points_Z2;
    ecdf_valuesY2(i) = sum(Y2_chosen <= sorted_Y2(i)) / num_points_Z2;
end

figure;
plot(sorted_X, ecdf_valuesX, 'LineWidth', 2);
hold on;
plot(sorted_X1, ecdf_valuesX1, 'LineWidth', 2);
hold on;
plot(sorted_X2, ecdf_valuesX2, 'LineWidth', 2);
title('Cumulative distribution function of X_1, X_2 and X (of points)');
xlabel('Value');
ylabel('Cumulative probability');

```

```
legend('X', 'X1', 'X2', 'Location', 'best');

% Creating copula values matrix for 3500 simulated points
copula_values = zeros(n_samples);

for i = 1:n_samples
    for j = 1:n_samples
        count_points_in_quadrant = sum(X <= sorted_X(i)
            & Y <= sorted_Y(j));
        copula_values(i, j) = count_points_in_quadrant / n_samples;
    end
end

%disp(size(copula_values));
%disp(copula_values);

% CDF copula
[X_mesh, Y_mesh] = meshgrid(ecdf_valuesX, ecdf_valuesY);
figure;
contour3(X_mesh, Y_mesh, copula_values, 'EdgeColor', 'k',
    'LineWidth', 1.5, 'LevelStep', 0.05);
hold on;
surf(X_mesh, Y_mesh, copula_values, 'FaceAlpha', 0.5,
    'EdgeColor', 'none');
title('3D Copula with contours');
xlabel('X');
ylabel('Y');
zlabel('Copula Value');
grid on;
saveas(gcf, 'copula_cdf.png');

% PDF manual calculations for 10 * 10 intervals in 2D
step = 10;
pdf_copula_values = zeros(step, step);
area = step / n_samples;

% Calculations for the first rectangle X[0, 100]
for j = 2:step
    x_max = sorted_X(1/area);
    x_min = sorted_X(1);
    y_max = sorted_Y(j * (1/area));
```



```

y_min = sorted_Y((j-1) * (1/area));
count_points_quadrant = sum(X >= x_min & X <= x_max &
                            Y > y_min & Y <= y_max);
% density is probability divided by area
pdf_copula_values(1, j) = count_points_quadrant / n_samples / 0.01;
end
% Calculations for the first rectangle Y[0, 100]
for i = 2:step
    y_max = sorted_Y((1/area));
    y_min = sorted_Y(1);
    x_max = sorted_X(i * (1/area));
    x_min = sorted_X((i-1) * (1/area));
    count_points_quadrant = sum(X > x_min & X <= x_max &
                                Y >= y_min & Y <= y_max);
    pdf_copula_values(i, 1) = count_points_quadrant / n_samples / 0.01;
end
% Calculations for the other points
for i = 2:step
    for j = 2:step
        x_max = sorted_X(i * (1/area));
        y_max = sorted_Y(j * (1/area));
        x_min = sorted_X((i-1) * (1/area));
        y_min = sorted_Y((j-1) * (1/area));
        count_points_quadrant = sum(X > x_min & X <= x_max
                                    & Y > y_min & Y <= y_max);
        pdf_copula_values(i, j) = count_points_quadrant / n_samples / 0.01;
    end
end

% Check that the sum of the probabilities is 100
total_probability = sum(pdf_copula_values(:));
disp(total_probability);

% uniform axes X and Y,
% linspace granted equal division into 10 blocks
% on the interval [0, 1] for the X and Y axes
[X_mesh, Y_mesh] = meshgrid(linspace(0, 1, step), linspace(0, 1, step));

% PDF copula plot manually
light_blue_color = [173, 216, 230] / 255;
%
figure;
surf(X_mesh, Y_mesh, pdf_copula_values, 'EdgeColor', 'none',
     'FaceColor', light_blue_color, 'FaceAlpha', 0.6);
hold on;
scatter3(X_mesh(:), Y_mesh(:), pdf_copula_values(:), 'filled',

```

```
        'MarkerFaceAlpha', 0.6);
title('PDF Copula');
xlabel('X');
ylabel('Y');
zlabel('Probability Density');
grid on;
saveas(gcf, 'copula_pdf.png');

% Table of notations:
%
% n          - number of dividing intervals (7)
% intervals  - uniform distribution for n intervals on [0, 1]
% points     - indices for selecting values from the copula matrix
% interp_values - array of length (n+1)*(n+1) with values
              - of the copula for the selected points
% A          - matrix (n+1, n+1) with values
              - of the copula for the selected points

% Discretizing
n = 7;
intervals = linspace(0, 1, n+5);
points = zeros((n+1)^2, 2);
counter = 1;
for i = 5:n+5
    for j = 5:n+5
        points(counter, 1) = intervals(i);
        points(counter, 2) = intervals(j);
        counter = counter + 1;
    end
end

interp_values = interp2(ecdf_valuesX, ecdf_valuesY, copula_values,
                        points(:, 1), points(:, 2));

% Making matrix from points
A = reshape(interp_values, [n+1, n+1]);

rank_A = rank(A);
disp(rank_A);

% Singular Value Decomposition
[U, S, V] = svd(A);
disp(S);
```

The MATLAB code for operations with 3 random vectors with independent components.

```
n_samples = 3500;
mu = 10; % mean
sigma = 2; % standard deviation %5
X_1 = mu + sigma * randn(n_samples, 1);

mu = 0; % mean
sigma = 2; % standard deviation
Y_1 = mu + sigma * randn(n_samples, 1);

mu = -8; % mean
sigma = 4; % standard deviation
X_2 = mu + sigma * randn(n_samples, 1);

mu = 6; % mean
sigma = 3; % standard deviation
Y_2 = mu + sigma * randn(n_samples, 1);

mu = -15; % mean
sigma = 3; % standard deviation
X_3 = mu + sigma * randn(n_samples, 1);

mu = 10; % mean
sigma = 2; % standard deviation
Y_3 = mu + sigma * randn(n_samples, 1);

%
figure;
scatter(X_1, Y_1, 'o', 'DisplayName', 'Z_1', 'MarkerEdgeColor', 'b');
hold on;
scatter(X_2, Y_2, 'o', 'DisplayName', 'Z_2', 'MarkerEdgeColor', 'b');
hold on;
scatter(X_3, Y_3, 'o', 'DisplayName', 'Z_3', 'MarkerEdgeColor', 'b');
hold off;
title('RVIC Z_1, Z_2, Z_3');
xlabel('X-axis');
ylabel('Y-axis');
%legend('Location', 'best');
grid on;
saveas(gcf, 'question_3rv.png');
```

```
% for question
figure;
scatter(X_1, Y_1, 'o', 'DisplayName', 'Z_1', 'MarkerEdgeColor', 'r');
hold on;
scatter(X_2, Y_2, 'o', 'DisplayName', 'Z_2', 'MarkerEdgeColor', 'g');
hold on;
scatter(X_3, Y_3, 'o', 'DisplayName', 'Z_3', 'MarkerEdgeColor', 'b');
hold off;
title('RVIC Z_1, Z_2, Z_3');
xlabel('X-axis');
ylabel('Y-axis');
legend('Location', 'best');
grid on;
saveas(gcf, 'original_3rv.png');

% Mixture
num_points_Z1 = 0.4 * n_samples;
num_points_Z2 = 0.2 * n_samples;
num_points_Z3 = 0.4 * n_samples;

indices_Z1 = randperm(n_samples, num_points_Z1);
indices_Z2 = randperm(n_samples, num_points_Z2);
indices_Z3 = randperm(n_samples, num_points_Z3);

X1_chosen = X_1(indices_Z1);
X2_chosen = X_2(indices_Z2);
X3_chosen = X_3(indices_Z3);
Y1_chosen = Y_1(indices_Z1);
Y2_chosen = Y_2(indices_Z2);
Y3_chosen = Y_3(indices_Z3);

X = [X1_chosen; X2_chosen; X3_chosen];
Y = [Y1_chosen; Y2_chosen; Y3_chosen];

Z_mix = [X Y];

figure;
scatter(X1_chosen, Y1_chosen, 'x', 'DisplayName', 'Chosen from Z_1');
hold on;
scatter(X2_chosen, Y2_chosen, '.', 'DisplayName', 'Chosen from Z_2');
hold on;
scatter(X3_chosen, Y3_chosen, '.', 'DisplayName', 'Chosen from Z_3');
title('Chosen RVIC');
xlabel('X-axis');
```

```

ylabel('Y-axis');
legend('Location', 'best');
grid on;
saveas(gcf, 'chosen_rvic3.png');

% Estimation of density to create a smooth CDF
[f_X, x_X] = ksdensity(X);
[f_X1, x_X1] = ksdensity(X1_chosen);
[f_X2, x_X2] = ksdensity(X2_chosen);
[f_X3, x_X3] = ksdensity(X3_chosen);
%
figure;
plot(x_X, cumsum(f_X) / sum(f_X), 'LineWidth', 2);
hold on;
plot(x_X1, cumsum(f_X1) / sum(f_X1), 'LineWidth', 2);
hold on;
plot(x_X2, cumsum(f_X2) / sum(f_X2), 'LineWidth', 2);
hold on;
plot(x_X3, cumsum(f_X3) / sum(f_X3), 'LineWidth', 2);
title('Cumulative distribution function of X_1, X_2, X_3 and X');
xlabel('Value');
ylabel('Cumulative probability');
legend('X', 'X1', 'X2', 'X3', 'Location', 'best');
saveas(gcf, 'cdf_X3.png');

[f_Y, y_Y] = ksdensity(Y);
[f_Y1, y_Y1] = ksdensity(Y1_chosen);
[f_Y2, y_Y2] = ksdensity(Y2_chosen);
[f_Y3, y_Y3] = ksdensity(Y3_chosen);

%
figure;
plot(y_Y, cumsum(f_Y) / sum(f_Y), 'LineWidth', 2);
hold on;
plot(y_Y1, cumsum(f_Y1) / sum(f_Y1), 'LineWidth', 2);
hold on;
plot(y_Y2, cumsum(f_Y2) / sum(f_Y2), 'LineWidth', 2);
hold on;
plot(y_Y3, cumsum(f_Y3) / sum(f_Y3), 'LineWidth', 2);
title('Cumulative distribution function of Y_1, Y_2, Y_3 and Y');
xlabel('Value');
ylabel('Cumulative probability');
legend('Y', 'Y1', 'Y2', 'Y3', 'Location', 'best');
saveas(gcf, 'cdf_Y3.png');

```

```
% Sort
sorted_X = sort(X);
sorted_X1 = sort(X1_chosen);
sorted_X2 = sort(X2_chosen);
sorted_Y = sort(Y);
sorted_Y1 = sort(Y1_chosen);
sorted_Y2 = sort(Y2_chosen);

% CDF by hands for 3500 samples
ecdf_valuesX = zeros(size(X));
ecdf_valuesY = zeros(size(Y));
for i = 1:n_samples
    ecdf_valuesX(i) = sum(X <= sorted_X(i)) / n_samples;
    ecdf_valuesY(i) = sum(Y <= sorted_Y(i)) / n_samples;
end

copula_values = zeros(n_samples);
for i = 1:n_samples
    for j = 1:n_samples
        count_points_in_quadrant = sum(X <= sorted_X(i) &
                                        Y <= sorted_Y(j));
        copula_values(i, j) = count_points_in_quadrant / n_samples;
    end
end

%
[X_mesh, Y_mesh] = meshgrid(ecdf_valuesX, ecdf_valuesY);
figure;
contour3(X_mesh, Y_mesh, copula_values, 'EdgeColor',
         'k', 'LineWidth', 1.5);
hold on;
surf(X_mesh, Y_mesh, copula_values, 'FaceAlpha', 0.5,
     'EdgeColor', 'none');
title('3D Copula with contours');
xlabel('X');
ylabel('Y');
zlabel('Copula Value');
grid on;
```