



**SEBASTIÃO VARGAS RAPOSO MATEUS**

Licenciatura em Matemática

**CÓPULAS NA MODELAÇÃO DOS RISCOS DE  
RESPONSABILIDADE CIVIL E DANOS  
PRÓPRIOS NO SEGURO AUTOMÓVEL**

MESTRADO EM MATEMÁTICA ATUARIAL

Universidade NOVA de Lisboa  
Setembro, 2023



# CÓPULAS NA MODELAÇÃO DOS RISCOS DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS NO SEGURO AUTOMÓVEL

**SEBASTIÃO VARGAS RAPOSO MATEUS**

Licenciatura em Matemática

**Orientadores:** Pedro Alexandre da Rosa Corte Real

*Professor Auxiliar, Faculdade Ciências e Tecnologia da Univ. Nova de Lisboa*

Gracinda Rita Diogo Guerreiro

*Professora Associada, Faculdade Ciências e Tecnologia da Univ. Nova de Lisboa*

**Júri**

**Presidente:** Doutor Luís Pedro Carneiro Ramos

*Professor Auxiliar do Departamento de Matemática, NOVA FCT*

**Arguente:** Doutor Manuel Leote Tavares Inglês Esquível

*Professor Associado com Agregação do Departamento de Matemática, NOVA FCT*

MESTRADO EM MATEMÁTICA ATUARIAL

Universidade NOVA de Lisboa

Setembro, 2023

## **Cópuas na Modelação dos Riscos de Responsabilidade Civil e Danos Próprios no Seguro Automóvel**

Copyright © Sebastião Vargas Raposo Mateus, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

*Aos meus.*

## AGRADECIMENTOS

Para iniciar os agradecimentos, gostaria de mostrar gratidão pelos meus orientadores, Professor Doutor Pedro Corte Real e Professora Gracinda Rita Diogo Guerreiro, pela competência e rigor, por todo o apoio e disponibilidade demonstradas para responder às minhas dúvidas. De facto, esta dissertação não teria sido possível sem eles, e por isso, o meu mais sincero obrigado.

À minha família, agradeço a paciência e o carinho, sem vocês este percurso não seria de todo possível. Aos meus pais, Maria João e Francisco, que acreditaram em mim e me deram todos os recursos para que fosse possível tudo o que eu desejei. Ao meu irmão, Flávio, que foi morar para longe mas continua muito perto. Obrigado por sempre me apoiarem e motivarem, nunca me deixando duvidar das minhas capacidades, foi o vosso entusiasmo e encorajamento que me levaram à conclusão deste trabalho. Realmente as melhores pessoas do mundo!

À minha namorada, Beatriz Palhinhas, por estar sempre ao meu lado, quer nos melhores e piores momentos. Obrigado por me apoiares, incentivares-me e de me ajudares, mesmo quando estou impossível. Tenho a certeza que sem ti não tinha chegado onde cheguei.

Aos meus amigos que fiz ao longo do meu percurso académico, aos amigos feitos na licenciatura em Coimbra, Pedro, Kika, Sofia, Fátima, Sanches e à malta toda da República que para sempre estarão no meu coração.

## RESUMO

Todos os modelos são simplificações da realidade e a indústria do risco, nomeadamente a indústria seguradora, quando avalia o impacto financeiro da concretização do risco está obrigada a introduzir muitas simplificações, algumas de modo explícito, na forma de pressupostos. Um dos pressupostos mais comuns numa avaliação atuarial é a de independência entre alguns ou mesmo todos os riscos (eventos ou coberturas) que se procuram modelar.

A dependência entre a frequência e os montantes das indemnizações de vários riscos segurados é reconhecida como um dos principais fatores que podem influenciar as perdas agregadas da seguradora. Assumir erroneamente essa independência pode resultar em perdas inesperadas significativas.

Procurando compreender melhor a relação de dependência entre 2 coberturas que são bastantes estudadas na indústria seguradora, esta dissertação dedica-se ao estudo da relação de dependência do valor dos sinistros em duas coberturas geralmente consideradas independentes e presentes no seguro automóvel: Danos Próprios e Responsabilidade Civil.

O conceito matemático usado na dissertação para estimar a relação de dependência é o conceito de Cópula. As Cópulas têm uma tradução probabilística que usamos para descrever a estrutura de dependência entre variáveis aleatórias. Vários artigos científicos, como os citados em [2] e [3], adotam a abordagem de pares de Cópulas para analisar dados de sinistros devido à sua capacidade de modelar estruturas de alta dependência dimensional.

Na dissertação começamos por apresentar a fundamentação teórica das Cópulas para suportar os métodos aplicados neste estudo. Acrescentamos uma aplicação com dados reais, a componente original da dissertação, utilizamos uma base de dados resultante da experiência de um seguro automóvel.

No ajustamento das Cópulas, consideramos a Cópula Gaussiana, a Cópula t-Student, a Cópula de Husler-Reiss, a Cópula de Gumbel e a Cópula de Frank. Para estimar os parâmetros, utilizamos o método de Inferência para Margens.

**Palavras-chave:** Cópulas, Dependência, Danos Próprios, Responsabilidade Civil

## ABSTRACT

All models are simplifications of reality and the risk industry, in particular the insurance industry, when assessing the financial impact of the realization of risk is obliged to introduce many simplifications, some explicitly in the form of assumptions.

One of the most common assumptions in an actuarial valuation is that of independence between some or all of the risks (events or coverages) that are being modeled.

The dependence between the frequency and amounts of compensation for various insured risks is recognized as one of the main factors that can influence the insurer's aggregate losses. Erroneously assuming this independence can result in significant unexpected losses.

In an effort to better understand the relationship of dependence between two coverages that are widely studied in the insurance industry, this dissertation is dedicated to the study of the dependency relationship of the value of claims in two coverages generally considered to be independent and present in automobile insurance: Own Damage and Third Party liability.

The mathematical concept used in the dissertation to estimate the relationship of dependence is the Copulas. Copulas have a probabilistic translation that we use to describe the dependency structure between random variables. Several scientific articles, such as those cited in [2] and [3], adopt the pair-Copulas approach to analyzing claims data due to its ability to model high-dimensional dependency structures.

In the dissertation, we begin by presenting the theoretical foundation of Copulas to support the methods applied in this study. We added an application with real data, the original component of the dissertation, using a database resulting from the experience of a car insurance company.

To adjust the Copulas, we considered the Gaussian Copula, the t-Student Copula, the Husler-Reiss Copula, the Gumbel Copula and the Frank Copula. To estimate the parameters, we used the Inference to Margins method.

**Keywords:** Copulas, Dependence, Own damage, Third Party liability

# ÍNDICE

<b>Índice de Figuras</b>	<b>viii</b>
<b>Índice de Tabelas</b>	<b>ix</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Revisão da Literatura</b>	<b>4</b>
2.1 Definições . . . . .	6
2.2 Limites de <i>Fréchet-Hoeffding</i> para Funções de Distribuição Conjuntas . . .	10
2.3 Conceitos de Dependência . . . . .	12
2.3.1 Dependência Perfeita . . . . .	12
2.3.2 Concordância . . . . .	13
2.3.3 Tau de Kendall e Rho de Spearman . . . . .	14
2.4 Família de Cópulas . . . . .	15
2.4.1 Cópulas Elípticas . . . . .	15
2.4.2 Cópulas Gaussianas . . . . .	16
2.4.3 Cópula-t . . . . .	17
2.4.4 Cópulas Arquimedianas . . . . .	17
2.5 Métodos de ajustamento de Cópulas . . . . .	20
2.5.1 Método da Máxima Verossimilhança . . . . .	20
2.5.2 Método dos momentos . . . . .	20
2.5.3 Inferência para margens . . . . .	20
2.5.4 Máxima Verossimilhança em Dois Passos . . . . .	21
<b>3 Ajustamento de Cópulas às indenizações de Responsabilidade Civil e Danos Próprios</b>	<b>22</b>
3.1 Dados . . . . .	22
3.1.1 Descrição da Base de Dados . . . . .	22
3.1.2 Análise Descritiva . . . . .	25
3.1.3 Distribuição das Margens . . . . .	27

3.1.4	Seleção de Cópulas - Conjunto $A$ . . . . .	31
3.1.5	Seleção de Cópulas - Conjunto $B$ . . . . .	33
3.1.6	Testes de Ajustamento e de Independência . . . . .	36
<b>4</b>	<b>Conclusão</b>	<b>38</b>
	<b>Bibliografia</b>	<b>39</b>
	<b>Anexos</b>	

## ÍNDICE DE FIGURAS

2.1	Representação Gráfica do Exemplo anterior . . . . .	7
3.1	Esquema do tratamento de Dados . . . . .	24
3.2	Boxplot dos custos dos sinistros de Responsabilidade Civil . . . . .	25
3.3	Boxplot dos custos de sinistros de Danos Próprios . . . . .	26
3.4	Histogramas dos montantes dos sinistros de Responsabilidade Civil e Danos Próprios . . . . .	27
3.5	Histogramas dos custos de sinistros Responsabilidade Civil e Danos Próprios limitados . . . . .	28
3.6	Histogramas dos montantes de sinistros Responsabilidade Civil e Danos Próprios das 107 observações . . . . .	29
3.7	Boxplot dos montantes de sinistros Responsabilidade Civil e Danos Próprios das 107 observações . . . . .	29
3.8	Diagrama de Pareto . . . . .	30
3.9	Coefficientes Cauda . . . . .	32
3.10	Caudas da distribuição da perda total (Responsabilidade Civil + Danos Próprios) . . . . .	33
3.11	Coefficientes Cauda . . . . .	34
3.12	Caudas da distribuição da perda total (Responsabilidade Civil + Danos Próprios) . . . . .	35
3.13	Comparação das Cópulas de Frank e Gaussiana com Cópula Independente do Conjunto A . . . . .	37
3.14	Comparação das Cópulas de Frank e Cópula Independente do Conjunto B . . . . .	37

## ÍNDICE DE TABELAS

3.1	Coberturas e número de contractos por cobertura . . . . .	23
3.2	Dados iniciais para o nº de Contrato 6.375 . . . . .	24
3.3	Dados Finais . . . . .	24
3.4	Estatísticas elementares da severidade dos montantes de sinistros de Respon- sabilidade Civil . . . . .	25
3.5	Estatísticas elementares da severidade dos montantes de sinistros de Danos Próprios . . . . .	26
3.6	Teste de Ajustamento à distribuição Gama . . . . .	27
3.7	Resultados dos testes de ajustamento para as coberturas RC e DP . . . . .	28
3.8	Resultados dos testes de ajustamento para as coberturas RC e DP . . . . .	30
3.9	Estimação dos parâmetros de cada Cópula . . . . .	31
3.10	Tau de Kendall e Rho de Spearman de cada Cópula . . . . .	32
3.11	Estimação dos parâmetros de cada Cópula . . . . .	34
3.12	Tau de Kendall e Rho de Spearman de cada Cópula . . . . .	35
3.13	Resultados dos testes de Ajustamento e de Independência . . . . .	36

## INTRODUÇÃO

O regime de Solvência II “requer que as empresas de seguros e resseguros da União Europeia disponham de bens suficientes”, que assegurem a solvência da empresa em função dos riscos assumidos. Além disso, “estabelece regras em matéria de governação, gestão de riscos, transparência e supervisão”, [4]. Com a necessidade de avaliação de riscos presentes na atividade seguradora, de forma separada, e de entender como estes se relacionam entre si, de forma a que possam ser agregados, o regime de Solvência II induz a necessidade de existência de modelos que permitam analisar e quantificar a dependência entre vários tipos de riscos. A análise da dependência de variáveis de risco tem uma importância extrema nestes modelos, uma vez que influencia a análise e a avaliação do capital de uma seguradora.

Existem diversas formas de incorporar a dependência entre v.a.’s, com diferentes tipos de complexidade. A utilização de medidas de dependência, como o coeficiente de correlação linear, medidas de concordância como também considerar independência entre as variáveis aleatórias são consideradas formas simples de incorporar a dependência. Por outro lado, com níveis de complexidade mais elevados, há a hipótese de utilizar distribuições conjuntas que têm subjacente uma estrutura de dependência específica.

As Cópulas são um conceito probabilístico usado para descrever a estrutura de dependência entre variáveis aleatórias. Estas funções, ver [5], agregam funções de distribuição multivariadas com a sua função de distribuição marginal unidimensional. De forma mais concreta, as cópulas são funções de distribuição multivariadas cujas margens unidimensionais têm distribuição uniforme no intervalo  $[0, 1]$ , ou seja, conectam as distribuições marginais de cada variável aleatória à sua distribuição conjunta, permitindo assim modelar a dependência entre elas.

A função distribuição das Cópulas está definida no hipercubo unitário, com margens uniformes. Esta propriedade possibilita a modelação da estrutura de dependência separada das distribuições marginais, o que é muito útil quando é necessário compreender a estrutura de dependência entre variáveis. Utilizando Cópulas como ferramenta de modelação, é possível construir diversos e variados modelos que permitem explicar alguns riscos específicos relacionados com a atividade seguradora, através da agregação de várias

Cópuas com diferentes distribuições marginais.

As Cópuas têm diversas funcionalidades, que permitem a modelação de vários tipos de dependência, incluindo dependência positiva, negativa ou assimétrica, e também permitem a incorporação de relações não lineares entre variáveis aleatórias.

Diversos autores publicaram um vasto número de artigos científicos sobre a teoria das Cópuas, tendo esta sido introduzida por Sklar em 1959, [6], contextualizado no tema de espaços métricos probabilísticos. Com a contribuição de Sklar, várias publicações como [7] e [8], permitiram detalhar e expandir o conceito de Cópuas. Para além destes, [9], [10] e [11] fornecem variadas aplicações práticas de Cópuas na área de Seguros e Finanças.

Ao nível do Atuariado, existem vários trabalhos relacionados com este tópico, como por exemplo a dissertação [12], onde a aplicação das Cópuas tem como objetivo ajustar a distribuição conjunta do número de sinistros de três grupos de risco homogéneos, sendo eles Responsabilidade Civil em danos materiais, Responsabilidade Civil para danos corporais e Danos Próprios no seguro automóvel, uma vez que, tal como é referido pela autora "as cópuas aparecem na literatura como uma boa ferramenta para ajustar a distribuição conjunta dos riscos de uma seguradora uma vez que permitem separar o comportamento marginal dos riscos da estrutura de dependência existente entre eles.". Neste trabalho, concluiu-se que o modelo multivariado para ajustar o número de sinistros entre os três grupos de risco deve ser a cópula de Nelsen com distribuições marginais Gama. Outro exemplo é o do artigo de *Gosh et al*, [3], onde são consideradas duas bases de dados diferentes, dados de sinistros de uma Seguradora Automóvel Australiana e dados de um Seguro Automóvel Sueco, com o objetivo de se ajustar diferentes famílias de Cópuas conhecidas, que permitem modelizar o risco relacionado com as coberturas típicas de uma apólice automóvel. Os autores procuram resultados que permitam fornecer informações importantes sobre a natureza da dependência e permitam satisfazer um dos objetivos primários das seguradoras visando avaliar o risco total de uma carteira agregada de perdas quando os componentes em análise, como por exemplo, o número de sinistros ou a população numa determinada área, são correlacionados.

Nesta dissertação, ir-se-á utilizar dados referentes a apólices de um Seguro Automóvel, tendo como o objetivo analisar se os montantes de sinistros pagos ao abrigo da cobertura de Responsabilidade Civil são independentes dos montantes pagos ao abrigo de algumas coberturas referentes a Danos Próprios. No caso de se verificar que estes montantes não são independentes, a dependência será modelada através de cópuas.

No Capítulo 2 é introduzida a origem da Teoria das Cópuas, uma breve explicação do conceito de Cópula, dos conceitos de dependência, exemplos clássicos de famílias de Cópuas, bem como os estudos mais recentes no âmbito da Matemática Atuarial.

No Capítulo 3 desenvolve-se a aplicação da Teoria das Cópuas na modelação dos riscos de Responsabilidade Civil e Danos Próprios no Seguro Automóvel, onde se procura analisar a dependência destas coberturas. Para tal, serão apresentados todos os tratamentos dos dados bem como os ajustamentos de distribuições para as margens quer ajustamentos das Cópuas.

---

O Capítulo 4 é dedicado às conclusões retiradas deste estudo.

## REVISÃO DA LITERATURA

Neste capítulo, seguindo as definições de [7] e com base nas referências [9] e [13], são apresentados e discutidos conceitos como dependência em termos de Cópulas, como também as várias famílias de Cópulas que irão ser aplicadas ao conjunto de dados em estudo nesta dissertação.

A designação “Cópula”, cujo significado em latim é “conexão”, “ligação”, remonta ao trabalho do matemático Abe Sklar em 1959, [6], onde o importante Teorema de Sklar foi primeiramente demonstrado. As Cópulas são ferramentas matemáticas utilizadas em diversas áreas científicas, nomeadamente Finanças e Atuariado, mais especificamente na gestão do risco, gestão de portfólios, otimização de portfólios e adequação de capital.

A maioria das aplicações de Cópulas na Estatística e Processos Estocásticos tem como base o Teorema de Sklar. O desenvolvimento da teoria de espaços métricos probabilísticos, entre o final de 1950 e o início de 1980, foi o principal motivo para o interesse acrescido no tema das Cópulas. O estudo das Cópulas foi bastante desenvolvido nessa época, tendo sido provados a maioria dos principais teoremas da teoria de Cópulas. A descoberta de novas aplicações na área das Finanças e da Estatística renovou o interesse do estudo das Cópulas no final dos anos 90, inclusive os trabalhos de Nelsen, em 1999, ver [7], e de Joe, em 1997, ver [8].

Como referido anteriormente, existem diversos trabalhos relacionados com este tema no âmbito da Matemática Atuarial que, de alguma forma, se relacionam com a investigação efetuada neste trabalho. Em [2], os autores colocam como objetivo testar a modelação da dependência bivariada entre perdas de diferentes linhas de negócio. Para além disso, neste trabalho os autores procuram estimar o risco de uma seguradora do Ramo Não-Vida usando medidas de risco como o *Value at Risk* (VaR) e o *Tail Value at Risk* (TVaR). Estas medidas de risco foram calculadas usando sinistros ocorridos e prémios em nove linhas de negócio, com base numa companhia de Seguros Não-Vida da Malásia. Numa primeira etapa, os autores procuram testar a modelação de dependência entre as perdas ajustando pares de Cópulas conhecidas, denominadas modelos de Cópula *C-Vine* e *D-Vine*. Para confirmar a escolha do modelo mais apropriado, computou-se os diversos critérios como o *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) nos modelos

---

ajustados. Além disso, este artigo também analisa o Loglik para cada modelo, concluindo-se, conforme os resultados, que a Cópula *D-Vine* tem uma estrutura de dependência mais alta e, conseqüentemente, é um modelo adequado para modelar a estrutura de dependência das perdas.

Numa segunda etapa, este artigo concentra-se na estimação de risco da companhia de seguros não-vida da Malásia. Este estudo concluiu que, enquanto que a medida de risco VaR revelou não ser uma medida coerente para este caso de estudo, o TVaR fornece uma melhor estimação sobre o risco da companhia de Seguros Não-Vida. Para além disso, visualizaram a uma tendência na variável de sinistros ocorridos, pelo que a linha de negócio de Seguro de Motores tem um valor mais alto de sinistros ocorridos do que os outros tipos de seguro.

Os autores de [14], utilizam Cópulas para entender qual é o tipo de relação entre os riscos de uma carteira de grupo. O problema que desencadeou este estudo foi motivado pela complexidade do *design* dos produtos de seguros modernos, que dificultam a modelação dos custos de seguros. Um deles é a estrutura multinível, maioritariamente encontrada nos seguros do ramo Não-Vida, onde um contrato contém um grupo de segurados, cada segurado tem vários tipos de coberturas, e o contrato é repetidamente observado ao longo do tempo. Na modelação de sinistros, espera-se capturar os efeitos de *cluster*, a dependência transversal entre vários tipos de sinistros, bem como a correlação em série no contexto longitudinal. Trabalhando com os dados de sinistros de seguros automóveis de Ontário, Canadá, os autores especificam três fontes de dependência: a correlação entre sinistros de vários veículos dentro do mesmo domicílio, a dependência transversal entre vários tipos de sinistros e a associação temporal para o custo de sinistros longitudinal de cada tipo. Essas relações explícitas e suas associações implícitas são especificadas através da aplicação da Cópula Gaussiana, onde retiram várias conclusões dos seus dados.

As Cópulas também foram utilizadas no artigo [3], onde foram utilizadas para modelar a estrutura de dependência de dados provenientes de Ramo de Seguros. Este estudo demonstra que certos tipos de Cópulas Arquimedianas com a propriedade de dependência de cauda pesada é uma estrutura aceitável para começar em termos de modelação de dados de sinistros, tanto no caso bivariado como no caso multivariado.

Também o trabalho [12] utiliza Cópulas para modelação, neste caso, do número de sinistros de três grupos de risco homogêneos do ramo Automóvel: Responsabilidade Civil (Danos Materiais), Responsabilidade Civil (Danos Corporais) e Danos Próprios. O objetivo é propor um modelo multivariado para ajustar o número de sinistros destes três riscos. A amostra utilizada é proveniente de uma carteira de uma seguradora portuguesa composta pelo número de sinistros mensais ocorridos entre 2000 e 2008. Os resultados finais do estudo revelam que o modelo multivariado para ajustar o número de sinistros entre os três grupos de risco deve ser a Cópula de Nelsen com distribuições marginais Gama.

Seguidamente, apresentar-se-ão as propriedades mais importantes da teoria da Cópulas para o caso bivariado, que pode ser estendido para uma dimensão superior. [1]

## 2.1 Definições

Antes de se definir o que é uma Cópula, são necessários alguns conceitos preliminares.

**Definição 2.1.1.** *Sejam  $S_1, \dots, S_n$  subconjuntos não vazios de  $\overline{\mathbb{R}}$ , onde  $\overline{\mathbb{R}}$  denota a linha real  $[-\infty, \infty]$ . Seja  $H$  uma função real de  $n$  variáveis tal que  $\text{Dom } H = S_1 \times \dots \times S_n$  e para  $a \leq b$  ( $a_k \leq b_k, \forall k$ ) seja  $B = [a, b] (= [a_1, b_1] \times \dots \times [a_n, b_n])$  uma  $n$ -caixa cujos vértices pertencem a  $\text{Dom } H$ . Os vértices de uma  $n$ -caixa  $B$  são os pontos  $c=(c_1, c_2, \dots, c_n)$ , onde cada  $c_k$  é igual ou a  $a_k$  ou a  $b_k$  para todo o  $k$ . Então o  $H$ -volume de  $B$  é dado por*

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

onde a soma é calculada sobre todos os vértices  $c$  de  $B$ , e  $\text{sgn}(c)$  é definido por

$$\text{sgn}(c) = \begin{cases} 1, & \text{se } c_k = a_k, \text{ para um número par de } k\text{'s,} \\ -1, & \text{se } c_k = a_k, \text{ para um número ímpar de } k\text{'s.} \end{cases}$$

Equivalentemente, o  $H$ -volume de uma  $n$ -caixa  $B=[a,b]$  é a diferença de  $n$ -ésima ordem de  $H$  em  $B$

$$V_H(B) = \Delta_a^b H(t) = \Delta_{a_n}^{b_n} \dots \Delta_{a_1}^{b_1} H(t),$$

onde as  $n$  diferenças de primeira ordem são definidas como

$$\Delta_{a_k}^{b_k} H(t) = H(t_1, \dots, t_{k-1}, b_k, t_{k+1}, \dots, t_n) - H(t_1, \dots, t_{k-1}, a_k, t_{k+1}, \dots, t_n).$$

Nesta dissertação, irá ser analisado o caso bivariado, pelo que a definição 2.1.1 reduz-se para a seguinte situação:

**Exemplo 2.1.1.** *Sejam  $S_1$  e  $S_2$  subconjuntos não vazios de  $\overline{\mathbb{R}}$ , e seja  $H$  uma função real de duas variáveis tal que  $\text{Dom } H = S_1 \times S_2$ . Seja  $B = [x_1, x_2] \times [y_1, y_2]$  um retângulo cujos vértices pertencem a  $\text{Dom } H$ . Então o  $H$ -volume de  $B$  é dado por*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \quad (2.1.1)$$

Note que se for definida a diferença de primeira ordem de  $H$  no retângulo  $B$  como

$$\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y) \quad e \quad \Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1),$$

então o  $H$ -volume do retângulo  $B$  é a diferença de segunda ordem de  $H$  em  $B$ ,

$$V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y).$$

*Demonstração.*

$$\begin{aligned} \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) &= \Delta_{x_1}^{x_2} \Delta_{y_1}^{y_2} H(x, y) \\ &= \Delta_{x_1}^{x_2} (H(x, y_2) - H(x, y_1)) \\ &= (H(x_2, y_2) - H(x_2, y_1)) - (H(x_1, y_2) - H(x_1, y_1)) \\ &= H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \end{aligned}$$

□

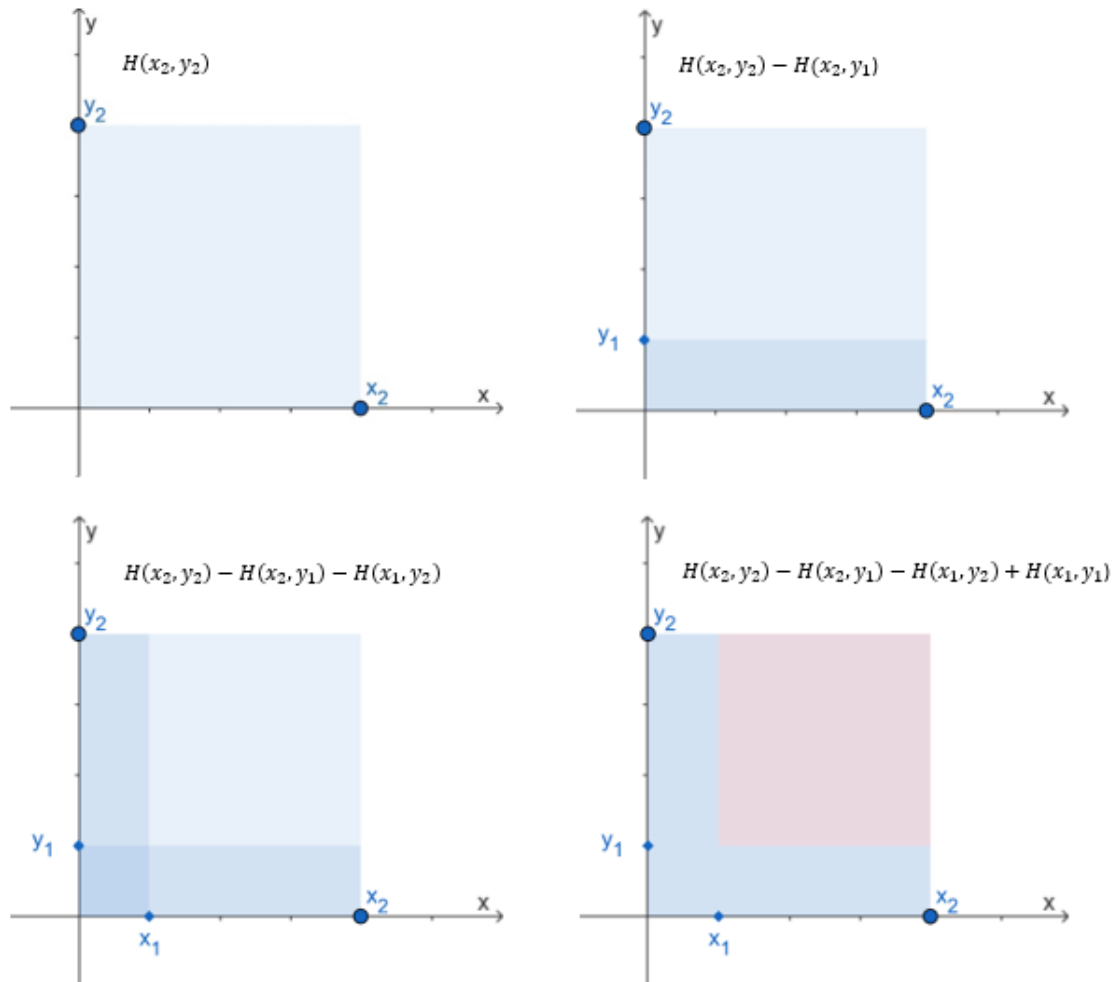


Figura 2.1: Representação Gráfica do Exemplo anterior

Na figura 2.1 estão representados graficamente as etapas no processo de cálculo de um volume a 2-dimensões. O  $H$ -volume é o volume contido dentro da caixa a vermelho.

**Definição 2.1.2.** Uma função real  $H$  de  $n$  variáveis diz-se  $n$ -crescente se  $V_H(B) \geq 0$ , para todas as  $n$ -caixas  $B$  cujos vértices se encontram no  $\text{Dom } H$ .

Suponha-se que o domínio de uma função real  $H$  de  $n$  variáveis é dado por  $\text{Dom } H = S_1 \times \dots \times S_n$ , onde cada  $S_k$  tem um elemento mais pequeno  $a_k$ . Diz-se que  $H$  é grounded se  $H(t) = 0, \forall t \in \text{Dom } H$ , tal que  $t_k = a_k$ , para pelo menos um  $k$ .

Se cada  $S_k$  é não vazio e tem um maior elemento  $b_k$ , então  $H$  tem margens e as margens de uma dimensão de  $H$  são funções  $H_k$  com  $\text{Dom } H_k = S_k$  e com  $H_k(x) = H(b_1, \dots, b_{k-1}, x, b_{k+1}, \dots, b_n), \forall x \in S_k$ .

As definições 2.1.1 e 2.1.2 permitem então a formalização da definição de Cópula. A seguinte definição pode ser encontrada em [7] e [9].

**Definição 2.1.3.** Uma Cópula é uma Função de Distribuição Conjunta multivariada no cubo unitário  $n$ -dimensional  $[0, 1]^n$  tal que

1. todas as distribuições marginais são uniformes no intervalo  $[0, 1]$ ;
2.  $C$  é grounded e  $n$ -crescente;
3.  $C$  tem margens  $C_k, k = 1, 2, \dots, n$ , que satisfazem  $C_k(u) = u, \forall u \in [0, 1]$ .

É possível observar que  $\forall n$ -Cópula,  $n \geq 3$ , cada margem  $k$ -dimensional de  $C$  é uma  $k$ -Cópula. Equivalentemente, uma  $n$ -Cópula é uma função  $C$  de  $[0, 1]^n$  para  $[0, 1]$  que segue as seguintes propriedades:

1.  $\forall u \in [0, 1]^n, C(u) = 0$  se pelo menos uma coordenada de  $u$  tem o valor 0, e  $C(u) = u_k$  se todas as coordenadas de  $u$  são iguais a 1 exceto  $u_k$ ;
2.  $\forall a, b \in [0, 1]^n$  tal que  $a_i \leq b_i$ , para todos os  $i, V_C([a, b]) \geq 0$

Como referido anteriormente, os dados que irão ser analisados são para o caso bivariado e, portanto, é um caso particular da definição anterior.

Com base em [5], tem-se a seguinte definição com algumas propriedades de Cópulas.

**Definição 2.1.4.** Uma Cópula é uma função  $C$  definida em  $[0, 1]^2 \rightarrow [0, 1]$ , que verifica as seguintes condições:

1.  $\forall u \in [0, 1], C(u, 0) = 0 = C(0, u)$ ;
2.  $\forall u, v \in [0, 1], C(u, 1) = u$  e  $C(1, v) = v$ ;
3.  $\forall (u_1, u_2), (v_1, v_2) \in [0, 1]^2$ , com  $u_1 \leq u_2, v_1 \leq v_2$ ,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (2.1.2)$$

**Observação 2.1.2.** O primeiro membro da desigualdade 2.1.2 denomina-se  $C$ -volume do retângulo  $B = [u_1, u_2] \times [v_1, v_2]$  e é representado por  $V_C(B)$ , como foi visto em 2.1.1.

O seguinte teorema apresentado é denominado de Teorema de Sklar. A maioria dos artigos e referências de Cópulas destacam a importância deste teorema, sendo muitas vezes considerado o mais importante em relação a resultado de Cópulas.

O teorema e demonstração podem ser vistos em [5] e [13],

**Teorema 2.1.1. (Teorema de Sklar)** *Seja  $H$  uma função de distribuição  $n$ -dimensional com margens  $F_1, \dots, F_n$ . Então existe uma  $n$ -Cópula  $C$  tal que, para todo o  $x \in \overline{\mathbb{R}}^n$*

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (2.1.3)$$

Da equação (2.1.3), retira-se que a Cópula  $C$  é uma função que relaciona a função distribuição conjunta e as suas marginais.

O Lema 2.1.3 e a Propriedade 2.1.4, que podem ser consultados em [13], irão permitir a demonstração do Teorema 2.1.1 para duas variáveis.

**Lema 2.1.3.** *Dada uma função distribuição  $F_X$ , define-se a função inversa  $F_X^{(-1)}$  e  $F_X^{(-1)+}$  de  $F_X$  como*

$$F_X^{(-1)}(p) = \inf \{x \in \mathbb{R} \mid F_X(x) \geq p\} = \sup \{x \in \mathbb{R} \mid F_X(x) < p\}, \quad (2.1.4)$$

e

$$F_X^{(-1)+}(p) = \inf \{x \in \mathbb{R} \mid F_X(x) > p\} = \sup \{x \in \mathbb{R} \mid F_X(x) \leq p\}, \quad (2.1.5)$$

para  $p \in [0, 1]$ , onde, por convenção,  $\inf = +\infty$  e  $\sup = -\infty$ .

Para qualquer  $x \in \mathbb{R}$  e probabilidade  $p$ , verifica-se a seguinte equivalência:

$$F_X^{(-1)}(p) \leq x \Leftrightarrow p \leq F_X(x).$$

*Demonstração.* A implicação  $\Rightarrow$  é demonstrada se for possível mostrar que

$$p > F_X(x) \Rightarrow x < F_X^{(-1)}(p).$$

Assuma-se que  $p > F_X(x)$ . Então existe um  $\epsilon > 0$  tal que  $p > F_X(x+\epsilon)$ . De (2.1.5), obtém-se  $x+\epsilon \leq F_X^{(-1)}(p)$  que, por sua vez, implica que  $x < F_X^{(-1)}(p)$ .

Agora prova-se a implicação  $\Leftarrow$ . Se  $p \leq F_X(x)$  então tem-se que  $p \leq F_X(x+\epsilon)$ , para todo  $\epsilon > 0$ . De 2.1.4, conclui-se que  $F_X^{(-1)}(p) \leq x+\epsilon$ , para todo  $\epsilon > 0$ .

Retirando o limite para  $\epsilon \downarrow 0$ , obtém-se  $F_X^{(-1)}(p) \leq x$ . □

**Propriedade 2.1.4.** *Se uma variável aleatória  $X$  tem uma função distribuição contínua  $F_X$ , então  $F_X(X) \sim \mathbb{U}(0, 1)$ .*

*Demonstração.* Esta prova é seguimento do Lema 2.1.3, que assegura que para todo  $0 < u < 1$ ,

$$\Pr[F_X(X) \geq u] = \Pr\left[X \geq F_X^{(-1)}(u)\right] = \bar{F}_X\left(F_X^{(-1)}(u)\right) = 1 - u,$$

do qual se conclui que  $F_X(X) \sim \mathbb{U}(0, 1)$ . □

Posto isto, torna-se possível demonstrar o Teorema de *Sklar* para duas variáveis, que se pode ser consultado em [13].

**Teorema 2.1.2.** *Seja  $F_X \in \mathbb{R}_2(F_1, F_2)$  com margens com funções distribuição  $F_1$  e  $F_2$ . Então existe uma única Cópula  $C$  tal que, para todo  $x \in \mathbb{R}^2$ ,*

$$F_X(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \quad (2.1.6)$$

Por outro lado, se  $C$  é uma Cópula e  $F_1$  e  $F_2$  são funções de distribuição, então a função  $F_X$  definida em (2.1.6) é uma função distribuição bivariada com margens  $F_1$  e  $F_2$ .

*Demonstração.* Uma vez que  $F_i$  são contínuas, a propriedade 2.1.4 garante que ambas  $F_1(X_1)$  e  $F_2(X_2)$  são  $\mathbb{U}(0, 1)$ .

Seja  $C$  uma função distribuição conjunta para  $(F_1(X_1), F_2(X_2))$ , isto é,

$$\begin{aligned} C(u_1, u_2) &= \Pr[F_1(X_1) \leq u_1, F_2(X_2) \leq u_2] \\ &= \Pr[X_1 \leq F_1^{(-1)}(u_1), X_2 \leq F_2^{(-1)}(u_2)] \\ &= F_X(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2)) \end{aligned}$$

onde o Lema 2.1.3 foi utilizado. Então verifica-se a representação (2.1.6), uma vez que

$$\begin{aligned} F_X(x_1, x_2) &= \Pr[X_1 \leq x_1, X_2 \leq x_2] \\ &= \Pr[F_1(X_1) \leq F_1(x_1), F_2(X_2) \leq F_2(x_2)] \\ &= C(F_1(x_1), F_2(x_2)). \end{aligned}$$

□

**Corolário 2.1.1.** *Seja  $H$  uma função de distribuição  $n$ -dimensional com margens contínuas  $F_1, \dots, F_n$  e Cópula  $C$  (onde  $C$  satisfaz (2.1.3)). Então para qualquer  $u \in [0, 1]^n$ ,*

$$C(u_1, \dots, u_n) = H(F_1^{(-1)}(u_1), \dots, F_n^{(-1)}(u_n)).$$

O Corolário 2.1.1 afirma ser possível estudar a dependência entre variáveis sem fixar as suas distribuições marginais.

## 2.2 Limites de *Fréchet-Hoeffding* para Funções de Distribuição Conjuntas

Nesta secção são introduzidos alguns conceitos e propriedades importantes para a análise de dependência de Cópulas.

Com base nas referências [7] e [9], e com mais informação sobre estes conceitos, incluindo interpretações geométricas, na referência [13], começa-se por introduzir as seguintes funções de Cópulas.

Considerem-se as funções  $M^n$ ,  $\Pi^n$  e  $W^n \in [0, 1]^n$  definidas como:

$$\begin{aligned} M^n(u) &= \min(u_1, \dots, u_n) \\ \Pi^n(u) &= u_1 \dots u_n \\ W^n(u) &= \max(u_1 + \dots + u_n - n + 1, 0). \end{aligned}$$

As funções  $M^n$  e  $\Pi^n$  são  $n$ -Cópulas,  $\forall n \geq 2$ , enquanto que a função  $W^n$  não é uma Cópula, para qualquer  $n \geq 3$ .

O seguinte exemplo demonstra que  $\Pi^n(u, v) = uv$  é uma Cópula, com  $u, v \in [0, 1]$ . Para tal, irão ser estabelecidas a validade das condições da Definição 2.1.4:

**Exemplo 2.2.1.** Para  $\forall u, v \in [0, 1]$  e  $n = 2$ ,

1.  $\Pi(u, 0) = u \times 0 = 0 = \Pi(0, u)$ ;
2.  $\Pi(u, 1) = u \times 1 = u$  e  $\Pi(1, v) = 1 \times v = v$ ;
3.  $\forall (u_1, u_2), (v_1, v_2) \in [0, 1]^2$ , com  $u_1 \leq u_2, v_1 \leq v_2$

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

$$\Leftrightarrow u_2 v_2 - u_2 v_1 - u_1 v_2 + u_1 v_1 \geq 0$$

$$\Leftrightarrow (u_2 - u_1)(v_2 - v_1) \geq 0 \Leftrightarrow V_{\Pi}(u, v) \geq 0$$

Portanto,  $\Pi^2(u)$  é uma Cópula para  $n = 2$ .

O seguinte teorema é denominado de Limites de Fréchet-Hoeffding.

**Teorema 2.2.1.** Se  $C$  é qualquer  $n$ -Cópula, então para todos  $u \in [0, 1]^n$ ,

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

A demonstração pode ser consultada em [5] p. 11.

Apesar do limite inferior de Fréchet-Hoeffding,  $W^n$ , nunca ser uma Cópula para  $n \geq 3$ , no seguinte sentido, ela revela ser o melhor limite inferior possível.

**Teorema 2.2.2.** Para qualquer  $n \geq 3$  e qualquer  $u \in [0, 1]^n$ , existe uma  $n$ -Cópula  $C$  (que depende de  $u$ ) tal que

$$C(u) = W^n(u).$$

Para consultar a demonstração, ver [5] p.48.

Denota-se por  $\bar{C}$  a função de sobrevivência para  $n$  variáveis aleatórias com função de distribuição conjunta  $C$ , isto é, se  $(U_1, \dots, U_n)^T$  tem função de distribuição  $C$ , então  $\bar{C}(u_1, \dots, u_n) = \mathbb{P}\{U_1 > u_1, \dots, U_n > u_n\}$ .

A seguinte definição pode ser consultada em [5].

**Definição 2.2.1.** Se  $C_1$  e  $C_2$  são Cópulas,  $C_1$  é menor que  $C_2$  (com notação  $C_1 < C_2$ ) se

$$C_1(u) \leq C_2(u) \quad e \quad \bar{C}_1(u) \leq \bar{C}_2(u),$$

para todo  $u \in [0, 1]^n$ .

É de notar que, no caso bivariado,

$$\begin{aligned} \bar{C}_1(u_1, u_2) \leq \bar{C}_2(u_1, u_2) &\Leftrightarrow 1 - u_1 - u_2 + C_1(u_1, u_2) \leq 1 - u_1 - u_2 + C_2(u_1, u_2) \\ &\Leftrightarrow C_1(u_1, u_2) \leq C_2(u_1, u_2). \end{aligned}$$

A Cópula do limite inferior de *Fréchet-Hoeffding*,  $W$ , é menor do que qualquer Cópula, e a Cópula do limite superior de *Fréchet-Hoeffding*,  $M$ , é maior do que qualquer Cópula.

Posto isto, é possível voltar a aplicar o Teorema de Sklar, agora em termos de variáveis aleatórias e as suas respectivas funções de distribuição.

Sejam  $X_1, \dots, X_n$  variáveis com funções de distribuição contínua  $F_1, \dots, F_n$ , respetivamente, e função de distribuição conjunta  $H$ . Então  $(X_1, \dots, X_n)^T$  tem uma Cópula única  $C$ , onde  $C$  é dado por (2.1.3). A representação da Cópula padrão da distribuição do vetor aleatório  $(X_1, \dots, X_n)^T$  torna-se:

$$H(x_1, \dots, x_n) = \mathbb{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\} = C(F_1(x_1), \dots, F_n(x_n)).$$

O Teorema 2.2.3 tem uma importância enorme para esta dissertação, visto que permitirá tirar conclusões sobre os dados obtidos.

**Teorema 2.2.3.** Seja  $(X_1, \dots, X_n)^T$  um vetor de variáveis aleatórias contínuas com Cópula  $C$ , então  $X_1, \dots, X_n$  são independentes se e só se  $C = \Pi^n$ .

## 2.3 Conceitos de Dependência

O conceito de Cópulas, atrás apresentado, permite o estudo da dependência entre variáveis aleatórias de um modo mais natural.

Uma das formas mais usuais de medir a dependência é a Correlação Linear, também denominada de Correlação de Pearson. No entanto, não é muito utilizado para a aplicação de Cópulas, uma vez que existem ferramentas mais apropriadas para o cálculo da dependência das Cópulas. Abaixo serão apresentadas algumas propriedades da Correlação Linear e serão introduzidas algumas medidas de dependência baseadas em Cópulas.

As definições e os resultados apresentados neste capítulo são essencialmente baseados em [9], [13] e [15].

### 2.3.1 Dependência Perfeita

Para todas as  $n$  Cópulas  $C$  sabe-se, pela desigualdade de *Fréchet-Hoeffding*, ver Teorema 2.2.1, que

$$W^n(u_1, \dots, u_n) \leq C(u_1, \dots, u_n) \leq M^n(u_1, \dots, u_n).$$

Para além disso, como referido anteriormente na secção 2.2, para  $n=2$ , os limites superior e inferior são eles próprios Cópulas. Sabe-se que  $W$  e  $M$  são funções de distribuição bivariadas de vetores aleatórios  $(U, 1 - U)$  e  $(U, U)$ , respetivamente, onde  $U \sim U(0, 1)$  ( $U(0, 1)$  representa a distribuição uniforme contínua no intervalo unitário  $[0, 1]$ ). Neste caso, diz-se que  $W$  descreve uma dependência negativa perfeita e  $M$  descreve uma dependência positiva perfeita.

**Definição 2.3.1.** *Se  $(X, Y)$  tem a Cópula  $M$ , então  $X$  e  $Y$  são Comonotónicos. Se tem a Cópula  $W$ , então são denominadas como contracomonotónicos.*

Para mais detalhes desta definição, ver [16].

### 2.3.2 Concordância

Sejam  $(x, y)^T$  e  $(\tilde{x}, \tilde{y})^T$  duas observações do vetor  $(X, Y)^T$  de uma variável aleatória contínua. Então  $(x, y)^T$  e  $(\tilde{x}, \tilde{y})^T$  são ditas concordantes se  $(x - \tilde{x})(y - \tilde{y}) > 0$  e discordantes se  $(x - \tilde{x})(y - \tilde{y}) < 0$ .

O seguinte teorema, e respetiva demonstração, encontram-se em [9], p.11, e introduz uma função que surge do conceito de concordância, designada de função  $Q$ , que terá uma grande importância na próxima secção. Esta função  $Q$  calcula a diferença da probabilidade de pares concordantes e discordantes entre dois vetores de variáveis aleatórias contínuas com distribuições conjuntas distintas, porém com mesmas marginais dependendo apenas das suas funções Cópulas.

**Teorema 2.3.1.** *Sejam  $(X, Y)^T$  e  $(\tilde{X}, \tilde{Y})^T$  vetores independentes de variáveis aleatórias contínuas com funções de distribuição conjunta  $H$  e  $\tilde{H}$ , respetivamente, com margens comuns  $F$  (de  $X$  e  $\tilde{X}$ ) e  $G$  (de  $Y$  e  $\tilde{Y}$ ). Sejam também  $C$  e  $\tilde{C}$  Cópulas de  $(X, Y)^T$  e  $(\tilde{X}, \tilde{Y})^T$ , respetivamente, tal que  $H(x, y) = C(F(x), G(y))$  e  $\tilde{H} = \tilde{C}(F(x), G(y))$ . Por fim, seja  $Q$  a diferença entre a probabilidade de concordância e discordância de  $(X, Y)^T$  e  $(\tilde{X}, \tilde{Y})^T$ , isto é, seja*

$$Q = \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0] - \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) < 0\}.$$

Então

$$Q = Q(C, \tilde{C}) = 4 \iint_{[0,1]^2} \tilde{C}(u, v) dC(u, v) - 1.$$

A Definição 2.3.2 pode ser encontrada em [5], onde são expostas as propriedades das medidas de concordância.

**Definição 2.3.2.** *Uma medida de valor real  $\kappa$  da dependência entre duas variáveis aleatórias contínuas  $X$  e  $Y$ , cuja Cópula é  $C$ , é uma medida de concordância se satisfaz as seguintes propriedades:*

1.  $\kappa$  é definido para todos os pares  $(X, Y)$  das variáveis aleatórias contínuas;
2.  $-1 \leq \kappa_{X,X} \leq 1$ ,  $\kappa_{X,X} = 1$  e  $\kappa_{X,-X} = -1$ ;
3.  $\kappa_{X,Y} = \kappa_{Y,X}$ ;

4. Se  $X$  e  $Y$  são independentes, então  $\kappa_{X,Y} = K_{\Pi} = 0$ ;
5.  $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$ ;
6. Se  $C$  e  $\tilde{C}$  são Cópulas tal que  $C < \tilde{C}$ , então  $\kappa_C \leq \kappa_{\tilde{C}}$ ;
7. Se  $\{(X_n, Y_n)\}$  é uma sequência de variáveis aleatórias contínuas com Cópulas  $C_n$ , e se  $\{C_n\}$  converge pontualmente para  $C$ , então  $\lim_{n \rightarrow \infty} \kappa_{C_n} = \kappa_C$ .

A definição 2.3.2 permitirá tirar conclusões sobre as duas medidas de concordância da próxima secção.

### 2.3.3 Tau de Kendall e Rho de Spearman

O Tau de Kendall e o Rho de Spearman são duas medidas de dependência (concordância) importantes. Estes fornecem talvez as melhores alternativas para o coeficiente de correlação linear como uma medida de dependência para distribuições não elípticas.

**Definição 2.3.3.** *Sejam  $(X_1, \dots, X_n)$  variáveis aleatórias independentes.*

*Diz-se que  $X'_i$  é uma cópia independente de  $X_i$  quando a distribuição de  $X'_i$  é igual à distribuição de  $X_i$ . Para além disso,  $X'_i$  e  $X_i$  são independentes, para  $i \in \mathbb{R}$ .*

**Definição 2.3.4.** *O Tau de Kendall para o vetor aleatório  $(X, Y)^T$  é definido como*

$$\tau(X, Y) = \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) > 0\} - \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) < 0\},$$

onde  $(\tilde{X}, \tilde{Y})^T$  é uma cópia independente de  $(X, Y)^T$ .

O Tau de Kendall para  $(X, Y)^T$  corresponde à diferença entre a probabilidade de concordância e a probabilidade de discordância.

**Teorema 2.3.2.** *Seja  $(X, Y)^T$  um vetor de variáveis aleatórias contínuas com Cópula  $C$ . Então o Tau de Kendall para  $(X, Y)^T$  é dado por*

$$\tau(X, Y) = Q(C, C) = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1.$$

Note-se que o integral acima é o valor esperado da variável aleatória  $C(U, V)$ , onde  $U, V \sim U(0, 1)$  com função distribuição conjunta  $C$ , isto é,  $\tau(X, Y) = 4\mathbb{E}(C(U, V)) - 1$ .

Para a demonstração, ver [15] p. 208.

**Definição 2.3.5.** *O Rho de Spearman para o vetor aleatório  $(X, Y)^T$  é definido por*

$$\rho_s(X, Y) = 3 [\mathbb{P}[(X - \tilde{X})(Y - Y') > 0] - \mathbb{P}[(X - \tilde{X})(Y - Y') < 0]],$$

onde  $(X, Y)^T$ ,  $(\tilde{X}, \tilde{Y})^T$  e  $(X', Y')$  são cópias independentes.

Note-se que  $\tilde{X}$  e  $Y'$  são independentes. Usando o Teorema 2.3.1, obtemos o seguinte resultado:

**Teorema 2.3.3.** *Seja  $(X, Y)^T$  um vetor de variáveis aleatórias contínuas com Cópula  $C$ . Então o Rho de Spearman para  $(X, Y)^T$  é dado por*

$$\rho_S(X, Y) = 3Q(C, \Pi) = 12 \iint_{[0,1]^2} uv dC(u, v) - 3 = 12 \iint_{[0,1]^2} C(u, v) dudv - 3.$$

Portanto, se  $X \sim F$  e  $Y \sim G$ , e tem-se  $U = F(X)$  e  $V = G(Y)$ , então

$$\begin{aligned} \rho_S(X, Y) &= 12 \iint_{[0,1]^2} uv dC(u, v) - 3 = 12\mathbb{E}(UV) - 3 \\ &= \frac{\mathbb{E}(UV) - \frac{1}{4}}{\frac{1}{12}} = \frac{Cov(U, V)}{\sqrt{Var(U)}\sqrt{Var(V)}} \\ &= \rho(F(X), G(Y)). \end{aligned}$$

**Teorema 2.3.4.** *Se  $X$  e  $Y$  são variáveis aleatórias contínuas cuja Cópula é  $C$ , então o Tau de Kendall e o Rho de Spearman satisfazem as propriedades na definição (2.3.2) para a medida de concordância.*

A demonstração do Teorema pode ser consultada em [7], p. 137.

## 2.4 Família de Cópulas

### 2.4.1 Cópulas Elípticas

Para se definir o conceito de Cópulas Elípticas, começar-se-á por introduzir a definição de distribuições Elípticas. Uma distribuição elíptica, que constitui uma generalização da lei Normal multivariada, é uma distribuição de probabilidade multivariada em que a distribuição de cada variável aleatória é especificada por um parâmetro comum denominado de *forma* e a estrutura de dependência entre as variáveis é descrita por uma matriz de correlação. Exemplos de distribuições Elípticas incluem as distribuições Normal e de Student-t multivariadas.

As Cópulas Elípticas são uma classe de Cópulas baseadas em distribuições elípticas. Estas são úteis na modelação de variáveis aleatórias dependentes quando a distribuição conjunta das variáveis não é conhecida, mas as distribuições marginais e a matriz de correlação estão disponíveis. Possuem várias propriedades, como simetria, independência da cauda e a capacidade de capturar dependência positiva e negativa.

As Cópulas Elípticas mais utilizadas são as Cópulas Gaussianas, a t-Cópula e a Cópula de Clayton.

As seguintes definições e teoremas têm como base [5], [8], [9] e [17].

**Definição 2.4.1.** *Se  $\mathbf{X}$  é um vetor aleatório  $n$ -dimensional, para algum  $\mu \in \mathbb{R}^n$  e existir alguma matriz simétrica  $n \times n$  definida não negativa,  $\Sigma$ , então a função característica  $\varphi_{\mathbf{X}-\mu}(\mathbf{t})$  de  $\mathbf{X} - \mu$  é uma função da forma quadrática  $\mathbf{t}^T \Sigma \mathbf{t}$ ,  $\varphi_{\mathbf{X}-\mu}(\mathbf{t}) = \phi(\mathbf{t}^T \Sigma \mathbf{t})$ , diz-se que  $\mathbf{X}$  tem uma distribuição Elíptica com parâmetros  $\mu$ ,  $\Sigma$  e  $\phi$ , e tem a notação  $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$ .*

Quando  $n=1$ , a classe das distribuições elípticas coincide com a classe de distribuições simétricas unidimensionais. A função  $\phi$  é denominada de gerador característico.

A definição 2.4.1 estabelece a expressão de distribuição Elíptica, que é utilizada no teorema seguinte.

**Teorema 2.4.1.**  $X \sim E_n(\mu, \Sigma, \phi)$  com  $\text{rank}(\Sigma) = k$  se e só se existir uma variável aleatória  $R$  não negativa e  $\mathbf{U}$ , um vetor aleatório  $k$ -dimensional uniformemente distribuído no conjunto  $\{\mathbf{z} \in \mathbb{R}^k \mid \mathbf{z}^T \mathbf{z} = 1\}$ , independente de  $R$  e uma matriz real  $\mathbf{A}$ , de dimensões  $n \times k$ , com  $\Sigma = \mathbf{A}\mathbf{A}^T$ , tal que

$$\mathbf{X} =_d \mu + R\mathbf{A}\mathbf{U}$$

Para a demonstração, ver [18], p.369.

## 2.4.2 Cópulas Gaussianas

Uma Cópula gaussiana é uma Cópula que é resultante de uma distribuição conjunta de variáveis aleatórias com distribuição Normal e, por consequência, é usualmente utilizada para modelar a dependência entre duas ou mais variáveis que seguem essa distribuição.

Duas das principais vantagens destas Cópulas são a sua facilidade de trabalhar matematicamente e a detenção de propriedades como a sua simetria, que as tornam úteis e potenciais candidatas para utilização em aplicações práticas.

No entanto, é importante notar que a Cópula gaussiana nem sempre é a melhor escolha para modelação de dependência, particularmente em situações onde as distribuições subjacentes são não gaussianas ou exibem caudas pesadas.

De forma geral, as Cópulas gaussianas são ferramentas eficazes para modelar a estrutura de dependência entre variáveis em muitas aplicações, mas é importante considerar cuidadosamente as suposições e limitações do modelo ao usá-las.

Segundo [9], a Cópula da distribuição Normal  $n$ -variada, com matriz de correlação linear  $\mathbf{R}$  é definida por

$$C_{\mathbf{R}}^{Ga}(u) = \Phi_{\mathbf{R}}^n \left( \Phi^{(-1)}(u_1), \dots, \Phi^{(-1)}(u_n) \right), \quad (2.4.1)$$

onde  $\Phi_{\mathbf{R}}^n$  denota a função de distribuição conjunta da função de distribuição Normal reduzida  $n$ -variada com matriz de correlação linear  $\mathbf{R}$ , e  $\Phi^{(-1)}$  denota a inversa da função de distribuição da distribuição Normal reduzida univariada. Cópulas com a forma da expressão (2.4.1) são denominadas Cópulas Gaussianas. No caso bivariado, a expressão da Cópula pode ser escrita como

$$C_{\mathbf{R}}^{Ga}(u, v) = \int_{-\infty}^{\Phi^{(-1)}(u)} \int_{-\infty}^{\Phi^{(-1)}(v)} \frac{1}{2\pi(1-R_{12}^2)^{\frac{1}{2}}} \exp \left\{ -\frac{s^2 - 2\mathbf{R}_{12}st + t^2}{2(1-R_{12}^2)} \right\} ds dt,$$

com  $R_{12}$  o usual coeficiente de correlação da distribuição Normal bivariada.

### 2.4.3 Cópula-t

Uma Cópula-t é uma Cópula que é resultante de uma distribuição conjunta multivariada *t-student*. São normalmente usadas para modelar a dependência entre duas ou mais variáveis aleatórias que seguem uma distribuição *t-student*. Este tipo de Cópulas é particularmente útil porque permite a modelação de estruturas de dependência que apresentam caudas mais pesadas do que as Cópulas Gaussianas, o que pode ser útil em determinadas situações em que eventos extremos são mais prováveis.

Isto é particularmente importante em aplicações financeiras e de gestão de risco, onde eventos extremos podem ter um impacto significativo no risco global de um portfólio ou investimento. No entanto, é importante observar que a Cópula-t pode ser mais difícil de trabalhar matematicamente do que a Cópula Gaussiana devido à sua complexidade.

Segundo [19], sejam  $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$  e  $R = \frac{\sqrt{v}}{\sqrt{S}}$ , com  $S \sim \chi_v^2$  (Qui-Quadrado com  $v$  graus de liberdade), independentes. Então o vetor aleatório  $\mathbf{Y}$  com valores  $\mathbb{R}^d$  dado por

$$\mathbf{Y} = R\mathbf{Z}, \quad (2.4.2)$$

tem uma distribuição *t-student* centrada com  $v$  graus de liberdade. É de notar que para  $v > 2$ ,  $\text{Cov}(\mathbf{Y}) = \frac{v}{v-2}\Sigma$ .

Pelo Teorema de Sklar, Teorema 2.1.1, a Cópula de  $\mathbf{Y}$ , definida em (2.4.2), pode ser escrita como

$$C_{v,\rho}^t(u) = t_{v,\rho}^n \left( t_v^{(-1)}(u_1), \dots, t_v^{(-1)}(u_n) \right),$$

onde  $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$  para  $i, j \in \{1, \dots, n\}$  e onde  $t_{v,\rho}^n$  denota a função distribuição de  $\frac{\sqrt{v}}{\sqrt{S}}\mathbf{Z}$ , onde  $S \sim \chi_v^2$  e  $\mathbf{Z} \sim \mathcal{N}_n(0, \rho)$  são independentes. Aqui  $t_v$  denota a função de distribuição da margem de  $t_{v,\rho}^n$ .

No caso bivariado, a expressão da Cópula pode ser escrita como

$$C_{v,\rho}^t(u, v) = \int_{-\infty}^{t_v^{(-1)}(u)} \int_{-\infty}^{t_v^{(-1)}(v)} \frac{1}{2\pi(1-\rho_{12}^2)^{\frac{1}{2}}} \exp \left\{ 1 + \frac{s^2 - 2\rho_{12}st + t^2}{v(1-\rho_{12}^2)} \right\}^{-\frac{v+2}{2}} ds dt.$$

É de notar que  $\rho_{12}$  é simplesmente o usual coeficiente de correlação da correspondente  $t_v$ -distribuição bivariada se  $v > 2$ .

### 2.4.4 Cópulas Arquimedianas

De acordo com [9], utilizando o Teorema de Sklar, Teorema (2.1.1), foi possível até agora analisar diferentes tipos de famílias de Cópulas, que foram derivadas de certas famílias de funções de distribuição multivariadas. Como foi visto na secção 2.4.1, as Cópulas Elípticas são simplesmente as funções de distribuição de componentes de vetores aleatórios distribuídos de forma elíptica.

As Cópulas Elípticas apresentam algumas restrições que as tornam, por vezes, desadequadas, tais como não terem expressões de forma fechada e estarem restritas a terem simetria radial ( $C = \hat{C}$ ). Em muitas aplicações da área Financeira e Actuarial, existe uma grande possibilidade de haver uma dependência mais forte entre grandes perdas (por exemplo, geradas por uma quebra do mercado de ações) do que entre grandes ganhos. Tais assimetrias não podem ser modeladas com Cópulas Elípticas.

Nesta secção, discutir-se-á uma importante classe de Cópulas designadas de Cópulas Arquimedianas. Muitas famílias paramétricas interessantes de Cópulas são Arquimedianas e esta classe permite uma grande variedade de diferentes estruturas de dependência. Além disso, em contraste com as Cópulas Elípticas, todas as Cópulas Arquimedianas estudadas têm expressões de forma fechada. Contrariamente às Cópulas discutidas até agora, a origem destas Cópulas não está associada a funções de distribuição multivariadas com base no Teorema de Sklar. Uma consequência disso é que são necessárias algumas condições técnicas para afirmar que extensões multivariadas de 2-Cópulas Arquimedianas são  $n$ -Cópulas adequadas. Outra desvantagem é que as extensões multivariadas de Cópulas Arquimedianas, em geral, sofrem de falta de escolha livre de parâmetros no sentido de que algumas das entradas na matriz de correlação resultante são forçadas a serem iguais.

Estas Cópulas são construídas com base numa função auxiliar  $\varphi : [0, 1] \rightarrow [0, +\infty]$  que verifica as seguintes condições:

1. É contínua;
2. É estritamente decrescente;
3.  $\varphi(1) = 0$ .

Ir-se-á incluir nesta classe as funções definidas em  $]0, 1]$  que verificam  $\lim_{x \rightarrow 0^+} \varphi(x) = +\infty$ , convencendo-se que  $\varphi(0) = +\infty$  e, neste caso, considerar-se-á que o contradomínio de  $\varphi$  é  $[0, +\infty]$ .

Em relação a esta família, as três Cópulas mais utilizadas são a Cópula de Gumbel, a Cópula de Frank e Cópula de Clayton, sendo que nesta dissertação serão apenas as duas primeiras Cópulas desta família utilizadas.

Antes de definir estes tipos de Cópulas, introduzem-se algumas definições que serão úteis na exposição a seguir.

**Definição 2.4.2.** *Seja  $\varphi$  uma função que satisfaz as condições 1, 2 e 3 atrás mencionadas. Denomina-se pseudo-inversa de  $\varphi$  à função  $\varphi^{[-1]} : [0, +\infty] \rightarrow [0, 1]$  definida por*

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

onde  $\varphi^{-1}$  denota a função inversa de  $\varphi$ .

É fácil verificar que  $\varphi^{[-1]}$  é uma função decrescente e contínua em  $[0, +\infty]$ , sendo estritamente decrescente em  $[0, \varphi(0)]$ . Para além disso,  $\varphi^{[-1]}(\varphi(u)) = u$  em  $[0, 1]$  e

$$\varphi\left(\varphi^{[-1]}(t)\right) = \begin{cases} t & , 0 \leq t \leq \varphi(0) \\ \varphi(0) & , \varphi(0) \leq t \leq +\infty. \end{cases}$$

Finalmente, se  $\varphi(0) = +\infty$ , então  $\varphi^{[-1]} = \varphi^{-1}$ .

Através da definição 2.4.2, torna-se possível a formalização da expressão das Cópulas Arquimedianas.

**Teorema 2.4.2.** *Seja  $\varphi$  uma função que verifica as condições acima referidas e  $\varphi^{[-1]}$  a sua pseudo-inversa. A função  $C: [0, 1]^2 \rightarrow [0, 1]$  definida por*

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (2.4.3)$$

é uma Cópula se e só se  $\varphi$  é convexa.

Cópulas da forma (2.4.3) são designadas Cópulas Arquimedianas. A função  $\varphi$  corresponde ao gerador da Cópula. Se  $\varphi(0) = \infty$ ,  $\varphi$  diz-se gerador estrito. Neste caso,  $\varphi^{[-1]} = \varphi^{-1}$  e  $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$  é dito ser uma Cópula Arquimediana estrita.

Os exemplos seguintes são casos particulares da expressão do Teorema 2.4.2, onde são apresentados dois membros da família de Cópulas Arquimedianas.

**Exemplo 2.4.1.** *Seja  $\varphi(t) = (-\ln t)^\theta$ , com  $\theta \geq 1$ . Como é possível observar,  $\varphi(t)$  é uma função contínua e  $\varphi(1) = 0$ .  $\varphi'(t) = -\theta(-\ln t)^{\theta-1}$ , portanto  $\varphi$  é uma função estritamente decrescente de  $[0, 1]$  até  $[0, \infty]$ .  $\varphi''(t) \geq 0$  no intervalo  $[0, 1]$  e, portanto,  $\varphi$  é convexo. Para além disso,  $\varphi(0) = +\infty$ , logo  $\varphi$  é um gerador estrito. Da equação (2.4.3), obtém-se*

$$C_\theta(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right).$$

Sabe-se ainda que  $C_1 = \Pi$  e  $\lim_{\theta \rightarrow +\infty} C_\theta = M$ . Esta família de Cópulas é denominada de família de Cópulas de Gumbel.

**Exemplo 2.4.2.** *Seja  $\varphi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$ , onde  $\theta \in \mathbb{R} \setminus \{0\}$ . Com esta função obtém-se a família de Cópulas de Frank, definidas por*

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right).$$

As Cópulas de Frank são Cópulas Arquimedianas estritas. Para além disso,  $\lim_{\theta \rightarrow -\infty} C_\theta = W$ ,  $\lim_{\theta \rightarrow 0} C_\theta = \Pi$  e  $\lim_{\theta \rightarrow +\infty} C_\theta = M$ . Os membros da família de Cópulas de Frank são as únicas Cópulas Arquimedianas que satisfazem a equação  $C(u, v) = \hat{C}(u, v)$  devido à chamada simetria radial.

## 2.5 Métodos de ajustamento de Cópulas

Com base nas referências [15], [17] e [20], ir-se-á apresentar os vários métodos de ajustamentos de Cópulas.

Existem quatro métodos principais para ajustar Cópulas, que diferem na forma como as marginais são consideradas no processo de ajustamento.

Considere-se uma amostra de vetores aleatórios  $(X_1, \dots, X_n)$  e observações correspondentes  $(x_1, \dots, x_n)$ , onde a  $i$ -ésima marginal tem densidade  $f_i(\cdot; \theta_i)$  e função de distribuição  $F_i(\cdot; \theta_i)$ .

### 2.5.1 Método da Máxima Verosimilhança

O primeiro método de ajustamento de Cópulas é o método da máxima verosimilhança, também conhecido como *Full Maximum Likelihood* (FML), que consiste em maximizar a verosimilhança, ver [17],

$$\begin{aligned} \mathcal{L}(\alpha, \theta_1, \dots, \theta_d, x_1, \dots, x_n) &= \prod_{i=1}^n c(F_1(x_{1,i}; \theta_1), \dots, F_d(x_{d,i}; \theta_d); \alpha) \cdot f_1(x_{1,i}; \theta_1) \dots f_{d,i} \\ &= f_j(x_{d,i}; \theta_d) \end{aligned} \quad (2.5.1)$$

sendo  $\alpha$  o parâmetro da Cópula  $C$  e  $\theta_i$  o parâmetro para a  $i$ -ésima distribuição marginal. A otimização é realizada em torno do espaço do parâmetro.

### 2.5.2 Método dos momentos

O segundo método é o método dos momentos, que consiste em igualar os momentos teóricos e momentos empíricos. Os parâmetros marginais  $\theta_i$  são definidos igualando os momentos empíricos da amostra  $(X_{i,1}, \dots, X_{i,n})$ , enquanto que os parâmetros de Cópula  $\alpha$  são conhecidos através da combinação do tau de Kendall ou do rho de Spearman.

### 2.5.3 Inferência para margens

O terceiro método de estimação, designado por inferência para margens, é um procedimento em 2 passos. Primeiro, as distribuições das marginais são ajustadas pelo método de máxima verosimilhança e, de seguida, uma pseudo amostra é definida como

$$\hat{u}_i = (\hat{u}_{1,i}, \dots, \hat{u}_{d,i}) = \left( F(x_{1,i}, \hat{\theta}_1), \dots, F(x_{d,i}, \hat{\theta}_d) \right),$$

para  $i = 1, \dots, n$ . A Cópula é então ajustada em  $\hat{u}_1, \dots, \hat{u}_n$  através do método da máxima verosimilhança

$$\mathcal{L}(\alpha, \hat{u}_1, \dots, \hat{u}_n) = \prod_{i=1}^n c(\hat{u}_{1,i}, \dots, \hat{u}_{d,i}; \alpha).$$

O método inferência das margens beneficia dos dois passos para reduzir a dimensão da verossimilhança de  $(\alpha, \theta_1, \dots, \theta_d)$  para  $\alpha$ .

#### 2.5.4 Máxima Verossimilhança em Dois Passos

A estimação através maximização da função log-verossimilhança tal como (2.5.1), devido à sua complexidade, poderá causar algumas complicações do ponto de vista computacional. Este ultimo método permite ultrapassar estes possíveis problemas .

Também conhecido como *Two-step Maximum Likelihood*, este método é bastante semelhante ao método da inferência das margens. Consiste em substituir a estimação paramétrica pela estimação não paramétrica na pseudo data, isto é,  $\hat{u}_i = (F_n(x_{1,i}), \dots, F_n(x_{d,i}))$  é simplificado para  $\hat{u}_i = (\text{rank}(x_{1,i})/n, \dots, \text{rank}(x_{d,i})/n)$ .

Na próxima secção, o método utilizado para a estimação de parâmetros será o método de Inferência para margens, apresentado na secção 2.5.3.

# AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

Como referido anteriormente, existem diversos trabalhos na área do Atuariado que abordam o tema de uso de Cópulas no ponto de vista de Seguro Automóvel. Como exemplos, referimos a dissertação [12], onde são utilizadas Cópulas para modelação do número de sinistros de três grupos de risco homogéneos do ramo Automóvel, o artigo [2], que testa a modelação da dependência bivariada entre perdas, como também estima o risco de uma seguradora do ramo não-vida usando o VaR e o TVaR. Para além desses, também os artigos [14] e [3] utilizam as Cópulas, para entender qual é a relação entre os riscos de uma carteira de grupo, e para modelar a estrutura de dependência de dados provenientes de Ramo de Seguros, respetivamente.

Na indústria de seguros, as dependências de risco entre as perdas foram reconhecidas como os principais fatores que influenciam as perdas agregadas da seguradora. Para fazer face a este problema, diversos autores sugeriram o uso de Cópulas para modelar dependências entre diferentes tipos de perdas. Nesta dissertação, serão realizados, como referido atrás, ilustrações numéricas de Cópulas, o seu ajustamento e a sua estimativa sobre um conjunto de dados real de um Seguro Automóvel, a partir do qual se pretende estudar a dependência dos montantes das indemnizações pagas no âmbito das coberturas de Danos Próprios e Responsabilidade Civil.

## 3.1 Dados

### 3.1.1 Descrição da Base de Dados

O objetivo desta aplicação é analisar uma carteira de seguro automóvel que possui contractos com diferentes tipos de coberturas, entre elas as coberturas de Responsabilidade civil e diferentes coberturas de Danos Próprios que, num mesmo sinistro, podem ser acionadas em simultâneo.

Os dados trabalhados nesta dissertação contém um conjunto de 7.569 observações, sendo as variáveis disponíveis o Número de Contrato, o Tipo de Cobertura acionada e o Montante do sinistro. Para além desta variáveis, para cada apólice há também informação sobre o Tipo de Combustível e Idade do veículo, a Zona de Residência, a Potência do veículo, a Marca do veículo e Idade do condutor.

Existem 6 tipos diferentes de coberturas, Responsabilidade Civil (RC) e no âmbito de Danos Próprios, as coberturas Choque, Colisão e Capotamento (CCC), Incêndio, Raio ou Explosão (IRE), Quebra Isolada de Vidros (QIV), Fenómenos da Natureza (FN) e Actos de Vandalismo (AV). A tabela 3.1.1, resume o número de contratos que inclui cada uma das coberturas referidas.

Coberturas	RC	CCC	IRE	QIV	FN	AV
nº de contratos	2.697	2.225	448	2.152	40	7

Tabela 3.1: Coberturas e número de contractos por cobertura

Tendo em consideração que o âmbito do estudo consiste na análise da eventual dependência entre montantes de sinistros de Danos Próprios e Responsabilidade Civil, os montantes decorrentes das várias coberturas referentes aos danos do veículo do segurado serão agregados, não fazendo distinção entre as diferentes coberturas.

### 3.1.1.1 Tratamento de Dados

A base de dados inicial possui 7.569 observações, pelo que, com base em diversos critérios que serão posteriormente referidos, foi reduzindo de dimensão até ficar com 664 observações.

A explicação para a redução foi a seguinte:

1. Uma vez que se quer trabalhar com custos de sinistros, só devem ser consideradas observações com o valor superior a 0 u.m. Os montantes dos dados iniciais variam entre -3.811,23 u.m. e 152.449 u.m, pelo que os dados foram reduzidos para 6.196 observações neste primeiro passo.
2. As apólices foram agregadas pelo seu número de contrato, tendo sido excluídas todas aquelas que possuem apenas uma cobertura acionada, uma vez que se quer analisar quando ocorrem RC e DP em simultâneo.
3. Foram excluídas aquelas que não possuem pelo menos uma cobertura acionada de RC
4. Finalmente, para além de terem de possuir uma cobertura de RC, também necessitam de possuir uma cobertura acionada de DP.

Nas tabelas 3.2 e 3.3 torna-se visível as alterações feitas na base de dados, passo a passo.

CAPÍTULO 3. AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

Número de Contrato	Cobertura	Montante
6.375	CCC	83,80
6375	CCC	3.348,50
6.375	RC	2.031,77

Tabela 3.2: Dados iniciais para o nº de Contrato 6.375

Número de Contrato	Cobertura RC	Montante RC	Cobertura DP	Montante DP
6.375	RC	2.031,77	CCC;CCC	83,8; 3.348,5
Soma de Montantes RC		Soma de Montantes DP		
2.031,77 u.m.		3.432,3 u.m.		

Tabela 3.3: Dados Finais

O processo de exclusão de observações pode ser visualizado no esquema 3.1.

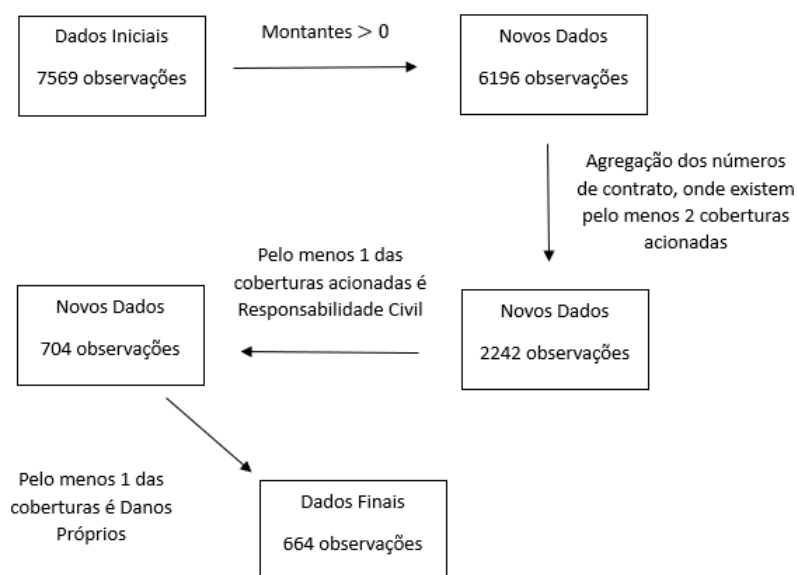


Figura 3.1: Esquema do tratamento de Dados

A amostra final contém, desta forma, um conjunto de 664 observações, onde as coberturas de RC e DP (correspondente a qualquer uma das coberturas de DP referidas na Tabela 3.1.1) podem ter sido acionadas múltiplas vezes, desde que tenham sido acionadas em simultâneo.

O mais importante deste tratamento de dados é a coluna das Somas de Montantes de RC e de DP, como ilustrado na Tabela 3.3 a título de exemplo para um contrato em particular, pois serão estes valores que serão analisados de forma global para a totalidade das observações em estudo.

### 3.1.2 Análise Descritiva

#### 3.1.2.1 Responsabilidade Civil

Realizou-se seguidamente uma análise exploratória da severidade dos sinistros relativos à cobertura de Responsabilidade Civil, como se pode ver na tabela 3.4, verificando que o custo médio de resolução de um sinistro, cujos danos estão a abrigo da cobertura de Responsabilidade Civil é de 2.330,71 u.m., com um desvio padrão de 4.486,97 u.m. Verifica-se, ainda, que o coeficiente de variação é de 1,92.

Valor Médio	Desvio Padrão	Coeficiente de Variação		
2.330,71	4.486,97	1,92		
Mínimo	Q1/4	Q1/2	Q3/4	Máximo
0,12	1.036,87	1.204	1.965,38	58.711

Tabela 3.4: Estatísticas elementares da severidade dos montantes de sinistros de Responsabilidade Civil

Observa-se ainda que as indemnizações variam entre 0,12 e 58.711 u.m., 25% dos custos dos sinistros tem um valor inferior ou igual a 1.036,87 u.m. e 75% dos custos dos mesmos tem um valor inferior ou igual a 1.965,38 u.m. Para além disso, 50% dos custos dos sinistros tem um valor inferior ou igual a 1.204 u.m., que é inferior à média, indicando que a distribuição do custo de sinistros será assimétrica à direita.

O boxplot da figura 3.2 ilustra o que foi mencionado anteriormente. É possível também observar que existem um número elevado de *outliers*.

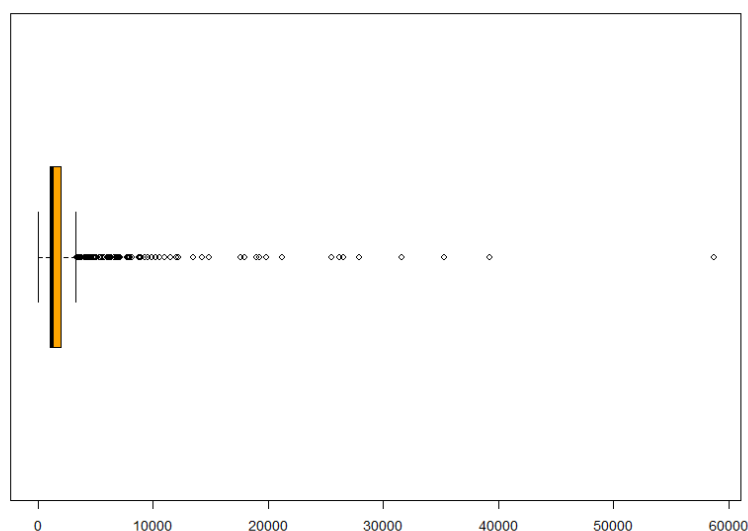


Figura 3.2: Boxplot dos custos dos sinistros de Responsabilidade Civil

### 3.1.2.2 Danos Próprios

Procedeu-se com a análise descritiva do custo de sinistros relativos aos Danos Próprios onde, de forma análoga, obteve-se o custo médio das apólices e o desvio padrão, com os valores de 2.098,64 e 5.382,19, respetivamente, como é possível observar na tabela 3.5. Verifica-se também que o desvio padrão corresponde a, aproximadamente, é 2.6 vezes a média.

Valor Médio	Desvio Padrão	Coeficiente de Variação		
2.098,64	5.382,19	2,56		
Mínimo	Q1/4	Q1/2	Q3/4	Máximo
8,67	348,82	1.013,78	2.167,79	115.626,2

Tabela 3.5: Estatísticas elementares da severidade dos montantes de sinistros de Danos Próprios

Na observação da tabela 3.5, constata-se que as indemnizações variam entre 8,67 e 115.626,2 u.m., 25% dos custos dos sinistros tem um valor inferior ou igual a 348,82 u.m. e 75% dos custos dos mesmos tem um valor inferior ou igual a 2.167,79 u.m. Para além disso, 50% dos custos dos sinistros tem um valor inferior ou igual a 1.013,79 u.m., que é inferior ao valor médio, indicando que a distribuição do custo de sinistros será assimétrica à direita.

O boxplot da figura 3.3 ilustra o que foi mencionado anteriormente. É possível também observar que existem bastantes outliers, sendo um indicador de que, posteriormente, poderá ser necessário realizar uma separação das indemnizações de menor e maior custo.

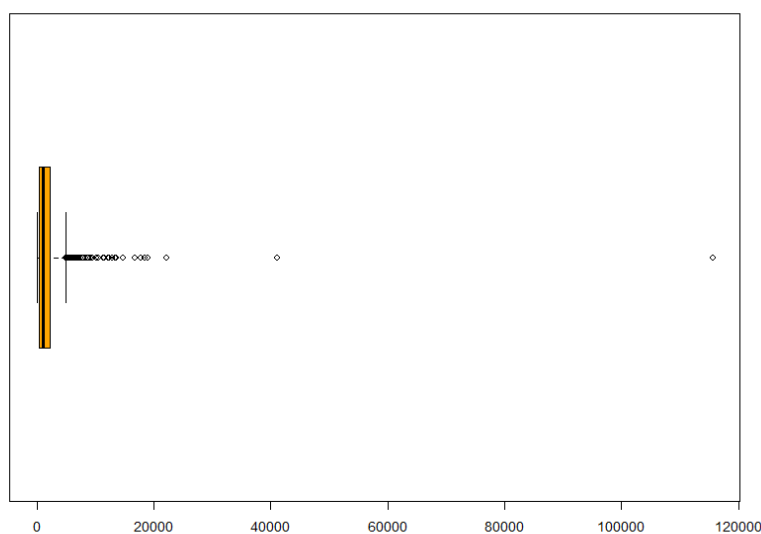


Figura 3.3: Boxplot dos custos de sinistros de Danos Próprios

### 3.1.3 Distribuição das Margens

Conforme descrito na secção 2.1, para além de se procurar a cópula mais apropriada para se ajustar aos dados, também é necessário encontrar a distribuição adequada das margens.

Inicia-se a análise das variáveis, examinando-se quais as distribuições candidatas para as marginais.

Os histogramas dos montantes de sinistros referentes às coberturas de Responsabilidade Civil e Danos Próprios, apresentados na figura 3.4, permitem observar que ambas as distribuições apresentam caudas pesadas.

Em ambas as situações, como é possível ver na tabela 3.6 o teste de ajustamento com base no rácio de dois estimadores de variância, ver [21], para o qual se observou um p-value de  $2,2 \times 10^{-16}$ , leva-nos a rejeitar a hipótese de que os montantes dos sinistros poderão ser modelados através uma distribuição Gama.

	Responsabilidade Civil	Danos Próprios
p-value	$2,2 \times 10^{-16}$	$2,2 \times 10^{-16}$

Tabela 3.6: Teste de Ajustamento à distribuição Gama

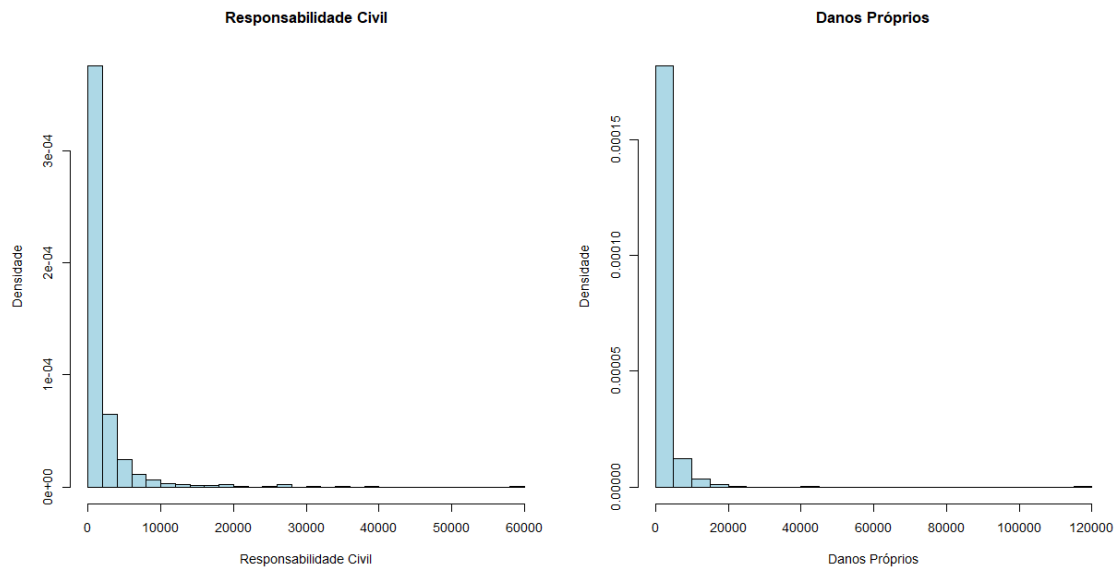


Figura 3.4: Histogramas dos montantes dos sinistros de Responsabilidade Civil e Danos Próprios

### CAPÍTULO 3. AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

Posto isto, realizaram-se várias tentativas do ponto de truncatura que permitissem identificar se as margens seguem uma distribuição de probabilidade. Apesar deste tratamento dos dados terem fornecido distribuições para as variáveis, mesmo assim não foi possível ajustar uma cópula com o procedimento efetuado. No final, o melhor resultado que fornecesse distribuições para as margens e que permitisse o ajustamento de cópulas foi o seguinte.

Foram criados dois grupos, aos quais serão denominados *A* e *B*.

O grupo *A* é uma interseção dos montantes das variáveis Responsabilidade Civil e Danos Próprios, pelo qual foram limitados com a seguinte expressão:

$$A = \{\text{Montante de RC} < 5.100 \text{ u.m.} \wedge \text{Montante de DP} < 4.500 \text{ u.m.}\} \quad (3.1.1)$$

Com a criação deste novo conjunto *A*, cujo número de observações é de 557, realizou-se novamente o teste de ajustamento para a distribuição Gama.

	p-value	$\alpha$	$\beta$
Responsabilidade Civil	0,1286	1,7164	785,54
Danos Próprios	0,0704	1,0653	1.034,71

Tabela 3.7: Resultados dos testes de ajustamento para as coberturas RC e DP

Realizando o mesmo teste de ajustamento atrás referido, [21], os resultados apresentados na tabela 3.7 leva-nos a não rejeitar a hipótese de que os montantes dos sinistros deste Conjunto *A* poderão ser modelados através uma distribuição Gama.

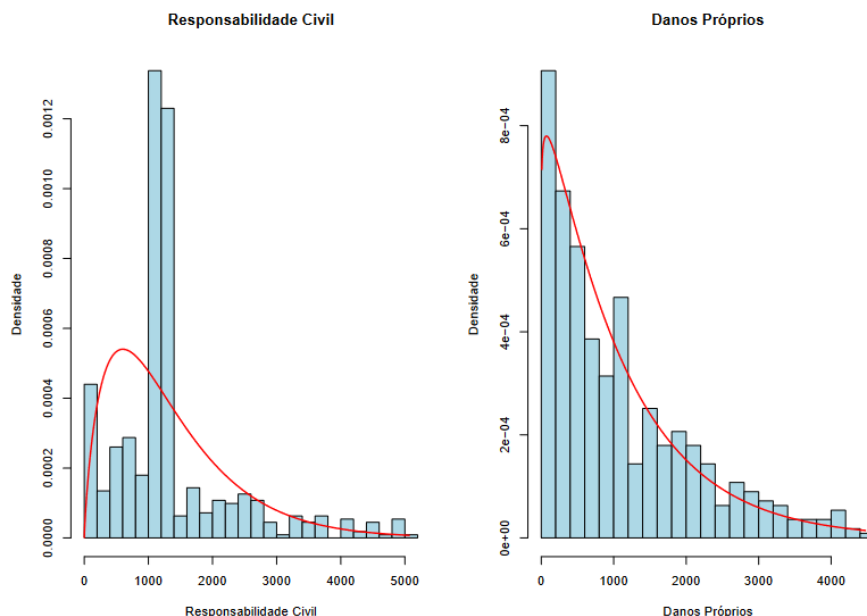


Figura 3.5: Histogramas dos custos de sinistros Responsabilidade Civil e Danos Próprios limitados

Para o grupo  $B$ , a expressão para limitar os dados é o complementar do Conjunto  $A$ , isto é

$$B = \overline{A} = \{\text{Montante de RC} \geq 5.100 \text{ u.m.} \vee \text{Montante de DP} \geq 4.500 \text{ u.m.}\} \quad (3.1.2)$$

O número de observações presentes no grupo  $B$  são as 107 observações que excederam os limites considerados para o ajustamento das distribuições Gama atrás referidas.

As figuras 3.6 e 3.7 ilustram o histograma e boxplot das 107 observações do Conjunto  $B$ .

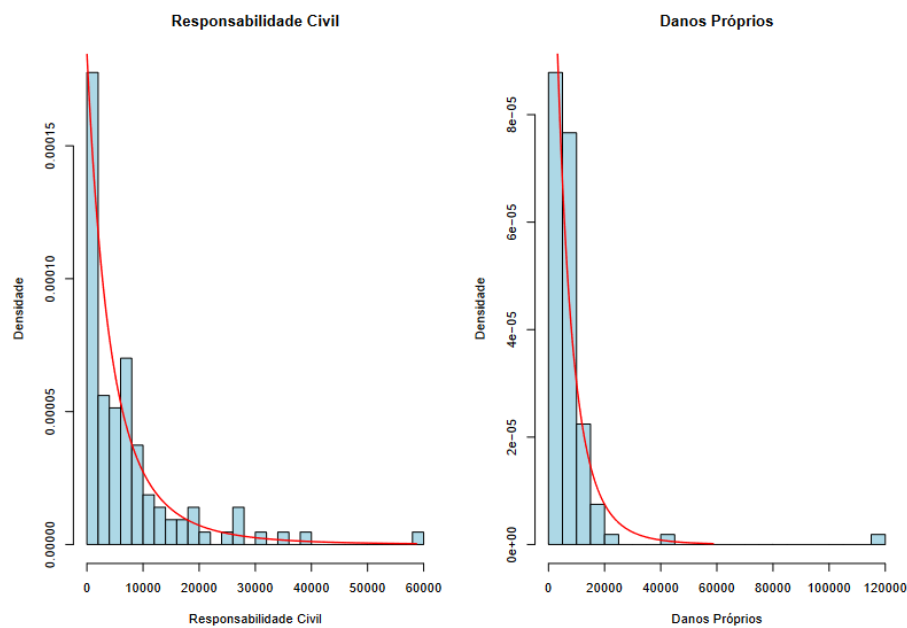


Figura 3.6: Histogramas dos montantes de sinistros Responsabilidade Civil e Danos Próprios das 107 observações

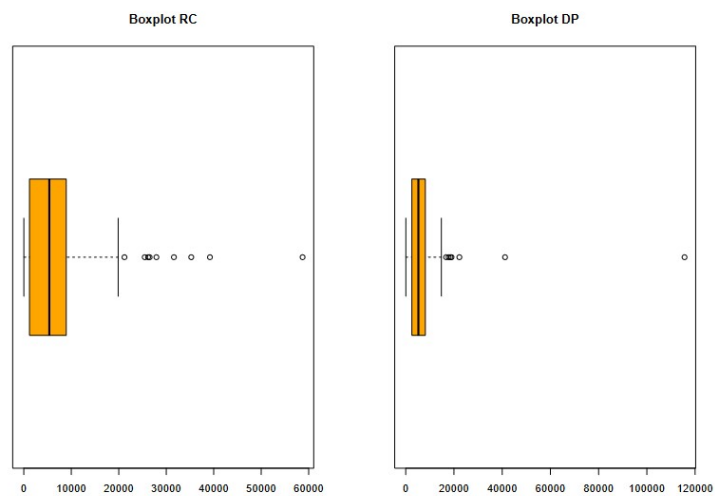


Figura 3.7: Boxplot dos montantes de sinistros Responsabilidade Civil e Danos Próprios das 107 observações

CAPÍTULO 3. AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

Construindo-se diagramas de Pareto para ambas as marginais, como se apresenta na figura 3.8, observa-se um alongamento da cauda, pelo que se conclui que que ambas variáveis possuem distribuições de cauda pesada.

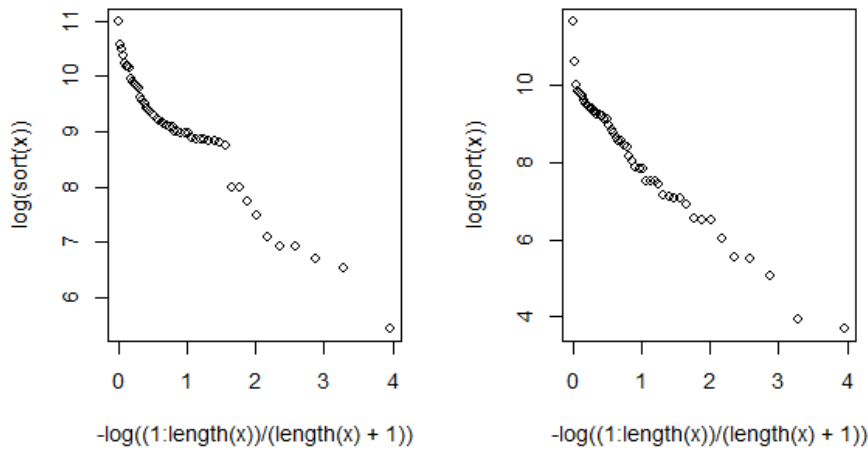


Figura 3.8: Diagrama de Pareto

Efetuando-se o teste de ajustamento de *Bootstrap* para a distribuição Pareto Generalizada, ver [22], obteve-se os resultados da tabela 3.8.

	p-value	$\alpha$	$\beta$
Responsabilidade Civil	0,1391	3,465	18.712,3
Danos Próprios	0,0711	6,620	39.852,9

Tabela 3.8: Resultados dos testes de ajustamento para as coberturas RC e DP

Sendo assim, com um nível de significância  $\alpha \geq 0.05$ , leva-nos a não rejeitar a hipótese de que os montantes dos sinistros poderão ser modelados através de uma distribuição Pareto, para ambas as margens.

Agora encontradas as distribuições das marginais para os conjuntos  $A$  e  $B$ , sendo a distribuição Gama e a distribuição Pareto, respetivamente, torna-se possível iniciar a aplicação das Cópulas.

### 3.1.4 Seleção de Cópulas - Conjunto A

Nesta aplicação, decidiu-se ajustar as Cópulas bivariadas Gaussiana, de Student, de Gumbel, de Frank e de Husler-Reiss, trabalhando-se com os dados tratados anteriormente.

Efetuando-se o ajustamento das Cópulas para os dados do Conjunto A, obtém-se os parâmetros com os seguintes valores:

	Indep.	Gaussian	Student	Husler-Reiss	Gumbel	Frank
alpha	0,5	0,1611	0,1608	0,6047	1,0776	0,9823
	-	-	90,631	-	-	-
shape	1,7164	1,7164	1,7164	1,7164	1,7164	1,7164
scale	785,672	785,672	785,672	785,672	785,672	785,672
shape	1,0656	1,0656	1,0656	1,0656	1,0656	1,0656
scale	1.034,726	1.034,726	1.034,726	1.034,726	1.034,726	1.034,726

Tabela 3.9: Estimação dos parâmetros de cada Cópula

O nível de dependência, observando a tabela, parece apresentar valores baixos, uma vez que para o primeiro parâmetro, o valor é próximo de 1, para as Cópulas não elípticas, ou próximo de 0 para as Cópulas elípticas. Os parâmetros ajustados das distribuições das margens são idênticos, uma vez que foram utilizados os mesmos parâmetros para todas as Cópulas. De maneira a avaliar a qualidade dos ajustamentos, ir-se-á observar os coeficientes cauda  $\lambda_U(X, Y)$  e  $\lambda_L(X, Y)$ , que são obtidos dado uma cópula. Estes podem ser estimados nos dados através de uma estimativa não-paramétrica de  $C(t, t)$ , isto é, a função de distribuição empírica bivariável

$$C_n(t, t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x_i \leq t} \mathbf{1}_{y_i \leq t}$$

Na figura 3.9 estão representadas as diferentes curvas das Cópulas estudadas, podendo-se observar quais as cópulas que se melhor aproximam dos coeficientes cauda.

Verifica-se que os coeficientes de cauda são melhor aproximados pelas Cópulas Gaussiana e de Frank. Para além disso, é possível observar também que a curva da cópula independente passa longe dos dados empíricos, pelo que reforça a hipótese de que as variáveis não são independentes.

CAPÍTULO 3. AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

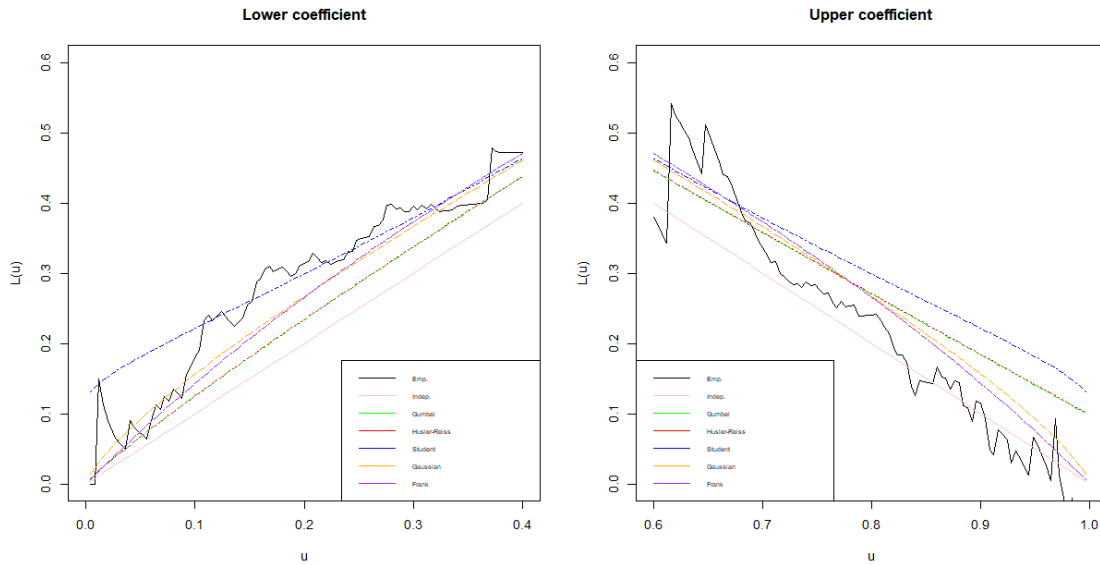


Figura 3.9: Coeficientes Cauda

Observando as duas medidas de dependência na tabela , o Tau de Kendall e o Rho de Spearman, comprova-se que as Cópulas anteriormente referidas refletem bons modelos uma vez que possuem os valores mais próximos do valor empírico.

	Emp.	Gaussian	Student	Husler-Reiss	Gumbel	Frank
Tau	0,09911422	0,1029891	0,1028408	0,07563426	0,07203415	0,108106
Rho	0,1403931	0,15	0,15	0,1130143	0,1073843	0,1616451

Tabela 3.10: Tau de Kendall e Rho de Spearman de cada Cópula

Desta forma, o estudo seguirá utilizando exclusivamente com as Cópulas de Frank e Gaussiana. Dada uma Cópula, quantidades de interesse podem ser estimadas pelo método de Monte-Carlo. Focando-se na distribuição de custo total, isto é, a soma do custo de Responsabilidade Civil e Danos Próprios, e escolhendo-se  $n=10^4$  para o tamanho da amostra, constrói-se o gráfico das funções de distribuição empíricas da soma para as Cópulas de Frank, Gaussiana e das Cópulas independentes.

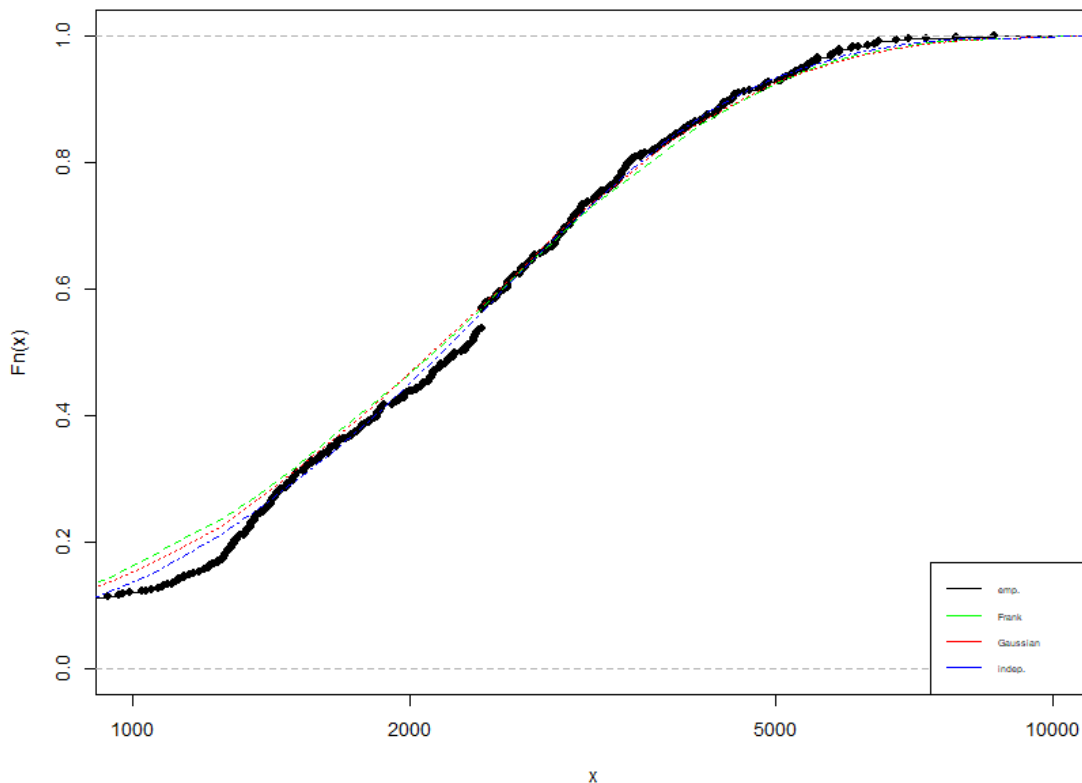


Figura 3.10: Caudas da distribuição da perda total (Responsabilidade Civil + Danos Próprios)

Pela figura 3.10, observa-se que ambas as cópulas refletem um bom ajustamento aos dados.

A cópula independente parece ser uma aproximação razoável para a distribuição, no entanto serão adiante apresentados resultados do teste de independência.

### 3.1.5 Seleção de Cópulas - Conjunto B

Finalmente, para os dados do Conjunto B, obtém-se os parâmetros com os valores na tabela 3.11, através de um processo análogo ao anterior:

O nível de dependência, observando a tabela, parece apresentar valores baixos, uma vez que para o primeiro parâmetro, o valor é 0, para as Cópulas elípticas, e para as Cópulas não elípticas variam bastante. Os parâmetros ajustados das distribuições das margens são, mais uma vez, idênticos, como era de esperar.

CAPÍTULO 3. AJUSTAMENTO DE CÓPULAS ÀS INDEMNIZAÇÕES DE RESPONSABILIDADE CIVIL E DANOS PRÓPRIOS

	Indep.	Gaussian	Student	Husler-Reiss	Gumbel	Frank
alpha	0,5	0,00	0,00	0,3506	1,0000	-2,6616
	-	-	8,5721	-	-	-
shape	3,4654	3,4654	3,4654	3,4654	3,4654	3,4654
scale	18.712,5893	18.712,5893	18.712,5893	18.712,5893	18.712,5893	18.712,5893
shape	6,6209	6,6209	6,6209	6,6209	6,6209	6,6209
scale	39.825,372	39.825,372	39.825,372	39.825,372	39.825,372	39.825,372

Tabela 3.11: Estimação dos parâmetros de cada Cópula

Avaliando a qualidade dos ajustamentos, observando-se os coeficientes cauda na figura 3.11, apesar dos resultados não refletirem um ajustamento não tão bom como os da secção 3.1.4, os coeficientes de cauda parecem ser melhor aproximados pelas Cópulas não elípticas, Cópula de Frank e Husler-Reiss.

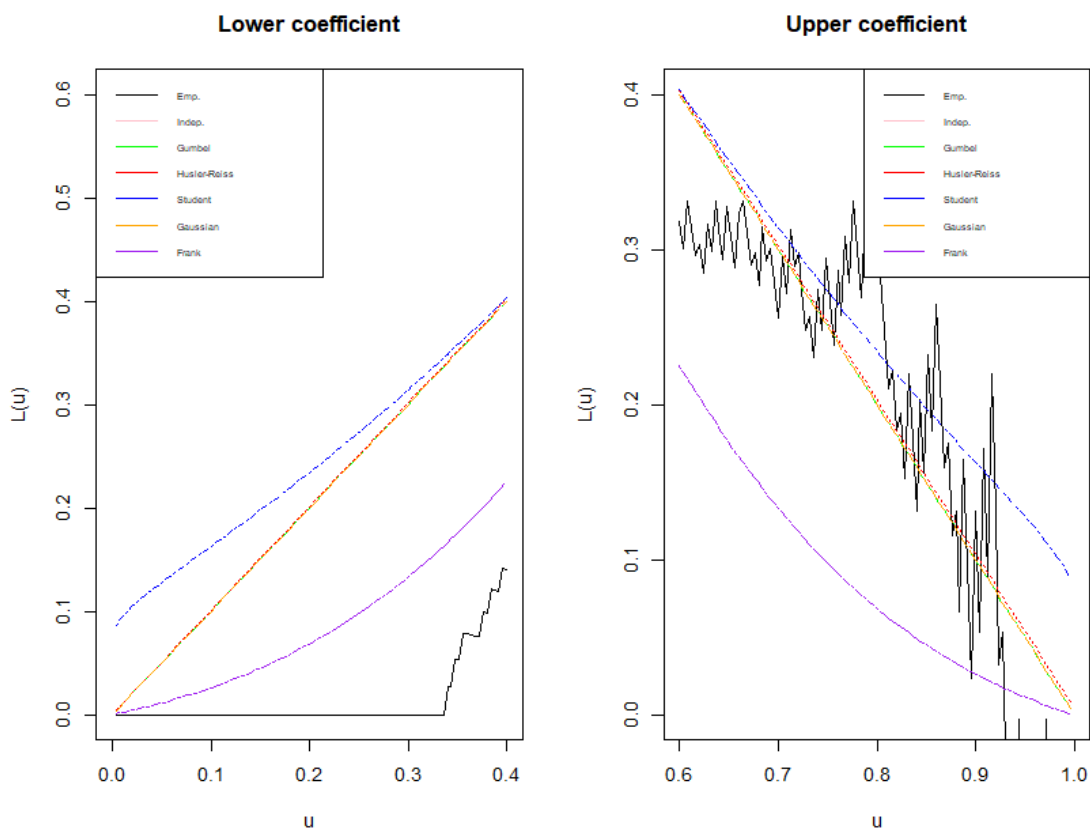


Figura 3.11: Coeficientes Cauda

Os valores do Tau de Kendall e do Rho de Spearman vêm comprovar que as Cópulas anteriores são admissíveis.

Uma vez que apenas uma Cópula parece apresentar valores aceitáveis, o estudo irá proceder-se exclusivamente com as Cópulas de Frank. Utilizando o mesmo método de

	Emp.	Gaussian	Student	Husler-Reiss	Gumbel	Frank
Tau	-0,2271	0	0	0,00334	0	-0,277
Rho	-0,3078	0	0	0,00501	0	-0,4067

Tabela 3.12: Tau de Kendall e Rho de Spearman de cada Cópula

Monte-Carlo, e focando-se novamente na distribuição de custo total, construiu-se o gráfico das funções de distribuição empíricas da soma para as Cópulas de Frank e das Cópulas independentes.

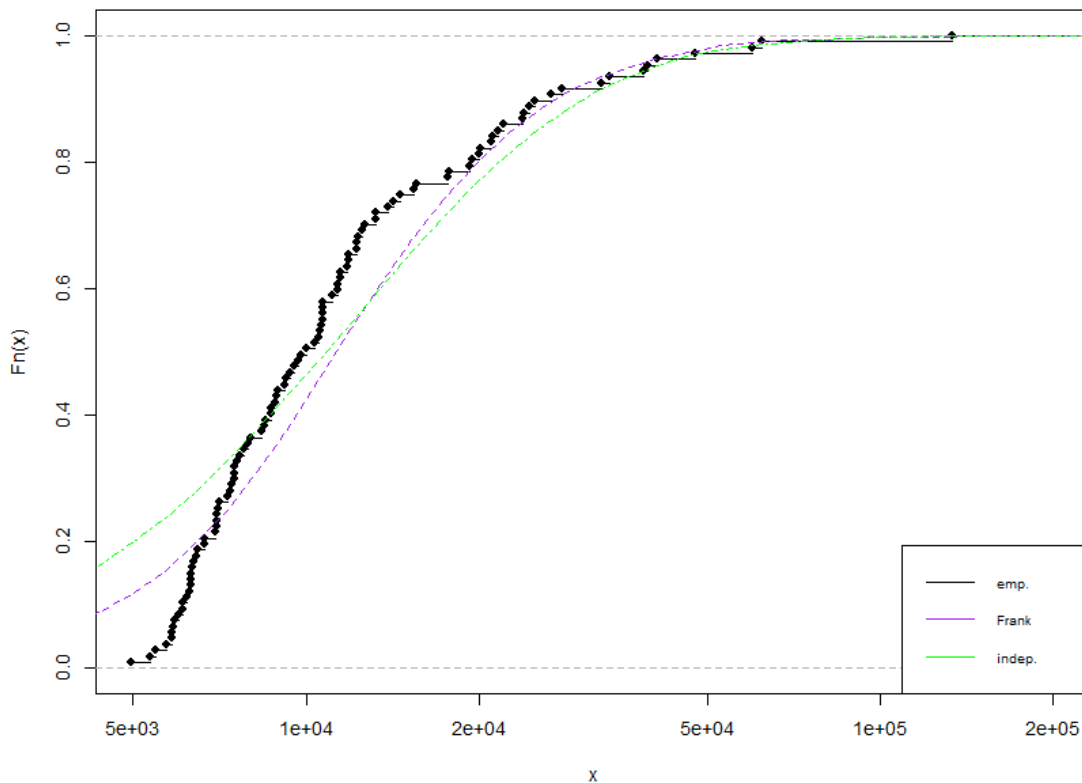


Figura 3.12: Caudas da distribuição da perda total (Responsabilidade Civil + Danos Próprios)

Na figura 3.12, consegue-se observar que as Cópulas apenas aparentam ser boas aproximações nos valores da cauda da distribuição. A cópula de Frank é a que melhor se ajusta aos dados, embora possa ser considerada com qualidade de ajustamento não satisfatória.

### 3.1.6 Testes de Ajustamento e de Independência

Com base nas referências [23] e [24], para se realizar o teste de qualidade de ajustamento, assumiu-se inicialmente quais as cópulas que ajustam aos dados.

Considere-se um vetor aleatório contínuo  $X = (X_1, \dots, X_d)$  com função de distribuição cumulativa conjunta  $H$  e margens  $F_1, \dots, F_d$ . A representação da cópula de  $H$  é dada por  $H(x_1, \dots, x_d) = C\{f_1(x_1), \dots, F_d(x_d)\}$ , onde  $C$  é uma função de distribuição conjunta única que possui margens uniformes em  $(0, 1)$ . Um modelo Cópula para  $X$  origina-se quando  $C$  é desconhecido mas assume-se que pertence a uma classe

$$C_0 = \{C_\theta : \theta \in \mathcal{O}\},$$

onde  $\mathcal{O}$  é o subconjunto aberto de  $\mathbb{R}^p$  para algum inteiro  $p \geq 1$ . O teste de hipóteses a considerar é:

$$H_0 : C \in C_0 \quad VS \quad H_1 : C \notin C_0$$

Já para o teste de independência, com base em [25], assume-se que a hipótese nula é a cópula ser independente, ou seja, o teste de hipóteses é definido como

$$H_0 : C \text{ é independente} \quad VS \quad H_1 : C \text{ não é independente}$$

Na seguinte tabela 3.13, estão apresentados os resultados dos testes de ajustamento de cópulas conjuntamente com os resultados dos testes de independência.

Cópulas	Conjunto A			Conjunto B	
	Frank	Gaussiana	Indep.	Frank	Indep.
Teste de ajustamento (p-value)	0,6648	0,6928	-	0,5779	-
Teste de independência (p-value)	0,00495	0,00495	0,40099	0,000499	0,8861

Tabela 3.13: Resultados dos testes de Ajustamento e de Independência

Portanto, através da observação da tabela 3.13, em relação ao teste de qualidade de ajustamento das cópulas, todas as Cópulas apresentam p-values superiores ao nível de significância de 5%, pelo que não se rejeita a hipótese de que

$$C_A \in C_0 \quad e \quad C_B \in C_0,$$

com  $C_A = \{\text{Frank}, \text{Gaussiana}\}$ ,  $C_B = \{\text{Frank}\}$  e  $C_0 = \{\text{Frank}, \text{Gaussiana}\}$ .

Em relação ao teste de independência, os p-values apresentam valores muito baixos, inferiores a 5%, e portanto, rejeita-se a hipótese de que estas Cópulas sejam independentes.

Observando as figuras 3.13 e 3.14, sendo a cópula independente representada pela curva azul em ambas, é evidente a diferença existente em relação à cópula de Frank e a Gaussiana em relação à independente, reforçando a hipótese de que não são independentes.

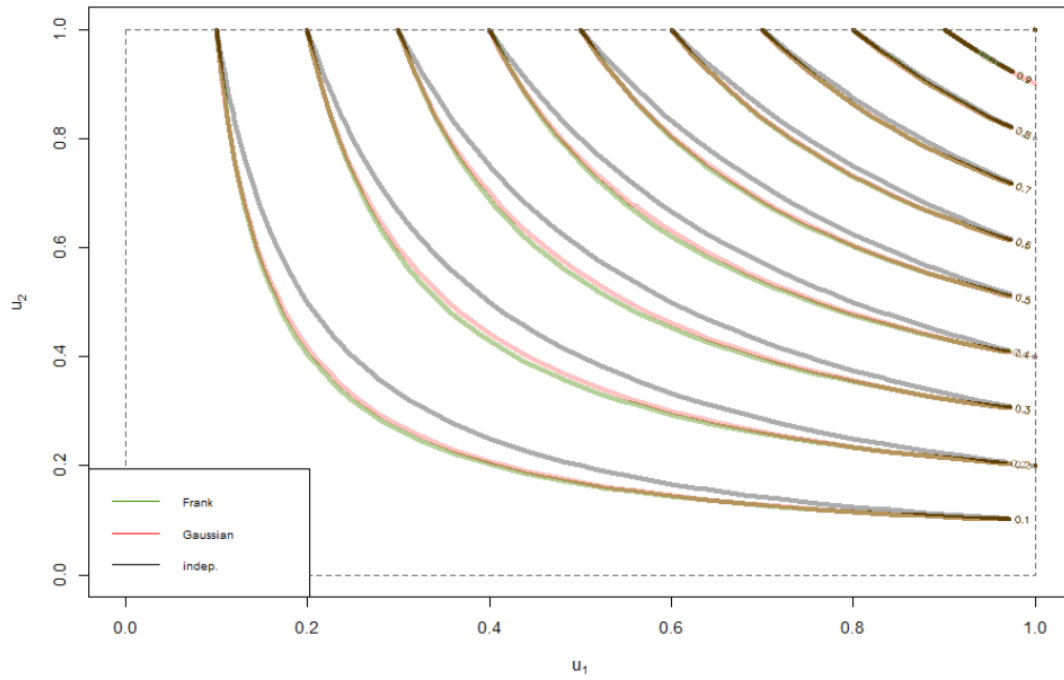


Figura 3.13: Comparação das Cópulas de Frank e Gaussiana com Cópula Independente do Conjunto A

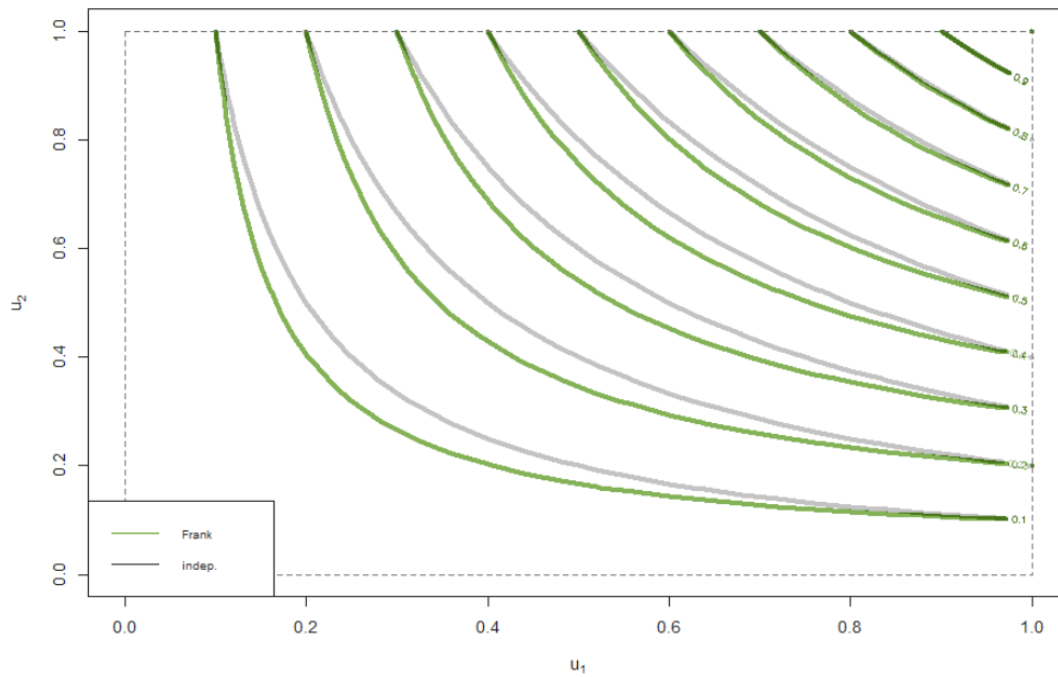


Figura 3.14: Comparação das Cópulas de Frank e Cópula Independente do Conjunto B

## CONCLUSÃO

A Modelação Matemática consiste no estabelecimento de um conjunto de ferramentas matemáticas que possibilitam a construção de uma análise teórica com referência na realidade envolvente.

Estes modelos originados da Modelação Matemática são simplificações dessa realidade envolvente e a indústria do risco, nomeadamente a indústria seguradora, quando avalia o impacto financeiro da concretização do risco, tem como exigência a introdução de diversas simplificações, algumas de modo explícito, na forma de pressupostos.

Um desses pressupostos é a independência entre alguns ou mesmo todos os riscos que se procuram modelar, pelo que, numa avaliação atuarial, usualmente se assume que são independentes entre si.

Com o teste de independência para as Cópulas quer com as ilustrações gráficas ao longo da dissertação, retira-se, de modo explícito, que as Cópulas ajustadas para os dados não são independentes.

Posto isto, visto que não são independentes, então as variáveis aleatórias analisadas também não o são e, por consequência, as coberturas de Responsabilidade Civil e de Danos Próprios são dependentes.

Estes resultados são bastantes importantes para a área do Atuariado, visto que, como foi referido atrás, este pressuposto de independência entre riscos ainda hoje é explicitamente considerado verdadeiro nas avaliações atuariais.

As duas últimas ilustrações desta dissertação revelam diferenças significativas entre a Cópula Independente e as outras Cópulas ajustadas, sendo que estas diferenças podem implicar previsões imprecisas para a Seguradora.

Esta nova informação de dependência destas coberturas poderá trazer um impacto significativo para a indústria seguradora, pelo que terá de ser necessário reestruturar a metodologia do cálculo de prémios.

## BIBLIOGRAFIA

- [1] J. M. Lourenço. *The NOVAtHesis L<sup>A</sup>T<sub>E</sub>X Template User's Manual*. NOVA University Lisbon. 2021. URL: <https://github.com/joaomlourenco/novathesis/raw/main/template.pdf> (ver pp. i, 5).
- [2] N. F. Rozainee e I. Ismail. *Bivariate Dependence Modelling of a Non-Life Insurance Company: Pair-Copula Construction*. International Journal of Science e Research (IJSR), 2022 (ver pp. iv, v, 4, 22).
- [3] I. Ghosh, D. Watts e S. Chakraborty. *Modeling Bivariate Dependency in Insurance Data via Copula: A Brief Study*. Journal of Risk e Financial Management 15: 329, 2022 (ver pp. iv, v, 2, 5, 22).
- [4] D. do Parlamento Europeu e do Conselho. *Seguro e resseguro*. 2009. URL: <https://eur-lex.europa.eu/PT/legal-content/summary/insurance-and-reinsurance.html> (ver p. 1).
- [5] R. B. Nelsen. *An introduction to copulas*. New York: Springer, 2006 (ver pp. 1, 8, 9, 11, 13, 15).
- [6] A. Sklar. *Fonctions de répartition à n dimensions et leurs merges*. l'Institut de Statistique de l'Université de Paris, 8: 229-231s, 1959 (ver pp. 2, 4).
- [7] R. B. Nelsen. *An introduction to copulas*. New York: Springer, 1999 (ver pp. 2, 4, 8, 10, 15).
- [8] H. Joe. *Multivariate Models and Dependence Concepts*. Springer Science+Business Media, 1997 (ver pp. 2, 4, 15).
- [9] P. Embrechts, F. Lindskog e A. McNeil. *Modelling Dependence with Copulas and Applications to Risk Management*. 2001 (ver pp. 2, 4, 8, 10, 12, 13, 15–17).
- [10] A. Dias. *Copulae Inference for Finance and Insurance*. Doctoral Thesis. Swiss Federal Institute of Technology, Zurich., 2004 (ver p. 2).
- [11] E. Frees e E. Valdez. *Understanding Relationships Using copulas*. North American Journal, 2: 1-25., 1998 (ver p. 2).

- [12] M. F. dos Santos. *Modelling claim counts of homogeneous risk groups using copulas*. Instituto Superior de Economia e Gestão, 2010 (ver pp. 2, 5, 22).
- [13] D. M. et al. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley Sons, 2005 (ver pp. 4, 9, 10, 12).
- [14] P. Shi, X. Feng e J.-P. Boucher. *Multilevel Modeling of Insurance Claims Using Copulas*. *The Annals of Applied Statistics*, 2016 (ver pp. 5, 22).
- [15] A. J. McNeil, F. Rudiger e P. Embrechts. *Quantitative Risk Management - Concepts, Techniques and Tools*. *Princeton Series in Finance*. Princeton University Press, 2005 (ver pp. 12, 14, 20).
- [16] N. Kolev, B. V. de M. Mendes e U. dos Anjos. *Copulas: a Review and Recent Developments*. 2005 (ver p. 13).
- [17] A. Charpentier. *Computational Actuarial Science with R*. Chapman Hall/CRC, 2005 (ver pp. 15, 20).
- [18] S. Cambanis, S. Huang e G. Simons. *On the theory of elliptically contoured distributions*. *J. Multivar. Anal.* 11, 368–385, 1981 (ver p. 16).
- [19] S. Daul et al. *The Grouped t-Copula with an Application to Credit Risk*. 2003 (ver p. 17).
- [20] H. JOE e J. XU. *The estimation method of inference function for margins for multivariate models*. *Technical Report no 166*. Vancouver: Univ. of British Columbia, Dept. of Statistics, 1996 (ver p. 20).
- [21] V. J.A. e E. Gonzalez-Estrada. *A variance ratio test of fit for Gamma distributions*. *Statistics e Probability Letters*, 2015 (ver pp. 27, 28).
- [22] V. J.A. e E. Gonzalez-Estrada. *A bootstrap goodness of fit test for the generalized Pareto distribution*. *Computational Statistics e Data*, 2015 (ver p. 30).
- [23] C. Genest, W. Huang e J.-M. Dufour. *A regularized goodness-of-fit test for copulas*. *Journal de la Société française de statistique* 154, 64–77, 2013 (ver p. 36).
- [24] C. Genest, B. Rémillard e D. Beaudoin. *Goodness-of-fit tests for copulas: A review and a power study*. *Insurance: Mathematics e Economics* 44, 199–214., 2009 (ver p. 36).
- [25] C. Genest e B. Rémillard. *Tests of independence and randomness based on the empirical copula process*. *Test* 13, 2004 (ver p. 36).



