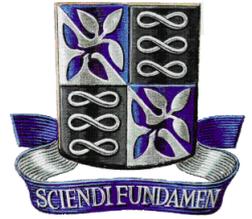




UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
DISSERTAÇÃO DE MESTRADO



CÓPULA PVF: REVISÃO DE LITERATURA E NOVOS RESULTADOS

SUÊDE SANTOS BARBOSA

Área de Concentração: ESTATÍSTICA

Salvador - Bahia
SETEMBRO DE 2023

CÓPULA PVF: REVISÃO DE LITERATURA E NOVOS RESULTADOS

SUÊDE SANTOS BARBOSA

Dissertação de Mestrado apresentada ao Colegiado da Pós-Graduação em Matemática da Universidade Federal da Bahia (UFBa), como parte dos requisitos para obtenção do título de Mestre em Matemática. Área de concentração: Estatística.

Orientador: Prof. Dr. Paulo Henrique Ferreira da Silva

Coorientadora: Profa. Dra. Rosemeire Leovigildo Fiaccone

Salvador - Bahia

Setembro de 2023

Sistema de Bibliotecas da UFBA

Barbosa, Suêde Santos.

Cópula PVF: revisão de literatura e novos resultados / Suêde Santos
Barbosa / Salvador, 2023.

39p. : il

Orientador: Prof. Dr. Paulo Henrique Ferreira da Silva.

Coorientadora: Profa. Dra. Rosemeire Leovigildo Fiaccone.

Dissertação (mestrado) – Universidade Federal da Bahia, Instituto de
Matemática e Estatística, Salvador, 2023.

1.cópula BB9, cópula PVF, família Arquimediana, medidas de con-
cordância. I. Silva, Paulo Henrique Ferreira da; Fiaccone, Rosemeire
Leovigildo. II. Universidade Federal da Bahia. Instituto de Matemática e
Estatística. III. Título.

CDD - 519.72

CDU - 519.72 vtres

CÓPULA PVF: REVISÃO DE LITERATURA E NOVOS RESULTADOS

SUÊDE SANTOS BARBOSA

Dissertação de Mestrado apresentada ao Colegiado da Pós-Graduação em Matemática da Universidade Federal da Bahia (UFBa), como parte dos requisitos para obtenção do título de Mestre em Matemática. Área de concentração: Estatística.

Banca Examinadora:



Prof. Dr. Paulo Henrique Ferreira da Silva (Orientador)
IME - UFBA

Profa. Dra. Giovana Oliveira Silva
IME - UFBA



Prof. Dr. Adriano Kamimura Suzuki
ICMC - USP

Ao meu pai João Carlos (in memoriam), à minha mãe Sônia e irmão Josimar.

Agradecimentos

A Deus, pela minha vida, por todas as oportunidades que me concedeu e pelas pessoas que tem colocado em meu caminho.

Aos meus pais, João Carlos e Sônia, pelo amor incondicional e por tudo que fizeram em prol da minha educação. Ao meu irmão Josimar, pelo apoio em vários momentos e incentivo. Suas palavras de encorajamento e compreensão foram fundamentais para superar os desafios encontrados durante a elaboração deste trabalho.

Ao professor Paulo Henrique, por TODA orientação, pelo conhecimento transmitido, pela imensa paciência, dedicação e preocupação. À professora Rosemeire, pela coorientação deste trabalho e orientações transmitidas. Suas sugestões valiosas e conhecimento foram fundamentais para o desenvolvimento desta pesquisa.

À minha querida amiga e irmã Meire, por todo apoio e carinho. Ao meu *best* Jorge Augusto, pelo incentivo e correções feitas.

Por fim, agradeço a todas as fontes de inspiração que encontrei ao longo do caminho no Instituto de Matemática e Estatística da Universidade Federal da Bahia, sejam elas autores de estudos relevantes, professores, pesquisadores ou profissionais da área. Suas contribuições acadêmicas e profissionais foram fundamentais para a construção do conhecimento apresentado nesta dissertação.

Expresso minha gratidão a todos os envolvidos, direta ou indiretamente, por seu apoio e contribuições ao longo deste processo. Cada um de vocês teve um papel importante na realização deste trabalho.

Muito obrigada!

Suêde S. Barbosa

*“Quanto mais aumenta nosso
conhecimento, mais evidente fica
nossa ignorância.”
(John F. Kennedy)*

Resumo

Toda distribuição de probabilidade conjunta traz consigo informações a respeito do comportamento individual de cada variável (distribuições marginais) e da estrutura de dependência que guia a relação entre elas. A modelagem estatística por meio de funções cópula proporciona uma análise individual, possibilitando o estudo minucioso das estruturas de associação que orientam um vetor aleatório. Uma cópula de interesse (e objeto de estudo desta dissertação) é a cópula *Power Variance Function* (PVF). Ela parte da distribuição PVF univariada de três parâmetros, derivada como uma extensão da distribuição estável positiva. A classe de cópulas PVF abarca uma família de cópulas Arquimedianas, que inclui as cópulas de Clayton, Gumbel e Gaussiana Inversa como casos especiais. Inicialmente, neste trabalho foi exposta uma revisão geral sobre cópulas, contendo as principais definições, propriedades básicas, teorema de Sklar, limitantes de Fréchet-Hoeffding, medidas de dependência/associação e famílias paramétricas, bem como identificar os principais resultados conhecidos a respeito da cópula PVF, tanto no caso bivariado como multivariado. Além das medidas de dependência usuais (τ de Kendall e ρ de Spearman), outras medidas como β de Blomqvist e γ de Gini foram investigadas; neste sentido, simulações das medidas de dependência da cópula PVF com diferentes valores (isto é, diferentes combinações de parâmetros) foram realizadas. Em adição, foram desenvolvidas discussões sobre a relação da cópula BB9 com a cópula PVF, apresentando também uma demonstração de que a cópula PVF é uma subclasse da classe de cópulas Arquimedianas. Por fim, foram expostos três métodos para a geração de dados a partir da cópula PVF e seus casos especiais, e realizados um estudo de simulação e um estudo de caso (aplicação a dados reais).

Palavras-chave: cópula BB9, cópula PVF, família Arquimediana, medidas de concordância.

Abstract

Every joint probability distribution brings with it information about the individual behavior of each variable (marginal distributions) and the dependency structure that guides the relationship between them. Statistical modeling through copula functions provides an individual analysis of these elements, enabling a detailed study of the association structures that guide a random vector. A copula of interest (and object of study of this dissertation) is the *Power Variance Function* (PVF) copula. It starts from the univariate three-parameter PVF distribution, derived as an extension of the positive stable distribution. The PVF copula class comprises a family of Archimedean copula that includes Clayton, Gumbel and Inverse Gaussian copulas as special cases. Through a methodological approach of literature review, it was possible to present in this work a general review of copulas, containing the main definitions, basic properties, Sklar's theorem, Fréchet-Hoeffding bounds, dependence/association measures and parametric families, as well as identifying the main known results regarding the PVF copula, both in the bivariate and multivariate cases. In addition to the usual dependency measures (Kendall's τ and Spearman's ρ), other measures like Blomqvist's β and Gini's γ were investigated; in this sense, simulations of the PVF copula dependence measures with different values (i.e., different combinations of parameters) were performed. In addition, discussions were developed about the relationship of the BB9 copula with the PVF copula; specifically, we presented a demonstration that the PVF copula is a subclass of the Archimedean copula. Finally, three methods for generating data from the PVF copula and its special cases, as well as simulation and case studies, were provided.

Keywords: BB9 copula, PVF copula, Archimedean family, concordance measures.

Sumário

1	Introdução	1
1.1	Objetivos	1
1.2	Organização do Trabalho	1
2	Conceitos Básicos	2
2.1	Definição e Propriedades	3
2.2	Medidas de Dependência e de Concordância	5
2.3	Dependência Caudal	8
3	Cópula <i>Power Variance Function</i> (PVF)	10
3.1	Distribuição PVF	10
3.1.1	Transformada de Laplace, Casos Especiais e Limitante	11
3.2	Cópula PVF	13
3.2.1	Cópula PVF e Cópulas Arquimedianas	18
3.2.2	Cópula PVF Bivariada <i>versus</i> Cópula BB9 Bivariada	19
3.3	Densidade da Cópula PVF	20
3.4	Cópula Condicional Bivariada da PVF	21
3.5	Classificação da Cópula PVF	22
3.6	Medidas de Concordância	23
3.6.1	Dependência de Cauda	27
3.6.2	Função <i>Cross-Ratio</i>	31
4	Estimação de Cópulas	32
4.1	Estimação por Máxima Verossimilhança	32
4.2	Método de Inferência para as Marginais (IFM)	34
5	Simulação e Aplicação	36
5.1	Abordagem pela Distribuição Condicional	36
5.2	Algoritmo para Cópulas Arquimedianas em Geral	37
5.3	Abordagem de Marshall e Olkin	38

5.4	Simulação	39
5.4.1	Resultados e Discussão	40
5.5	Aplicação	41
6	Considerações Finais	43
	Referências	45

Lista de Figuras

3.1	Gráfico da FDP da distribuição PVF (Mallick e Ravishanker, 2005), considerando $k = 10$, $\alpha = 0,5$, $\delta = 2$ e $\theta = 1$	12
3.2	Gráfico da FDP da distribuição PVF (Mallick e Ravishanker, 2005), em que foi fixado o parâmetro $\alpha = 0,4$ e variou-se os parâmetros δ e θ	12
3.3	Gráficos e curvas de nível da cópula PVF, considerando diferentes combinações de valores dos parâmetros α e η	16
3.4	Gráficos da densidade das cópulas de Clayton (painel superior esquerdo), Gumbel (painel superior direito) e Gaussiana Inversa (painel inferior).	21
3.5	Gráficos de dispersão da cópula PVF com: $\tau = 0,9$ (painel superior esquerdo), $\tau = 0,7$ (painel superior direito), $\tau = 0,5$ (painel inferior esquerdo) e $\tau = 0,2$ (painel inferior direito).	24
3.6	Gráficos de dispersão da cópula PVF, considerando amostras de tamanho 1.000 e $\tau = 0,5$, além de diferentes combinações de α e η (painéis superiores). Nos painéis inferiores (da esquerda para a direita), $\tau = 0,33$; $0,20$ e $0,05$	29
3.7	Gráficos de dispersão da cópula PVF, considerando amostras de tamanho 1.000 e $\tau = 0,9$, além de diferentes combinações de α e η (painéis superiores). Nos painéis inferiores, $\tau = 0,7$, considerando diferentes combinações de α e η	29
3.8	Gráficos de dispersão das cópulas de Gumbel e Clayton, para $\tau = 0,4$ e $0,8$	30
5.1	Gráfico de dispersão das variáveis Co e Sc.	42

Lista de Tabelas

3.1	Função geradora e inversa da geradora da cópula PVF e de seus casos especiais e limitante.	17
3.2	Relação entre os parâmetros das cópulas BB9 e PVF.	20
3.3	Valores de β_C , considerando diferentes combinações dos parâmetros α e η da cópula PVF.	26
3.4	Valores de γ_C , considerando diferentes combinações dos parâmetros α e η da cópula PVF.	26
3.5	Função <i>cross-ratio</i> para os casos especiais e limite da cópula PVF.	31
5.1	Média, vício e EQM das estimativas dos parâmetros do modelo de cópula PVF (caso bivariado).	40
5.2	Coefficiente τ de Kendall para todas as variáveis do conjunto de dados <i>uranium</i> , tomadas duas a duas.	41
5.3	Valores AIC dos modelos PVF, Clayton, Gumbel e Gaussiana Inversa.	42

Capítulo 1

Introdução

Na área da Estatística, uma cópula desempenha um papel crucial ao formular uma distribuição multivariada que pode representar diversos tipos de dependência. A construção dessa distribuição envolve a utilização de uma cópula juntamente com uma transformação simples aplicada a cada variável. Essa transformação tem como objetivo garantir que cada variável, após ser transformada marginalmente, siga uma distribuição uniforme. Uma vez realizada essa transformação marginal, a estrutura de dependência pode ser expressa como uma distribuição multivariada sobre as variáveis uniformes obtidas. Em essência, uma cópula é precisamente isso: uma distribuição multivariada sobre variáveis aleatórias uniformes. (Nelsen, 2006)

Nos últimos anos tem crescido o interesse por modelos que incorporem mais informações no sentido de modelagem de eventos conjuntos. Como consequência, o uso da teoria de cópulas tem se tornado popular, uma vez que permite ligar distribuições marginais univariadas para formar uma distribuição conjunta. As cópulas têm sido bastante utilizadas para a modelagem de dependência entre dados multivariados, principalmente nas áreas biológicas, ciências atuariais e finanças (ver, por exemplo, Embrechts et al., 2003; Cherubini et al., 2004; Nelsen, 2006; Kolev et al., 2006; Jaworski, 2010). Gakes (2001) e Goethals et al. (2008) exibem algumas relações entre modelos com fragilidade e modelos de cópulas.

Já que uma cópula é uma função que liga as distribuições marginais das variáveis aleatórias à sua distribuição conjunta, em diversos casos, as distribuições marginais são conhecidas (ou podem ser estimadas) e a distribuição conjunta é desconhecida (ou pode ser difícil de estimar).

A cópula de interesse do presente estudo é a cópula *Power Variance Function*, também conhecida como cópula PVF. Essa é uma cópula versátil, por possuir casos particulares (Clayton, Gumbel e Gaussiana Inversa) comumente utilizados nos estudos de modelagem. Apesar dessa versatilidade, diante dos materiais pesquisados, pouco se en-

contra sobre essa cópula, como, por exemplo, suas medidas de dependência (além do τ de Kendall). Este estudo realizou uma revisão bibliográfica sobre a cópula PVF, de forma a reunir todos os materiais (desde o ano de 2017) que desenvolveram estudos dessa cópula ou qualquer menção sobre a mesma.

1.1 Objetivos

Neste trabalho será apresentada uma análise do comportamento da cópula PVF e dos seus casos especiais e limitante, das medidas de dependência usuais (τ de Kendall e ρ de Spearman) e de outras medidas como β de Blomqvist e γ de Gini. Além disso, será investigada a relação da cópula BB9 com a cópula PVF, buscando entender se essas cópulas são iguais e, caso sejam, qual a correspondência entre seus parâmetros. Por fim, será exibida uma demonstração de que a cópula PVF é uma subclasse da classe de cópulas Arquimedianas.

1.2 Organização do Trabalho

O presente trabalho está dividido em sete capítulos, sendo estruturado da seguinte forma:

- O Capítulo 2 contém uma breve revisão dos conceitos básicos de funções cópula, bem como suas medidas de dependência e concordância;
- O Capítulo 3 apresenta a distribuição PVF e seus casos especiais, bem como a cópula PVF e suas medidas de dependência e concordância. Também prova-se que a cópula PVF é uma subclasse da classe de cópulas Arquimedianas e discute-se a relação entre a cópula PVF bivariada e a cópula BB9 bivariada;
- O Capítulo 4 aborda dois métodos comumente utilizados para a estimação de modelos de cópulas: máxima verossimilhança e inferência para as marginais;
- O Capítulo 5 inicia com a exposição de três diferentes abordagens para a geração de dados utilizando a cópula PVF e seus casos especiais e limitante; em seguida, apresenta um estudo de simulação feito para avaliar o comportamento dos estimadores dos parâmetros da cópula PVF, bem como uma aplicação a dados reais, para analisar a qualidade do modelo PVF e seus casos particulares;
- O Capítulo 6 apresenta as considerações finais deste trabalho, além de propostas de trabalhos futuros.

Capítulo 2

Conceitos Básicos

Este capítulo inicia com a definição de cópula, bem como um breve contexto histórico. Em seguida, apresenta-se a definição de cópula bivariada e multivariada, o teorema de Sklar e os limitantes de Fréchet-Hoeffding. Por fim, são descritas as principais medidas de dependência: coeficiente de correlação linear, τ de Kendall, ρ de Spearman, β de Blomqvist, γ de Gini e dependência de cauda.

De acordo com os dicionários Michaelis (2008) e Priberam (2021), a palavra cópula origina-se do latim *copula*, possuindo os seguintes significados:

- i) união, junção;
- ii) ato sexual, coito, copulação.

De acordo com Lloréns (2017), a cópula é “o que resta” de uma função de distribuição conjunta, dada a exclusão do comportamento marginal das variáveis aleatórias.

A primeira referência em um artigo de Estatística, usando a palavra “cópula” no título ou como uma palavra-chave, ocorreu em 1981, por Schweizer e Wolff (ver Schweizer e Wolff, 1981).

Sob a perspectiva estatística e matemática, Nelsen (2006) define cópulas inicialmente de maneira simples, indicando elas como funções que juntam ou acoplam funções de distribuição multivariadas (ou multidimensionais) às suas funções de distribuição marginais univariadas (ou unidimensionais). De forma alternativa, ele compreende cópulas como sendo funções de distribuição multivariadas, cujas marginais unidimensionais são uniformes no intervalo $(0, 1)$.

De acordo com Fischer (1997), as cópulas são de interesse de estatísticos, visto que, por meio delas, é possível estudar medidas de dependência livres de escala (*free-scale*); e também por elas constituírem um ponto de partida para a construção de famílias de distribuições multivariadas.

2.1 Definição e Propriedades

Nesta seção, baseada em Cherubini (2004) e Nelsen (2006), será apresentada a definição analítica de função cópula e suas propriedades fundamentais, para os casos bivariado e multivariado.

Cópula Bivariada

Uma **cópula bivariada** (**cópula bidimensional** ou **2-cópula**) é uma função de distribuição bivariada $C : [0, 1]^2 \rightarrow [0, 1]$ com distribuições marginais $U(0, 1)$, isto é, $C(u, v) = \mathbb{P}(U \leq u, V \leq v)$, com $U, V \sim U(0, 1)$.

Uma cópula possui as seguintes propriedades:

- (1) $\forall u \in [0, 1], C(u, 0) = 0 = C(0, u)$;
- (2) $\forall u \in [0, 1], C(u, 1) = u$ e $C(1, u) = u$;
- (3) Desigualdade retangular: $\forall u_1, u_2, v_1, v_2 \in [0, 1]$, tais que $u_1 \leq u_2$ e $v_1 \leq v_2$, tem-se:

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

Cópula Multivariada

Sejam S_1, \dots, S_d subconjuntos não vazios de \mathbb{R} , H uma função real de d variáveis tal que o domínio de H é $DomH = S_1 \times S_2 \times \dots \times S_d$, e seja também $B = [\mathbf{a}, \mathbf{b}] = ([a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d])$ um cubo d -dimensional em que os vértices estão contidos em $DomH$ e $\mathbf{a} \leq \mathbf{b}$, para todo $a_k \leq b_k, k = 1, 2, \dots, d$. Então, o H -volume sobre o cubo d -dimensional, $B = [\mathbf{a}, \mathbf{b}]$, é dado pela aplicação sucessiva do operador diferença, isto é,

$$\text{Vol}_H(B) = \Delta_{\mathbf{a}}^{\mathbf{b}} H(\mathbf{x}) = \Delta_{a_1}^{b_1} \Delta_{a_2}^{b_2} \dots \Delta_{a_d}^{b_d} H(x_1, x_2, \dots, x_d),$$

em que os pontos sobre os quais são aplicadas as diferenças são os vértices do cubo d -dimensional.

Uma **cópula multivariada** é uma função de distribuição d -dimensional $C : [0, 1]^d \rightarrow [0, 1]$, em que as marginais são $U(0, 1)$, e satisfaz as seguintes propriedades:

- (1) $C(u_1, \dots, u_d)$ é crescente e contínua em cada um de seus componentes u_i , para todo $i = 1, \dots, d$;
- (2) $\forall \mathbf{u} \in [0, 1]^d, C(\mathbf{u}) = 0$ se pelo menos uma coordenada de \mathbf{u} tende a zero, e $C(\mathbf{u}) = 1$ se todas as coordenadas de \mathbf{u} tenderem para 1;
- (3) $\text{Vol}_{C(S_1, \dots, S_d)} \geq 0, \forall S_k = (a_k, b_k] \subseteq I = [0, 1], k = 1, \dots, d$.

Um importante teorema que relaciona as distribuições multivariadas com as funções cópula é o de Sklar.

O teorema de Sklar estabelece que qualquer função de distribuição multivariada pode ser decomposta em duas partes: a função de distribuição marginal de cada variável e uma cópula que descreve a estrutura de dependência entre as variáveis.

Essa decomposição é extremamente útil, pois permite tratar separadamente a modelagem das marginais e a modelagem da dependência. As marginais podem ser ajustadas independentemente, usando distribuições univariadas adequadas para cada variável. A cópula, por sua vez, pode ser escolhida de acordo com a forma desejada de dependência, permitindo uma flexibilidade maior no modelo.

Teorema de Sklar (Sklar, 1959): Seja $H(x_1, x_2, \dots, x_d)$ uma função de distribuição multivariada (ou d -variada) com distribuições marginais $F_1(x_1), F_2(x_2), \dots, F_d(x_d)$. Se estas forem todas contínuas, então existe uma única cópula $C : [0, 1]^d \rightarrow [0, 1]$, tal que:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

para todo $(x_1, \dots, x_d) \in (-\infty, +\infty)^d$. Considerando $F_j(x_j) = u_j$, com $u_j \in [0, 1]$, obtém-se $x_j = F_j^{-1}(u_j)$, $j = 1, 2, \dots, d$. Daí, segue a cópula:

$$C(u_1, \dots, u_d) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

para todo $(u_1, \dots, u_d) \in [0, 1]^d$.

No caso bidimensional, tem-se:

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \implies C(u_1, u_2) = H(F_1^{-1}(u_1), F_2^{-1}(u_2)).$$

Utilizando o teorema de Sklar, é possível separar a distribuição bidimensional H em duas partes:

- O conjunto das distribuições marginais: $F_1(x_1)$ e $F_2(x_2)$;
- A cópula C , que contém as informações sobre a estrutura de dependência de H .

Os Limitantes de Fréchet-Hoeffding são propriedades teóricas importantes que se aplicam às cópulas. Esses limitantes estabelecem condições sobre a função de cópula que garantem a sua viabilidade e coerência.

Para qualquer cópula existem dois limites entre os quais a cópula é verificada. O primeiro limitante, chamado de limite inferior de Fréchet (W), estabelece que qualquer cópula deve ser não decrescente em cada uma de suas variáveis marginais. O segundo

limitante é o limite superior de Fréchet (M), o qual estabelece que qualquer cópula deve ser limitada superiormente por uma função linear (Nelsen, 2006).

Se C é uma cópula d -dimensional qualquer, então segue que:

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}),$$

com:

$$W(\mathbf{u}) = \max \left\{ \sum_{j=1}^d u_j - d + 1, 0 \right\} \quad \text{e} \quad M(\mathbf{u}) = \min_{1 \leq j \leq d} \{u_j\},$$

para $\mathbf{u} \in [0, 1]^d$.

No caso da cópula bivariada, segue que:

$$W(u, v) = \max \{u + v - 1, 0\} \quad \text{e} \quad M(u, v) = \min\{u, v\}.$$

2.2 Medidas de Dependência e de Concordância

As medidas de dependência quantificam o grau de relação estatística entre as variáveis aleatórias. Cada medida de dependência aborda uma particularidade da interação entre as variáveis, verificando a força e a direção da relação entre elas. Essas medidas variam de -1 a 1, em que -1 indica uma dependência negativa perfeita, 1 indica uma dependência positiva perfeita e 0 indica ausência de dependência.

As medidas de concordância são definidas por Scarsini (1984) como sendo medidas que satisfazem algumas propriedades necessárias para examinar uma provável associação entre variáveis aleatórias. Essas medidas são utilizadas para avaliar a concordância ou similaridade entre as variáveis aleatórias, medindo o grau de concordância nas ordens das observações entre as variáveis, conseqüentemente mensurando o ajuste da cópula aos dados e sua capacidade de modelar a relação conjunta entre as variáveis. A ideia intuitiva dessa definição é que duas variáveis aleatórias, X_1 e X_2 , são concordantes quando altos (baixos) valores de X_1 estão associados com altos (baixos) valores de X_2 .

Tais medidas são mais robustas e invariantes a transformações estritamente crescentes das variáveis aleatórias; dessa forma, apresentam menos limitações em comparação às medidas de dependência. Além disso, as medidas de concordância não levam em consideração as informações das marginais das variáveis, focando exclusivamente na concordância relativa entre os valores observados.

A seguir é apresentada/descrita uma medida de dependência (coeficiente de Pearson) e, posteriormente, medidas de concordância (τ de Kendall, ρ de Spearman, β de Blomqvist e γ de Gini).

Coeficiente de Correlação de Pearson

O coeficiente de correlação linear, ou coeficiente de Pearson, é uma medida de dependência paramétrica. É importante destacar que ele não é uma medida de concordância, já que não satisfaz as propriedades de simetria e normalidade (ver seção “Concordância”). Esse coeficiente mensura apenas a relação linear entre as variáveis e não captura outras formas de associação, como relações não lineares ou dependências mais complexas.

Seja (X, Y) um vetor de variáveis aleatórias com variâncias finitas. O coeficiente de correlação linear de (X, Y) é definido como:

$$r = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \text{Var}[Y]}}$$

em que $\text{Cov}[X, Y]$ é a covariância entre X e Y , $\text{Var}[X]$ e $\text{Var}[Y]$ são as variâncias de X e Y , respectivamente.

Esse coeficiente varia no intervalo de -1 a 1, em que -1 indica uma correlação perfeitamente negativa, 0 indica ausência de correlação e 1 indica uma correlação perfeitamente positiva.

Concordância

Sejam (x_1, y_1) e (x_2, y_2) duas observações do vetor (X, Y) de variáveis aleatórias contínuas. Então, (x_1, y_1) e (x_2, y_2) são ditas concordantes quando $(x_1 - x_2)(y_1 - y_2) > 0$ e discordantes quando $(x_1 - x_2)(y_1 - y_2) < 0$.

Então, qualquer função r de uma distribuição bivariada de variáveis aleatórias (X, Y) é uma medida de concordância se satisfaz as seguintes propriedades:

- (1) Simetria: $r(X, Y) = r(Y, X)$;
- (2) Normalidade: $-1 \leq r(X, Y) \leq 1$;
- (3) Comonotonicidade: $r(X, Y) = 1$ se, e somente se, X e Y são comonotônicas, ou seja, têm correlação positiva máxima.

τ de Kendall

O τ de Kendall é uma medida definida em termos da concordância, pela diferença entre a probabilidade dos pares concordantes e discordantes.

A versão populacional do τ de Kendall é obtida por meio da função de distribuição de uma cópula d -dimensional C da seguinte forma:

$$\tau_d = \frac{1}{2^{d-1} - 1} \left(2^d \int \dots \int_{[0,1]^d} C(u_1, \dots, u_d) dC(u_1, \dots, u_d) - 1 \right),$$

com $u_1, \dots, u_d \in [0, 1]$.

Para $d = 2$, obtém-se:

$$\tau_2 = 4 \int_0^1 \int_0^1 C(u_1, u_2) c(u_1, u_2) du_1 du_2 - 1 = 4 \mathbb{E}[C(U, V)] - 1,$$

ρ de Spearman

O coeficiente de correlação de postos de Spearman (ou ρ de Spearman) representa uma das medidas mais conhecidas para quantificar o grau de associação entre duas ou mais variáveis e foi primeiramente estudado por Spearman (1904).

Cherubini (2004) indica que o ρ de Spearman pode ser interpretado como um valor esperado normalizado e que representa correlação de classificação.

Seja (X, Y) um vetor aleatório com cópula associada C . Então, a versão populacional do ρ de Spearman (ou coeficiente de Spearman), $\rho(X, Y)$, do vetor aleatório (X, Y) , em termos da cópula equivalente C , é dada por:

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3. \quad (2.1)$$

A versão amostral ρ_n de $\rho(X, Y)$ é dada por:

$$\rho_n = \frac{12}{n^3 - n} \sum_{i=1}^n R_{X_i} R_{Y_i} - \frac{3(n+1)}{n-1},$$

em que R_{X_i} e R_{Y_i} são os postos de X_i e Y_i , respectivamente.

A versão multivariada (d -dimensional) do coeficiente ρ de Spearman é dada por:

$$\rho(C) = h_\rho(d) \left\{ 2^d \int_{[0,1]^d} C(\mathbf{u}) d\mathbf{u} - 1 \right\},$$

em que $h_\rho(d) = (d+1)/[2^d - (d+1)]$.

β de Blomqvist

A medida β de Blomqvist é também chamada de coeficiente de correlação medial. No caso bivariado, o β de Blomqvist de uma cópula C é definido no centro de $[0, 1]^2$ como β_C e expresso como:

$$\beta_C = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1.$$

No caso multivariado (ou d -variado), o β de Blomqvist é dado por:

$$\frac{2^d C(1/2, 1/2) - 1}{2^{d-1} - 1},$$

em que d refere-se à dimensão.

Embora o β de Blomqvist dependa da cópula apenas através de seu valor no centro de $[0, 1]^2$, ele pode, ainda assim, fornecer uma aproximação precisa tanto para o ρ de Spearman quanto para o τ de Kendall.

γ de Gini

De acordo com Jaworski et al. (2009), define-se a medida γ de Gini para o vetor (X, Y) como:

$$\gamma_G = 2 \mathbb{E} [|U + V - 1| - |U - V|],$$

em que $U = F(X)$ e $V = G(Y)$, com F e G denotando as funções de distribuição de X e Y , respectivamente.

A medida γ de Gini ainda pode ser reescrita da seguinte forma:

$$\gamma_G = 2 \iint_{[0,1]^2} (|u + v - 1| - |u - v|) dC(u, v) = 4 \iint_{[0,1]^2} [M(u, v) + W(u, v)] dC(u, v) - 2.$$

Ubeda-Flores (2022) apresenta outra maneira de definir o coeficiente γ de Gini:

$$\gamma(C) = 4 \int_0^1 [C(u, u) + C(u, 1 - u)] du - 2. \quad (2.2)$$

Para a extensão multivariada do γ de Gini, define-se a função $A(\mathbf{u}) = [M(\mathbf{u}) + W(\mathbf{u})] / 2$, $\mathbf{u} \in [0, 1]^d$, com a função de sobrevivência correspondente \bar{A} . Dessa forma, o coeficiente será dado por:

$$\gamma(C) = \frac{1}{b(d) - a(d)} \left[\int_{[0,1]^d} [A(\mathbf{u}) + \bar{A}(\mathbf{u})] dC(\mathbf{u}) - a(d) \right],$$

em que as constantes de normalização $a(d)$ e $b(d)$ são dadas por:

$$a(d) = \frac{1}{d+1} + \frac{1}{2(d+1)!} + \sum_{i=0}^d (-1)^i \binom{d}{i} \frac{1}{(i+1)!}$$

e

$$b(d) = 1 - \sum_{i=1}^{d-1} \frac{1}{4^i}.$$

2.3 Dependência Caudal

Os coeficientes de dependência nas caudas são medidas de associação locais que mensuram o grau de dependência entre duas variáveis aleatórias reais quando elas assumem ao mesmo tempo valores extremos (muito elevados ou muito pequenos). Em outras palavras, a dependência caudal descreve como as probabilidades extremas em uma cópula estão relacionadas à dependência entre as variáveis aleatórias.

Existem algumas famílias de cópulas Arquimedianas que possibilitam a modelagem de diferentes níveis de dependência tanto para caudas inferiores quanto superiores. Essa característica é especialmente relevante em aplicações financeiras, uma vez que eventos extremos negativos geralmente apresentam um efeito distinto dos eventos extremos positivos. Com efeito, a assimetria na dependência das caudas manifesta-se como um elemento fundamental na escolha de uma cópula para a modelagem em cenários financeiros.

O coeficiente de dependência caudal é derivado como uma probabilidade condicional, de modo que esteja entre 0 e 1. Além disso, é invariante sob transformações estritamente crescentes de X e Y (Embrechts et al., 2001).

Sejam X e Y variáveis aleatórias com funções de distribuição $u = F(x)$ e $v = G(y)$, respectivamente, e distribuição conjunta dada pela cópula C . Os coeficientes de dependência nas caudas superior e inferior, λ_U e λ_L , respectivamente, podem ser definidos como:

$$\lambda_U = 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} \quad \text{e} \quad \lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u},$$

se ambos os limites existirem.

Os coeficientes de dependência na cauda têm as seguintes propriedades:

- (1) λ_U e $\lambda_L \in [0, 1]$;
- (2) Se $\lambda_U \in]0, 1]$, então X e Y são assintoticamente correlacionados na cauda superior.
Se $\lambda_U = 0$, então X e Y são assintoticamente independentes na cauda superior;
- (3) Se $\lambda_L \in]0, 1]$, então X e Y são assintoticamente correlacionados na cauda inferior.
Se $\lambda_L = 0$, então X e Y são assintoticamente independentes na cauda inferior.

Capítulo 3

Cópula *Power Variance Function* (PVF)

Neste capítulo serão apresentadas três versões da função densidade de probabilidade da distribuição PVF, sua transformada de Laplace, casos especiais e limitante. Posteriormente, será exposta a cópula PVF, sua densidade bivariada, condicionais bivariadas e medidas de concordância. Dois resultados do presente estudo serão apresentados neste capítulo: a prova de que a cópula PVF é uma subclasse da classe de cópulas Arquimedianas e as relações entre os parâmetros da cópula BB9 bivariada e os parâmetros da cópula PVF.

3.1 Distribuição PVF

A distribuição *Power Variance Function*, também conhecida como família PVF ou função de variância de potência, é uma família de distribuições estatísticas que foi proposta como uma alternativa flexível para modelar dados com caudas pesadas e assimetria. Por possuir a variância como sendo uma função de potência da média, essa família recebeu o nome de *Power Variance Function* (Hougaard, 2000).

O estudo da distribuição PVF teve início em 1982 com Morris. Posteriormente, em 1984, Tweedie apresentou uma análise mais abrangente dessa família, especificamente a versão univariada de três parâmetros. Em 1986, Hougaard derivou a família PVF como uma extensão da distribuição estável positiva.

Já em 2006, Mallick e Ravishanker desenvolveram uma abordagem Bayesiana para o modelo de fragilidade PVF, considerando distribuições exponenciais para o risco basal. Hougaard (2000), Duchateau e Janssen (2008) e Wienke (2010) também exploraram a família PVF utilizando abordagens clássicas em suas pesquisas.

Por fim, Hanagal (2009) apresentou uma mistura bivariada de Weibull-PVF. To-

dos esses marcos históricos contribuíram significativamente para o avanço do estudo da distribuição PVF.

A distribuição PVF univariada é uma família de distribuições estatísticas utilizada para modelar uma única variável aleatória, caracterizada por ter a variância como uma função de potência da média. Essa distribuição apresenta três parâmetros, α , δ e θ , sendo denotada por $PVF(\alpha, \delta, \theta)$. (Romeo, 2017 e Biondo, 2020)

Seja W uma variável aleatória com distribuição $PVF(\alpha, \delta, \theta)$, então a função densidade de probabilidade (FDP) da PVF, descrita por Hougaard (1986), é expressa da seguinte forma:

$$f(w|\alpha, \delta, \theta) = -\frac{1}{\pi w} \exp\left\{\frac{-\theta w + \delta\theta^\alpha}{\alpha}\right\} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} \left(\frac{-w^{-\alpha}\delta}{\alpha}\right)^k \sin(\alpha k\pi),$$

para $w > 0$, $0 < \alpha < 1$, $\delta > 0$ e $\theta > 0$.

Considerando que a função gama pode ser escrita como $\Gamma(k + 1) = k!$, Mallick e Ravishanker (2005) reescreveram a FDP da PVF como:

$$f(w|\alpha, \delta, \theta) = -\frac{1}{\pi w} \exp\left\{\frac{-\theta w + \delta\theta^\alpha}{\alpha}\right\} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{\Gamma(k + 1)} \left(\frac{-w^{-\alpha}\delta}{\alpha}\right)^k \sin(\alpha k\pi). \quad (3.1)$$

Para valores fixos de α e δ , a família PVF é dita exponencial natural em θ . Para α fixo, ela é um modelo exponencial de dispersão (Hougaard, 1986).

A FDP (3.1) é a mais utilizada na literatura, mas uma outra versão dessa função também foi apresentada por Aalen (1992). Essa FDP é dada por:

$$f_U(u) = \exp\left\{-\frac{\nu}{\theta} \left(\frac{u}{\mu} + \frac{1}{\nu - 1}\right)\right\} \times \frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} (u/\mu)^{k(\nu-1)} \Gamma(1 - k(\nu - 1)) \sin(\pi k(\nu - 1))}{k!(\nu - 1)^k},$$

com $\mu > 0$, $\theta > 0$ e $0 < \nu \leq 1$.

A Figura 3.1 apresenta o gráfico da FDP da distribuição PVF descrita por Mallick e Ravishanker (2005), considerando $k = 10$, $\alpha = 0,5$, $\delta = 2$ e $\theta = 1$. Já a Figura 3.2 exhibe o gráfico das curvas de nível da FDP da distribuição PVF (Mallick e Ravishanker, 2005), em que foi fixado o parâmetro $\alpha = 0,4$ e variou-se os parâmetros δ e θ .

3.1.1 Transformada de Laplace, Casos Especiais e Limitante

Considerando $W \sim PVF(\alpha, \delta, \theta)$, Hougaard (1986) mostrou que a transformada de Laplace da PVF, $\mathcal{L}_W(s) = \mathbb{E}[\exp\{-sW\}]$, é dada por:

$$\mathcal{L}_W(s) = \exp\left\{-\left(\frac{\delta}{\alpha}\right) [(\theta + s)^\alpha - \theta^\alpha]\right\}. \quad (3.2)$$

Calculando a primeira e segunda derivadas de (3.2), tem-se, respectivamente:

$$\mathcal{L}_W^{(1)}(s) = \mathcal{L}_W(s)(-\delta)(\theta + s)^{\alpha-1}$$

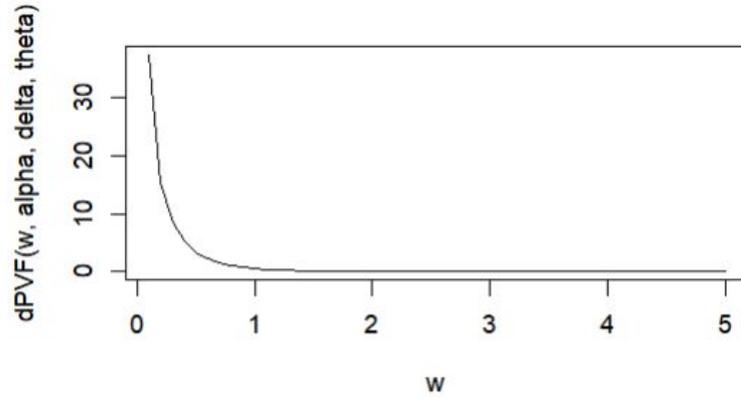


Figura 3.1: Gráfico da FDP da distribuição PVF (Mallick e Ravishanker, 2005), considerando $k = 10$, $\alpha = 0,5$, $\delta = 2$ e $\theta = 1$.

Fonte: Autoria própria.

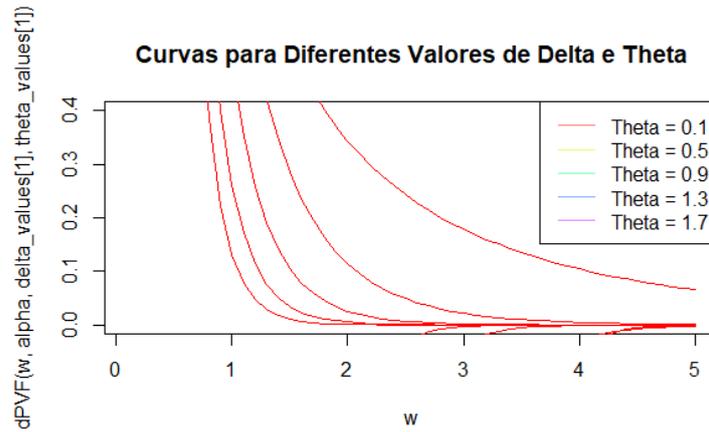


Figura 3.2: Gráfico da FDP da distribuição PVF (Mallick e Ravishanker, 2005), em que foi fixado o parâmetro $\alpha = 0,4$ e variou-se os parâmetros δ e θ .

Fonte: Autoria própria.

e

$$\mathcal{L}_W^{(2)}(s) = \delta^2 \mathcal{L}_W(s)(\theta + s)^{2\alpha-2} - \delta \mathcal{L}_W(s)(\alpha - 1)(\theta + s)^{\alpha-2}.$$

Se $\theta > 0$, todos os momentos existem e os valores da esperança e variância são expressos, respectivamente, por:

$$\mathbb{E}[W] = (-1)\mathcal{L}_W^{(1)}(0) = \delta\theta^{\alpha-1}$$

e

$$\text{Var}[W] = \mathcal{L}_W^{(2)}(0) - \left[-\mathcal{L}_W^{(1)}(0)\right]^2 = \delta(1 - \alpha)\theta^{\alpha-2}.$$

A partir da escolha específica dos parâmetros na distribuição PVF, são obtidos dois casos especiais:

- **Distribuição Estável Positiva:** Se $\theta = 0$ e $\delta = \alpha$, então $W \sim \text{Estável Positiva}(\alpha)$, sendo $\mathcal{L}_W(s) = \exp\{-s^\alpha\}$. Além disso, $\mathbb{E}[W] = 0$ e $\text{Var}[W] = 0$;
- **Distribuição Gaussiana Inversa:** Se $\alpha = 1/2$, então $W \sim \text{Gaussiana Inversa}(\delta, \theta)$, com $\mathcal{L}_W(s) = \exp\{-2\delta[(\theta + s)^{1/2} - \theta^{1/2}]\}$. Ademais, $\mathbb{E}[W] = \delta\theta^{-1/2}$ e $\text{Var}[W] = \delta\theta^{-1}(\delta + \theta^{-1/2}/2 - 1)$.

Além de indicar os casos especiais acima, Hougaard (1986) apresentou outro caso, o qual denominou de caso limite ou limitante:

- **Distribuição Gama:** Se $\alpha \rightarrow 0$, então $W \sim \text{Gama}(\delta, \theta)$, sendo $\mathcal{L}_W(s) = \theta^\delta(s + \theta)^{-\delta}$. Aqui, $\mathbb{E}[W] = \delta\theta^{-1}$ e $\text{Var}[W] = \delta$.

A transformada de Laplace da família PVF, expressa por Aalen (1992), é dada por:

$$\mathcal{L}(s) = \exp\left\{\frac{\nu}{\theta(1-\nu)}\left[1 - \left(1 + \frac{\theta\mu s}{\nu}\right)^{-\nu-1}\right]\right\}.$$

Duchateau (2008) mostrou que a relação entre a parametrização de Aalen (1992) e a de Hougaard (1986) é a seguinte:

$$\alpha_H = 1 - \nu, \quad \delta_H = \mu^{1-\nu} \left(\frac{\mu}{\theta}\right)^\nu \quad \text{e} \quad \theta_H = \frac{\nu}{\mu\theta}.$$

Ao trabalhar com modelos de fragilidade, Hougaard (2000) propôs, ao discutir a família PVF, uma restrição nos parâmetros dessa família, em virtude do problema de identicabilidade (em que a média tem que ser igual a 1), e a variância interpretada como uma medida de heterogeneidade da população. Dessa forma, como a média da PVF é $\mathbb{E}[W] = \delta\theta^{\alpha-1}$, tem-se a restrição: $\delta = \theta^{1-\alpha}$. A fim de facilitar a compreensão de quando a restrição está sendo aplicada, é estabelecido um novo parâmetro, representado por η . Consequentemente, $\delta = \eta^{1-\alpha}$ e $\theta = \eta$, com $\alpha \in (0, 1)$ e $\eta > 0$. (Romeo, 2017)

Com essa parametrização, o valor da média será:

$$\mathbb{E}[W] = \delta\theta^{\alpha-1} = \eta^{1-\alpha}\eta^{\alpha-1} = 1$$

e a expressão da variância será:

$$\text{Var}[W] = \delta(1-\alpha)\theta^{\alpha-2} = \eta^{1-\alpha}(1-\alpha)\eta^{\alpha-2} = (1-\alpha)\eta^{-1}.$$

3.2 Cópula PVF

Em dados de sobrevivência multivariados, um modelo de fragilidade é usado para incorporar a heterogeneidade não observada entre os indivíduos, a qual pode influenciar

os tempos de sobrevivência. Essa heterogeneidade é frequentemente modelada por meio de uma variável latente chamada de fragilidade, que captura as diferenças individuais na suscetibilidade aos eventos de interesse.

De acordo com Oakes (1989), qualquer modelo de fragilidade para dados de sobrevivência multivariados possui uma representação via cópula Arquimediana.

A representação via cópula Arquimediana ocorre ao considerar a relação entre a fragilidade latente e os tempos de sobrevivência observados. Tal cópula é uma escolha comum para esse tipo de modelagem devido às suas propriedades matemáticas e flexibilidade na representação de diferentes tipos de dependência.

Sejam T_1, \dots, T_d variáveis aleatórias contínuas não negativas com funções de sobrevivência S_1, \dots, S_d . Se T_1, \dots, T_d são condicionalmente independentes dado uma variável aleatória W , não negativa, em um modelo de fragilidade, tem-se (Romeo, 2017 e Biondo, 2020):

$$S(t_1, \dots, t_d | w) = \prod_{j=1}^d S_j(t_j | w) = \prod_{j=1}^d [\bar{G}_j(t_j)]^w,$$

em que \bar{G}_j são as funções de sobrevivência basais. Dessa forma, a função de sobrevivência conjunta tem a seguinte representação de cópula Arquimediana:

$$S(t_1, \dots, t_d) = C_\phi(S_1(t_1), \dots, S_d(t_d)) = \phi^{-1} \left[\sum_{j=1}^d \phi(S_j(t_j)) \right],$$

com o gerador da cópula C_ϕ dado por $\phi = \mathcal{L}_W^{-1}$, isto é, a inversa da transformada de Laplace de W . Essa expressão é observada a partir da representação da cópula de sobrevivência (Romeo, 2017):

$$C(u_1, \dots, u_d) = S(S_1^{-1}(u_1), \dots, S_d^{-1}(u_d)),$$

com:

$$\begin{aligned} S(t_1, \dots, t_d) &= \prod_{j=1}^d \int [\bar{G}_j(t_j)]^w f_W(w) dw \\ &= \int \exp \left\{ w \sum_{j=1}^d \log(\bar{G}_j(t_j)) \right\} f_W(w) dw \\ &= \mathbb{E} \left[\exp \left\{ W \sum_{j=1}^d \log(\bar{G}_j(t_j)) \right\} \right] \\ &= \mathcal{L}_W \left(- \sum_{j=1}^d \log(\bar{G}_j(t_j)) \right) \end{aligned}$$

e

$$\begin{aligned} S_j(t_j) &= \mathcal{L}_W(-\log(\bar{G}_j(t_j))), \quad \text{para } j = 1, \dots, d, \\ S_j^{-1}(u_j) &= \bar{G}_j^{-1}(\exp\{-\mathcal{L}_W^{-1}(u_j)\}), \quad \text{para } 0 < u_j < 1, j = 1, \dots, d. \end{aligned}$$

Considerando que a cópula PVF é uma cópula Arquimediana, tem-se que a transformada de Laplace de $W \sim \text{PVF}(\alpha, \eta)$, utilizando as restrições dos parâmetros apresentadas por Hougaard (2000), é dada por:

$$\phi^{-1}(s) = \exp \left\{ \frac{1}{\alpha} \left[\eta^{1-\alpha} (\eta + s)^\alpha - \eta \right] \right\}.$$

A média e a variância de W são, respectivamente:

$$\mathbb{E}[W] = 1 \quad \text{e} \quad \text{Var}[W] = (1 - \alpha)\eta^{-1}.$$

Considerando os casos especiais e o caso limite da distribuição PVF, segue que:

- **Distribuição Estável Positiva:**

Se $\eta = 0$ e $\alpha \in (0, 1)$, então $\mathcal{L}_W(s) = \exp\{-s^\alpha\}$. Nesse caso,

$$\mathbb{E}[W] \neq 1 \text{ e } \text{Var}[W] = 0;$$

- **Distribuição Gaussiana Inversa:**

Se $\alpha = 1/2$, então $\mathcal{L}_W(s) = \exp\{2\eta[1 - (1 + s\eta^{-1})^{1/2}]\}$. Nesse caso, $\mathbb{E}[W] \neq 1$ e $\text{Var}[W] > 0$.

- **Distribuição Gama:**

Se $\alpha \rightarrow 0$, então $\mathcal{L}_W(s) = (1 + s\eta)^{-1/\eta}$. Nesse caso, $\mathbb{E}[W] = 1$ e $\text{Var}[W] = \eta > 0$.

Romeo (2017) apresentou a inversa da transformada de Laplace de W , indicada por $g(s) = \eta^\alpha - \alpha\eta^{\alpha-1} \log(s)$, e o gerador da cópula PVF, expresso como $\phi(s) = [g(s)]^{1/\alpha} - \eta$. Por conseguinte, a função cópula multivariada PVF é expressa por:

$$C_{\alpha, \eta}(u_1, \dots, u_d) = \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left(\sum_{j=1}^d [g(u_j)]^{\frac{1}{\alpha}} - (d-1)\eta \right)^\alpha - \eta \right] \right\},$$

com $u_j = S_j(t_j)$. No caso bivariado, tal cópula é escrita como:

$$C_{\alpha, \eta}(u_1, u_2) = \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left([g(u_1)]^{\frac{1}{\alpha}} + [g(u_2)]^{\frac{1}{\alpha}} - \eta \right)^\alpha - \eta \right] \right\}. \quad (3.3)$$

Na Figura 3.3 pode-se observar o comportamento da cópula PVF por meio de gráficos e curvas de nível, considerando três diferentes combinações de valores dos parâmetros α e η . Nos painéis (a) e (b) foram utilizados $\alpha = 0,9$ e $\eta = 10$; visualmente, o gráfico e a curva de nível para essa combinação de parâmetros refletem uma dependência forte, com curvas mais acentuadas, indicando uma variação rápida ou uma maior sensibilidade nas mudanças de valores da variável representada. Nos painéis (c) e (d) foram considerados $\alpha = 0,5$ e $\eta = 0,001$. Por fim, nos painéis (e) e (f) foram escolhidos $\alpha = 0,001$ e

$\eta = 0,01$; o gráfico e a curva de nível para essa combinação de parâmetros exibem uma dispersão uniforme dos pontos ou curvas quase paralelas aos eixos, não havendo uma inclinação acentuada ou uma concentração notável de pontos em uma área específica, o que representa uma dependência muito fraca ou quase inexistente entre as variáveis. Ainda pode-se verificar na Figura 3.3 que as curvas de nível da cópula PVF são convexas, com uma forma de “U” voltada para cima. Isso significa que a dependência entre as variáveis aumenta à medida que se aproximam dos extremos.

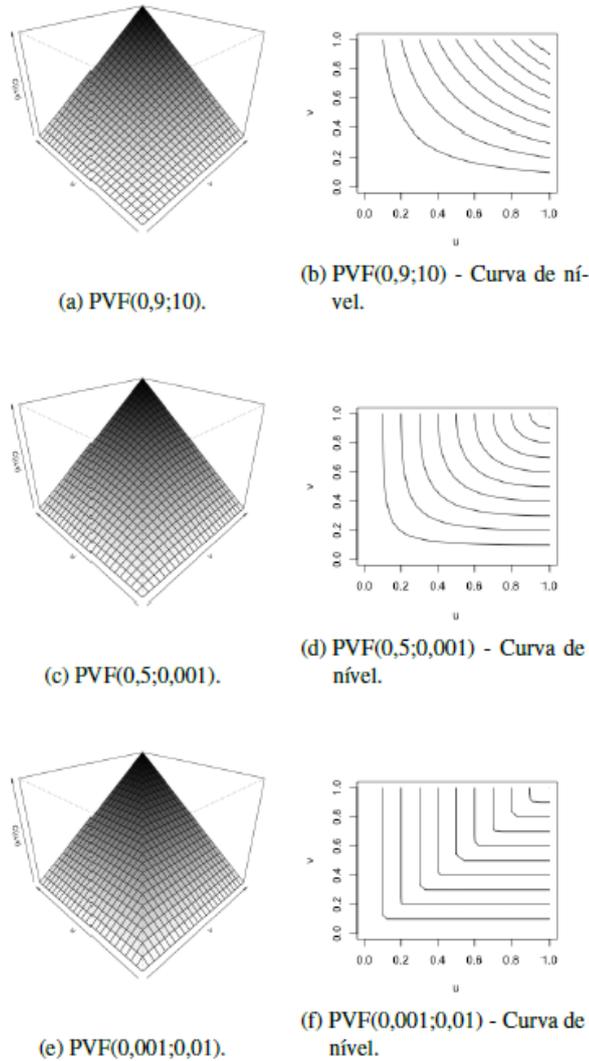


Figura 3.3: Gráficos e curvas de nível da cópula PVF, considerando diferentes combinações de valores dos parâmetros α e η .

Fonte: Biondo (2020).

Casos Especiais da Cópula PVF

A cópula PVF possui alguns casos especiais que podem ser úteis em diferentes contextos de modelagem. Esses casos especiais referem-se a valores específicos dos parâmetros

da cópula PVF, que levam a formas particulares da função de dependência.

A família de cópulas PVF possui como casos especiais as cópulas de Gumbel e Gaussiana Inversa, e como caso limitante a cópula de Clayton. Isso se torna uma vantagem para o uso da cópula PVF, já que essa flexibilidade permite ajustar a função de dependência de acordo com as necessidades e particularidades de cada análise.

Um ponto a ser destacado é o fato das cópulas de Gumbel ($\eta \rightarrow 0$) e Clayton ($\alpha \rightarrow 0$) serem obtidas quando se está na borda do espaço paramétrico. De acordo com Hougaard (1986, 2000) e Romeo (2017), seguem as expressões dos casos especiais e limitante da cópula PVF:

- **Cópula (bivariada) de Gumbel:**

$$C_\alpha(u_1, u_2) = \exp \left\{ - \left([-\log(u_1)]^{\frac{1}{\alpha}} + [-\log(u_2)]^{\frac{1}{\alpha}} \right)^\alpha \right\};$$

- **Cópula (bivariada) Gaussiana Inversa:**

$$C_\eta(u_1, u_2) = \exp \left\{ -2 \left[\eta^{\frac{1}{2}} \left([b(u_1)]^2 + [b(u_2)]^2 - \eta \right)^{\frac{1}{2}} - \eta \right] \right\},$$

com $b(s) = \eta^{1/2} - (1/2)\eta^{-1/2} \log(s)$;

- **Cópula (bivariada) de Clayton:**

$$C_\eta(u_1, u_2) = (u_1^{-\eta} + u_2^{-\eta} - 1)^{-\frac{1}{\eta}}.$$

Conforme a definição expressa por Oakes (1989), é a partir da função geradora e sua inversa que é possível escrever a função cópula de uma distribuição. A Tabela 3.1 apresenta, de forma resumida, a função geradora e sua respectiva inversa, para a cópula PVF e seus casos especiais e limitante.

Tabela 3.1: Função geradora e inversa da geradora da cópula PVF e de seus casos especiais e limitante.

Distribuição de W	$\phi^{-1}(s)$	$\phi(s)$
PVF(α, η)	$\exp \left\{ -\frac{1}{\alpha} [\eta^{1-\alpha}(\eta + s)^\alpha - \eta] \right\}$	$[g(s)]^{1/\alpha} - \eta$
Estável positiva(α)	$\exp\{-s^\alpha\}$	$[\log(-s)]^{1/\alpha}$
Gaussiana Inversa($\eta^{1/2}, \eta$)	$\exp \left\{ 2\eta \left[1 - (1 + s\eta^{-1})^{1/2} \right] \right\}$	$\frac{1}{4}\eta^{-1}[\log(s)]^2 - \log(s)$
Gama(η^{-1}, η^{-1})	$(1 + s\eta)^{-1/\eta}$	$\eta^{-1}(s^{-\eta} - 1)$

Fonte: Romeo (2017).

3.2.1 Cópula PVF e Cópulas Arquimedianas

As cópulas Arquimedianas são um tipo de cópulas largamente usado na modelagem de dependência entre variáveis aleatórias. Elas são chamadas de “Arquimedianas” em função da sua relação com funções Arquimedianas, que são funções decrescentes, não negativas e convexas.

Tais cópulas são caracterizadas por uma função geradora Arquimediana, também conhecida como função de cópula, que descreve a relação entre as distribuições marginais e a cópula conjunta das variáveis. Essa função de cópula é utilizada para modelar a dependência entre as variáveis, independentemente de suas distribuições marginais.

As cópulas Arquimedianas são de grande relevância, visto possuírem uma ampla variedade de aplicações (por exemplo, em Econometria, Finanças e Estatística), facilidade de construção, expressões analíticas extremamente simples para a maioria de seus momentos e parâmetros de dependência, além de várias famílias de cópulas pertencerem a essa classe (Nelsen, 2006).

Alguns exemplos de cópulas Arquimedianas são: Clayton, Ali-Mikhail-Haq (AMH), Gumbel, Frank e Nelsen. Na aplicação a dados financeiros, destacam-se as cópulas de Clayton e Gumbel, pois permitem a modelagem de diversas formas de dependência, incluindo assimetria e dependência nas caudas. A cópula de Clayton apresenta apenas dependência de cauda inferior e a cópula de Gumbel modela dependência de cauda superior.

A família de cópulas PVF foi considerada como uma subclasse das cópulas Arquimedianas por Romeo (2018). Nesse contexto, será apresentada uma demonstração de que a cópula PVF é, de fato, Arquimediana, usando o teorema de Ling (Ling, 1965). Nessa prova, a singularidade reside no fato de que não se encontrou qualquer indício semelhante nos materiais pesquisados. Portanto, seguem o enunciado do teorema de Ling e a demonstração para a cópula PVF.

Teorema de Ling (Ling, 1965): Seja C uma cópula associativa, tal que $C(u, u) < u$ para todo $u \in (0, 1)$. Então, C é Arquimediana.

Prova para a cópula PVF. Seja C uma cópula PVF. Então, segue que:

$$C(u, C(v, w)) = \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left([g(u)]^{1/\alpha} + [g(C(v, w))]^{1/\alpha} - \eta \right)^\alpha - \eta \right] \right\}.$$

Como $g(u) = \eta^\alpha - \alpha\eta^{\alpha-1} \log(u)$ e $g(C(v, w)) = ([g(v)]^{1/\alpha} + [g(w)]^{1/\alpha} - \eta)^\alpha$, então:

$$\begin{aligned} C(u, C(v, w)) &= \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left([g(u)]^{1/\alpha} + [g(v)]^{1/\alpha} + [g(w)]^{1/\alpha} - \eta - \eta \right)^\alpha - \eta \right] \right\} \\ &= \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left(([g(u)]^{1/\alpha} + [g(v)]^{1/\alpha} - \eta) + [g(w)]^{1/\alpha} - \eta \right)^\alpha - \eta \right] \right\} \\ &= C(C(u, v), w). \end{aligned}$$

Portanto, a cópula PVF é associativa.

Além disso, considerando a função geradora da cópula PVF, $\phi = [g(s)]^{1/\alpha} - \eta$, e sua pseudo-inversa, $\phi^{-1}(s) = \exp \{-\alpha^{-1} [\eta^{1-\alpha}(\eta + s)^\alpha - \eta]\}$, tem-se:

$$C(u, u) = \exp \left\{ \frac{1}{\alpha} \left[\eta^{1-\alpha} \left(2[g(u)]^{1/\alpha} \right)^\alpha - \eta \right] \right\} = \phi^{-1} (2(\phi(u))).$$

Para $0 < u < 1$, seguem 2 casos:

Caso 1: Para $2\phi(u) < \phi(0)$, tem-se que: $\phi^{-1} (2\phi(u)) < \phi^{-1}(\phi(0)) = u$;

Caso 2: Para $2\phi(u) \geq \phi(0)$, tem-se que: $\phi^{-1} (2\phi(u)) = \phi^{-1}(\phi(0)) = 0 < u$.

Logo, $C(u, u) < u$.

Assim, como C é associativa e $C(u, u) < u$, para todo $u \in (0, 1)$, então a cópula PVF é Arquimediana. \square

A família de cópulas PVF está em uma classe conhecida como *LT-Archimedean copulas* (Joe, 2015). A classe *LT-Archimedean copulas* refere-se a uma classe de cópulas Arquimedianas que são definidas em termos da transformada de Laplace. Por isso, como a família de cópulas PVF possui uma representação definida a partir de transformada de Laplace, justifica-se o fato dela pertencer a essa classe.

3.2.2 Cópula PVF Bivariada *versus* Cópula BB9 Bivariada

Durante o período de 2000 a 2022, autores como Hougaard (2000), Romeo (2017), Biondo (2020) e Cantoni (2021) produziram materiais com menção à cópula PVF. No ano de 2019, Jones e coautores, em seu artigo intitulado “*A bivariate power generalized Weibull distribution: A flexible parametric model for survival analysis*”, afirmaram que a cópula BB9 de Joe (1997, 2015) é a própria cópula PVF. Com exceção de Jones et al. (2019), nenhum outro artigo desse período (dentro os que foram encontrados no presente estudo) se refere à cópula BB9 como sendo a própria cópula PVF.

Ao desenvolver a expressão da cópula PVF (Equação (3.3)), obtém-se:

$$C_{\alpha, \eta}(u_1, u_2) = \exp \left\{ - \left[\left(\frac{\eta}{\alpha} - \log(u_1) \right)^{\frac{1}{\alpha}} + \left(\frac{\eta}{\alpha} - \log(u_2) \right)^{\frac{1}{\alpha}} - \left(\frac{\eta}{\alpha} \right)^{\frac{1}{\alpha}} \right]^\alpha + \frac{\eta}{\alpha} \right\}. \quad (3.4)$$

A partir da Equação (3.4), é possível estabelecer a relação entre os parâmetros empregados para a cópula BB9 com os apresentados na expressão da cópula PVF retratada por Romeo (2017). Além das duas nomenclaturas citadas anteriormente, PVF e BB9, Carole (2014) propôs uma terceira nomenclatura que também modifica os parâmetros. Com base nessa constatação, o presente estudo propõe, pela primeira vez na literatura sobre o assunto, uma comparação entre os três autores, concluindo que, de fato, trata-se da mesma cópula, embora com diferenças nos parâmetros. Vale ressaltar que Carole (2014), além de utilizar a nomenclatura BB9, também apresentou a nomenclatura Crowder. A Tabela 3.2 apresenta a relação entre os parâmetros das cópulas BB9 e PVF, não somente para a expressão da cópula de Jones et al. (2019), mas também para aquelas de Carole (2014) e Joe (1997, 2015), que exibem a BB9 com parâmetros distintos.

Tabela 3.2: Relação entre os parâmetros das cópulas BB9 e PVF.

Autores	Parâmetros da BB9	Relação com parâmetros da PVF
Joe (1997, 2015)	ν, δ	$\nu = 1/\alpha, \delta = \alpha/\eta$
Carole (2014)	δ, θ	$\delta = \eta/\alpha, \theta = 1/\alpha$
Jones et al. (2019, 2020)	λ, w	$\lambda = \eta/\alpha, w = \alpha$

Fonte: Autoria própria.

3.3 Densidade da Cópula PVF

Em uma cópula bivariada, a densidade é uma função de duas variáveis e está relacionada às probabilidades de ocorrência conjunta dos eventos. A densidade de uma cópula fornece informações sobre a relação de dependência entre as variáveis aleatórias em questão.

Se as funções de distribuição marginais, $F_X(x)$ e $F_Y(y)$, são contínuas com respeito às funções densidade marginais, $f_X(x)$ e $f_Y(y)$, então a função densidade conjunta de X e Y pode ser escrita como:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y),$$

em que $c(F_X(x), F_Y(y))$ representa a densidade da cópula, que pode ser calculada da seguinte forma:

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2), \quad \text{para } (u_1, u_2) \in [0, 1]^2.$$

Se a segunda derivada da cópula C existir, ela pode ser entendida como uma medida de dependência local (Cardoso, 2009).

Assim, a densidade da cópula PVF, definida em (3.3), pode ser expressa como:

$$\begin{aligned} c_{\alpha,\eta}(u_1, u_2) &= \frac{\partial^2}{\partial u_1 \partial u_2} C_{\alpha,\eta}(u_1, u_2) \\ &= \frac{1}{\eta u_1 u_2} q^{\frac{1}{\alpha}-1} r^{\frac{1}{\alpha}-1} \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} \right)^{\alpha-2} \left[\eta \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} - \eta \right)^\alpha - (\alpha - 1) \eta^\alpha \right] \\ &\quad \times \exp \left\{ \frac{\eta - \eta^{1-\alpha} \left(\sqrt[\alpha]{q} + \sqrt[\alpha]{r} - \eta \right)^\alpha}{\alpha} \right\}, \end{aligned}$$

com $q = \eta^{\alpha-1} [\eta - \alpha \log(u_1)]$ e $r = \eta^{\alpha-1} [\eta - \alpha \log(u_2)]$.

A Figura 3.4 ilustra o comportamento bivariado da densidade dos casos especiais e limitante da cópula PVF.

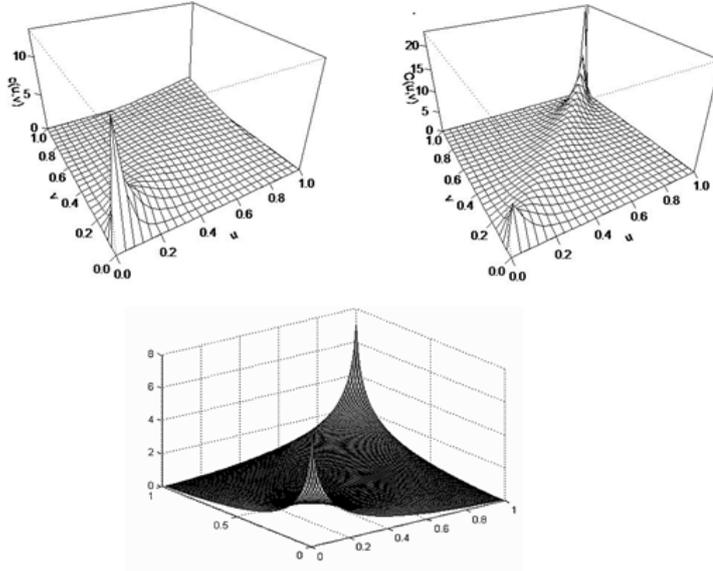


Figura 3.4: Gráficos da densidade das cópulas de Clayton (painel superior esquerdo), Gumbel (painel superior direito) e Gaussiana Inversa (painel inferior).

Fonte: Biondo (2020).

3.4 Cópula Condicional Bivariada da PVF

A cópula condicional bivariada se apresenta como uma poderosa ferramenta para a modelagem de dependência condicional entre duas variáveis aleatórias. Ela possibilita uma análise mais aprofundada e uma compreensão mais completa das relações entre as variáveis condicionadas em diferentes condições ou cenários.

Por meio da função cópula, é possível obter a distribuição condicional. A cópula condicional bivariada é definida da seguinte forma:

$$c_{u_1}(u_2) = \frac{\partial}{\partial u_1} C(u_1, u_2) \quad \text{e} \quad c_{u_2}(u_1) = \frac{\partial}{\partial u_2} C(u_1, u_2).$$

Logo, a cópula condicional bivariada da PVF é dada por:

$$c_{u_1}(u_2) = \frac{1}{u_1} q^{\frac{1}{\alpha}-1} \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} - \eta \right)^{\alpha-1} \exp \left\{ \frac{\eta}{\alpha} - \frac{\eta^{1-\alpha} \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} - \eta \right)^{\alpha}}{\alpha} \right\}$$

e

$$c_{u_2}(u_1) = \frac{1}{u_2} q^{\frac{1}{\alpha}-1} \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} - \eta \right)^{\alpha-1} \exp \left\{ \frac{\eta}{\alpha} - \frac{\eta^{1-\alpha} \left(q^{\frac{1}{\alpha}} + r^{\frac{1}{\alpha}} - \eta \right)^{\alpha}}{\alpha} \right\},$$

com $q = \eta^{\alpha-1} [\eta - \alpha \log(u_1)]$ e $r = \eta^{\alpha-1} [\eta - \alpha \log(u_2)]$.

Uma aplicação interessante da cópula condicional bivariada da PVF será apresentada na Seção 5.1, no que diz respeito ao desenvolvimento de algoritmos para a geração de dados pela distribuição condicional.

3.5 Classificação da Cópula PVF

A cópula pode ser classificada de acordo com a expressão matemática que a descreve e a estrutura de dependência por ela captada. Cardoso (2009) apresenta essa classificação, e a partir dela, pode-se classificar a cópula PVF como será visto a seguir.

Segundo a expressão matemática que descreve a cópula, a família PVF é classificada como **cópula explícita**, uma vez que possui forma fechada de apresentação. Além disso, como a sua expressão matemática é indexada por parâmetros, pode-se afirmar que a cópula PVF é uma **cópula paramétrica**. A vantagem da cópula PVF ser do tipo explícita reside no fato dela permitir uma representação clara e precisa da dependência entre as variáveis aleatórias. Através da especificação dos parâmetros, é possível controlar e modelar diferentes tipos de dependência, como dependência positiva, negativa ou nenhuma dependência.

A classe de cópulas PVF engloba três famílias de cópulas com diferentes e variadas características. Assim, cada família apresenta uma dependência de cauda específica. Conforme a estrutura de dependência por elas captadas, a cópula PVF e a cópula Gaussiana Inversa não possuem classificação, uma vez que não apresentam dependência de cauda. Em contrapartida, a cópula de Gumbel, com coeficiente de dependência de cauda superior positivo, é dita cópula do tipo “J” e a cópula de Clayton, com coeficiente de dependência de cauda inferior positivo, é dita cópula do tipo “L”.

Além disso, as cópulas também podem ser classificadas como côncavas ou convexas com base em sua forma. Como a cópula PVF é uma cópula Arquimediana, então ela é uma cópula convexa (conforme já observado na Figura 3.3). A convexidade das curvas de nível da cópula PVF reflete sua capacidade de modelar dependência positiva e não linear entre as variáveis aleatórias. Essa característica é particularmente útil quando se deseja capturar relações não lineares ou assimetrias nos dados.

3.6 Medidas de Concordância

Nesta seção será calculada e apresentada a expressão matemática e feitas algumas considerações sobre as medidas de concordância para a cópula PVF e para seus casos especiais e limitante.

τ de Kendall

Para cópulas Arquimedianas, o coeficiente τ de Kendall pode ser obtido a partir da função geradora ϕ . No caso da cópula PVF, esse coeficiente não possui uma expressão de forma fechada, mas pode ser calculado numericamente e é expresso como (Romeo, 2017):

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt = 1 - \frac{4}{\eta^{\alpha-1}} \int_0^1 \frac{[g(t)]^{1/\alpha} - \eta}{[g(t)]^{1/\alpha-1}} t dt,$$

sendo $\phi'(s) = -\eta^{\alpha-1}[g(s)]^{1/\alpha-1}s^{-1}$.

O modelo de cópula PVF somente admite modelagem da dependência positiva. As variáveis aleatórias bivariadas (e.g., tempos de sobrevivência bivariados) (T_1, T_2) são independentes se $\alpha \rightarrow 1$ para todo $\eta \in \mathbb{R}^+$, ou quando $\eta \rightarrow \infty$ para todo $\alpha \in (0, 1)$. A cópula de comonotonicidade (cópula especial que caracteriza a dependência positiva perfeita, isto é, $\tau = 1$) é obtida quando ambos os parâmetros (α, η) tendem a zero (Romeo, 2017).

A Figura 3.5 mostra quatro gráficos de dispersão de 1.000 simulações da cópula PVF para os seguintes valores do coeficiente τ : 0,9; 0,7; 0,5 e 0,2 (sendo os dois primeiros valores correspondentes a uma associação forte e os seguintes, a associação moderada e fraca, respectivamente).

Para os casos especiais e limitante da cópula PVF, seguem os respectivos coeficientes:

- **Cópula bivariada de Gumbel:**

$$\tau = 1 - \alpha.$$

Observa-se que, para $\alpha \rightarrow 1$, são obtidas cópulas independentes. Já para $\alpha \rightarrow 0$, é obtido $\tau = 1$, ou seja, é gerada a cópula de comonotonicidade;

- **Cópula bivariada de Clayton:**

$$\tau = \frac{\eta}{\eta + 2}.$$

Nota-se que, para $\eta \rightarrow 0$, são obtidas cópulas independentes. Já para $\eta \rightarrow \infty$, o coeficiente τ tende a assumir valor 1, ou seja, é gerada a cópula de comonotonicidade;

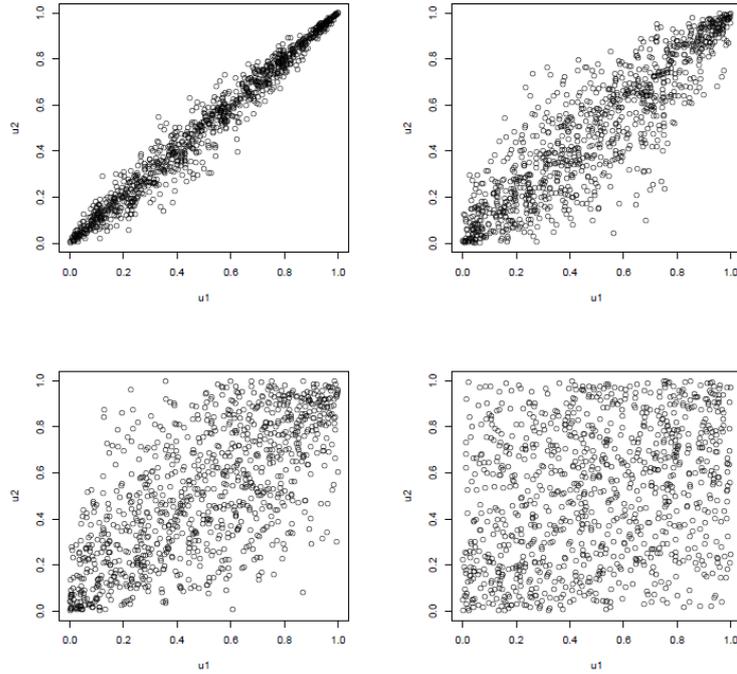


Figura 3.5: Gráficos de dispersão da cópula PVF com: $\tau = 0,9$ (painel superior esquerdo), $\tau = 0,7$ (painel superior direito), $\tau = 0,5$ (painel inferior esquerdo) e $\tau = 0,2$ (painel inferior direito).

Fonte: Biondo (2020).

- **Cópula bivariada Gaussiana Inversa:**

Diferente dos dois casos anteriores, o coeficiente τ de Kendall para a cópula bivariada Gaussiana Inversa pode ser expresso numericamente considerando $\alpha \rightarrow 1/2$. Dessa maneira, o valor de τ está no intervalo $(0, 1/2)$.

Observa-se que, para $\eta \rightarrow \infty$, são obtidas cópulas independentes. Já para $\eta \rightarrow 0$, tem-se que: $\tau \rightarrow 1/2$.

ρ de Spearman

Conforme a Equação (2.1) para o cálculo do ρ de Spearman e a Equação (3.3) da cópula bivariada PVF, o valor para o cálculo do coeficiente ρ de Spearman foi obtida utilizando o *software* R. Ao fazer uma combinação dos parâmetros α e η , obteve-se valores para o coeficiente ρ . Desse modo, seguem algumas considerações sobre o comportamento do ρ , de acordo com as combinações feitas:

- À medida que aumenta-se, de forma simultânea, os parâmetros α e η , foi possível observar que o coeficiente ρ tende a assumir o valor 0;
- Ao assumir para o parâmetro α valores tendendo a zero pela esquerda e mantendo o

valor do parâmetro η fixo (igual a 1), foi possível observar que o coeficiente ρ tende a assumir o valor 0,5;

- Ao assumir para o parâmetro α valores tendendo a 1 pela direita e mantendo o valor do parâmetro η fixo (igual a 1), foi possível observar que o coeficiente ρ tende a assumir o valor 0.

Para os casos especiais e limitante da cópula PVF, seguem as seguintes observações sobre os seus coeficientes:

- **Cópula bivariada de Gumbel:** Quando $\alpha \rightarrow 0$, o coeficiente ρ tende a assumir o valor 1. Quando α assume valores próximos a 1 pela esquerda, o ρ se aproxima do valor zero;
- **Cópula bivariada Gaussiana Inversa:** Ao considerar valores de $\eta \rightarrow 0$, observa-se que o coeficiente ρ tende a assumir o valor 0,68 e estabilizar nesse valor. Ao fazer $\eta \rightarrow \infty$, então o valor do $\rho \rightarrow 0$;
- **Cópula bivariada de Clayton:** Quando o parâmetro $\eta \rightarrow 0$, então o coeficiente ρ tende a 0. Mas um ponto interessante a se notar é que, quando η se aproxima de 130, o valor de ρ tende a 1; e depois do valor 130, o coeficiente ρ começa a diminuir.

β de Blomqvist

Considerando o caso bivariado da cópula PVF, a expressão para o cálculo da medida β de Blomqvist será:

$$\begin{aligned}\beta_C &= 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \\ &= 4 \exp \left\{ -\frac{1}{\alpha} \left[\eta^{1-\alpha} \left([g(1/2)]^{\frac{1}{\alpha}} + [g(1/2)]^{\frac{1}{\alpha}} - \eta \right)^\alpha - \eta \right] \right\} - 1 \\ &= 4 \exp \left\{ -\frac{1}{\alpha} \left(\eta^{1-\alpha} \left[2 \left(\eta^\alpha + 0,693147\alpha\eta^{\alpha-1} \right)^{1/\alpha} - \eta \right]^\alpha - \eta \right) \right\} - 1,\end{aligned}$$

utilizando $\log(1/2) = -0,693147$.

Ao fazer a combinação dos parâmetros α e η da cópula PVF, foi possível encontrar os seguintes valores para o coeficiente β de Blomqvist, dispostos na Tabela 3.3.

Para os casos especiais e limitante da cópula PVF, seguem os respectivos coeficientes:

- **Cópula bivariada de Gumbel:**

$$\beta_C = 4 \exp \{-2^\alpha \times 0,693147\} - 1.$$

Tabela 3.3: Valores de β_C , considerando diferentes combinações dos parâmetros α e η da cópula PVF.

α	η	β_C
$\alpha \rightarrow 0$	$\eta \rightarrow 0$	$\beta_C \rightarrow 1$
$\alpha \rightarrow 0$	$\eta \rightarrow \infty$	$\beta_C \rightarrow 0$
$\alpha \rightarrow 1$	$\eta \rightarrow 0$	$\beta_C \rightarrow 0$
$\alpha \rightarrow 1$	$\eta \rightarrow \infty$	$\beta_C \rightarrow 0$

Fonte: Autoria própria.

Verifica-se que, quando $\alpha \rightarrow 0$, tem-se: $\beta_C \rightarrow 1$. E, quando $\alpha \rightarrow 1$, tem-se: $\beta_C \rightarrow 0$;

- **Cópula bivariada Gaussiana Inversa:**

$$\beta_C = 4 \exp \left\{ -2 \left(\eta^{1/2} \left[2 \left(\eta^{1/2} + 0,346574 \right)^2 - \eta \right]^{1/2} - \eta \right) \right\} - 1;$$

- **Cópula bivariada de Clayton:**

$$\beta_C = 4 (2^{\eta+1} - 1)^{-1/\eta} - 1.$$

Verifica-se que, quando $\eta \rightarrow \infty$, tem-se: $\beta_C \rightarrow 1$. E, quando $\eta \rightarrow 0$, tem-se: $\beta_C \rightarrow 0$.

A principal vantagem de conhecer a medida β_C de Blomqvist para a cópula PVF, é por essa medida ser robusta a *outliers* e não faz suposições específicas sobre a forma da distribuição.

γ de Gini

A partir da Equação (2.2) do γ de Gini e da Equação (3.3) da cópula bivariada PVF, foram realizadas várias combinações utilizando os parâmetros α e η , registrando-se os valores resultantes do coeficiente γ , de modo que foram feitas algumas observações ilustradas na Tabela 3.4.

Tabela 3.4: Valores de γ_C , considerando diferentes combinações dos parâmetros α e η da cópula PVF.

α	η	γ_C
$\alpha \rightarrow 1^-$	$\eta = 1$	$\gamma_C \rightarrow 0,9$
$\alpha \rightarrow 0$	$\eta = 1$	$\gamma_C \rightarrow 0,9$

Fonte: Autoria própria.

Para os casos especiais e limitante da cópula PVF, seguem os respectivos coeficientes γ de Gini e observações:

- **Cópula bivariada de Gumbel:**

O coeficiente γ será calculado pela expressão:

$$\gamma_C = 4 \int_0^1 \left[\exp \{2^\alpha \log(t)\} + \exp \left\{ - \left([-\log(t)]^{1/\alpha} + [-\log(1-t)]^{1/\alpha} \right)^\alpha \right\} \right] dt;$$

- **Cópula bivariada Gaussiana Inversa:**

Para essa cópula, o γ de Gini apresenta um comportamento peculiar: para valores de η menores ou iguais a 1,1, o valor do γ tende a ser 1; porém, caso o η passe de 1,1, então o valor de γ ultrapassa o valor de 1, não sendo assim um valor do coeficiente a ser considerado e analisado;

- **Cópula bivariada de Clayton:**

O coeficiente γ será calculado pela expressão:

$$\gamma_C = 4 \int_0^1 \left[(2t^{-\eta} - 1)^{-1/\eta} + [t^\eta + (1-t)^{-\eta} - 1]^{-1/\eta} \right] dt - 2.$$

A principal vantagem de conhecer a medida γ de Gini para a cópula PVF é por ela ser útil para avaliar a assimetria ou desigualdade na dependência das cópulas, indicando se as maiores mudanças em uma variável estão associadas a maiores mudanças na outra.

3.6.1 Dependência de Cauda

De acordo com Nelsen (2006) e Romeo (2017), a dependência de cauda, no caso bivariado, está associada à intensidade de dependência entre duas variáveis aleatórias na cauda inferior ou cauda superior conjunta de uma distribuição bivariada.

Para as cópulas Arquimedianas, a dependência caudal pode ser expressa em termos dos geradores. Uma propriedade interessante é que a função geradora controla a intensidade da dependência caudal. Dependendo do comportamento específico da função geradora, a cópula pode exibir diferentes graus de dependência de cauda. Dessa forma, para as cópulas Arquimedianas, a dependência de cauda superior (λ_U) e a dependência de cauda inferior (λ_L) podem ser expressas, respectivamente, da seguinte forma:

$$\lambda_U = 2 - 2 \lim_{s \rightarrow 0} \frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)} \quad \text{e} \quad \lambda_L = 2 \lim_{s \rightarrow \infty} \frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)},$$

sendo que, para a cópula PVF,

$$\phi^{-1}(s) = \exp \left\{ -\alpha^{-1} \left[\eta^{1-\alpha} (\eta + s)^\alpha - \eta \right] \right\},$$

donde segue que:

$$(\phi^{-1})'(s) = -\exp\left\{-\frac{1}{\alpha} [\eta^{1-\alpha}(s+\eta)^\alpha - \eta]\right\} \left(\frac{\eta}{s+\eta}\right)^{1-\alpha}$$

e

$$\frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)} = \exp\left\{-\frac{1}{\alpha} \eta^{1-\alpha} [(2s+\eta)^\alpha - (s+\eta)^\alpha]\right\} \left(\frac{1+\eta/s}{2+\eta/s}\right)^{1-\alpha}.$$

Logo, tem-se que:

$$\lambda_U = 2 - 2 \lim_{s \rightarrow 0} \frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)} = 2 - 2 \exp\left\{-\frac{1}{\alpha} \eta^{1-\alpha} [(\eta)^\alpha - (\eta)^\alpha]\right\} \left(\frac{\eta}{\eta}\right)^{1-\alpha} = 0$$

e

$$\lambda_L = 2 \lim_{s \rightarrow \infty} \frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)} = 2 \times 0 \times (1/2)^{1-\alpha} = 0.$$

Pelos cálculos acima, como ambos os valores dos coeficientes de cauda, superior e inferior (respectivamente, λ_U e λ_L), assumem valor zero, concluímos que a cópula PVF não apresenta dependência em nenhuma das caudas.

As Figuras 3.6 e 3.7 apresentam gráficos de dispersão da cópula PVF gerados com amostras de tamanho 1.000, para diferentes valores de τ e combinações dos parâmetros α e η , a fim de ilustrar o comportamento da cauda da PVF. Na Figura 3.6, parte superior, foi considerado o coeficiente τ no valor de 0,5 e, na parte inferior, $\tau = 0,33; 0,20$ e 0,05 (seguindo da esquerda para a direita, respectivamente); pode-se observar que, numa associação positiva moderada entre as variáveis e os parâmetros tendendo a zero, existe uma tendência na cauda inferior da cópula PVF. Já na Figura 3.7, os quatro gráficos de dispersão mostram a simulação de variáveis que possuem um grau de dependência positiva forte ($\tau = 0,9$, considerando $(\alpha, \eta) = (\{0,1, 0,001\}, \{0,01, 0,05\}, \{0,097, 0,0001\})$), nos painéis superiores e $\tau = 0,7$, com $(\alpha, \eta) = (\{0,01, 0,20\}, \{0,20, 0,05\}, \{0,29, 0,001\})$, nos painéis inferiores) e nessas imagens fica nítida uma tendência nas caudas inferior e superior também quando os parâmetros se aproximam do valor zero. Em ambos os casos, é exatamente o momento em que a cópula PVF passa a ser a cópula de Clayton (caso $\eta \rightarrow 0$) ou de Gumbel (caso $\alpha \rightarrow 0$).

Similarmente à cópula PVF, os coeficientes de dependência de cauda da cópula Gaussiana Inversa são iguais a zero. Em contraste, de acordo com Cherubini (2004), segue que:

- **Cópula de Clayton:** A estrutura de dependência da cópula de Clayton é definida por exibir apenas dependência de cauda inferior; isso quer dizer que a chance da ocorrência de eventos extremos na cauda inferior da distribuição é maior que na cauda superior. A expressão para o cálculo do coeficiente de dependência na cauda inferior é dada por:

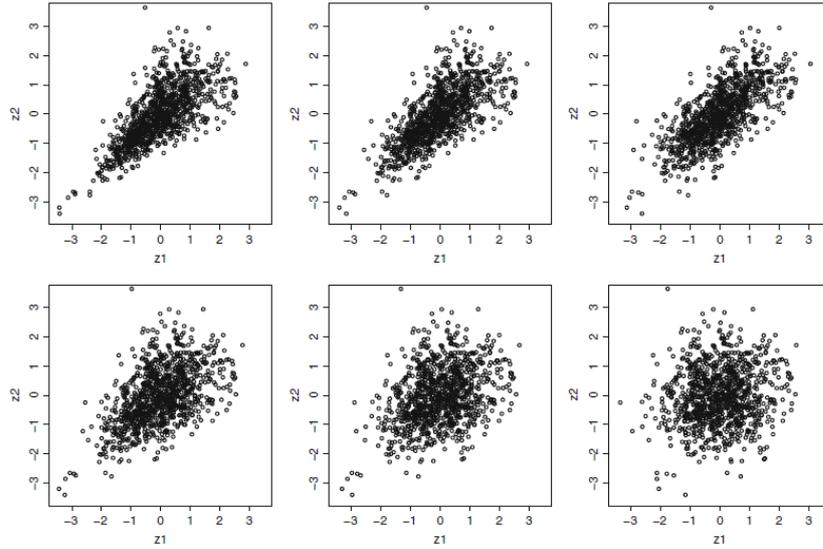


Figura 3.6: Gráficos de dispersão da cópula PVF, considerando amostras de tamanho 1.000 e $\tau = 0,5$, além de diferentes combinações de α e η (painéis superiores). Nos painéis inferiores (da esquerda para a direita), $\tau = 0,33; 0,20$ e $0,05$.

Fonte: Romeo (2017).

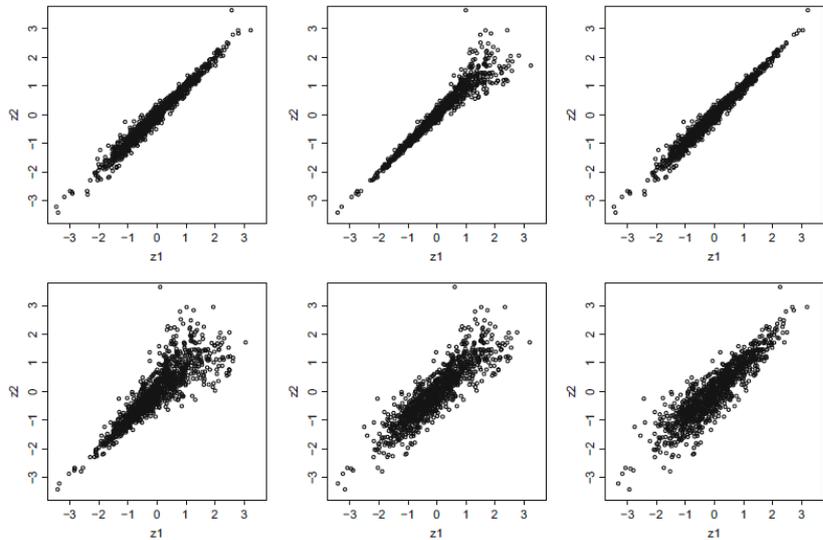


Figura 3.7: Gráficos de dispersão da cópula PVF, considerando amostras de tamanho 1.000 e $\tau = 0,9$, além de diferentes combinações de α e η (painéis superiores). Nos painéis inferiores, $\tau = 0,7$, considerando diferentes combinações de α e η .

Fonte: Romeo (2017).

$$\lambda_L = 2^{-1/\eta}.$$

Pode-se observar que, à medida que o valor de $\eta \rightarrow \infty$, λ_L tende a assumir valor igual a 1.

- **Cópula de Gumbel:** Já para a cópula de Gumbel, a estrutura de dependência é caracterizada somente pela dependência de cauda superior. A expressão para o cálculo desse coeficiente da cauda é dada por:

$$\lambda_U = 2 - 2^\alpha.$$

Analisando a expressão de λ_U , pode-se verificar que conforme $\alpha \rightarrow 0$, então $\lambda_U \rightarrow 1$. E, quando temos $\alpha \rightarrow 1^-$, ocorrerá $\lambda_U \rightarrow 0$.

A Figura 3.8 apresenta uma simulação com quatro gráficos de dispersão ilustrando a dependência de cauda das cópulas de Gumbel e Clayton. Em ambas as cópulas, foi utilizado o valor do coeficiente τ de Kendall igual a 0,4 (grau de associação positiva fraca a moderada) e 0,8 (grau de associação positiva forte; isso significa que há uma tendência de comportamento conjunto nas caudas das variáveis). Ainda é possível perceber que a cópula de Gumbel manifesta um comportamento de assimetria com uma maior dependência na cauda superior. Já a cópula de Clayton também apresenta-se de forma assimétrica e revela uma maior dependência na cauda inferior.

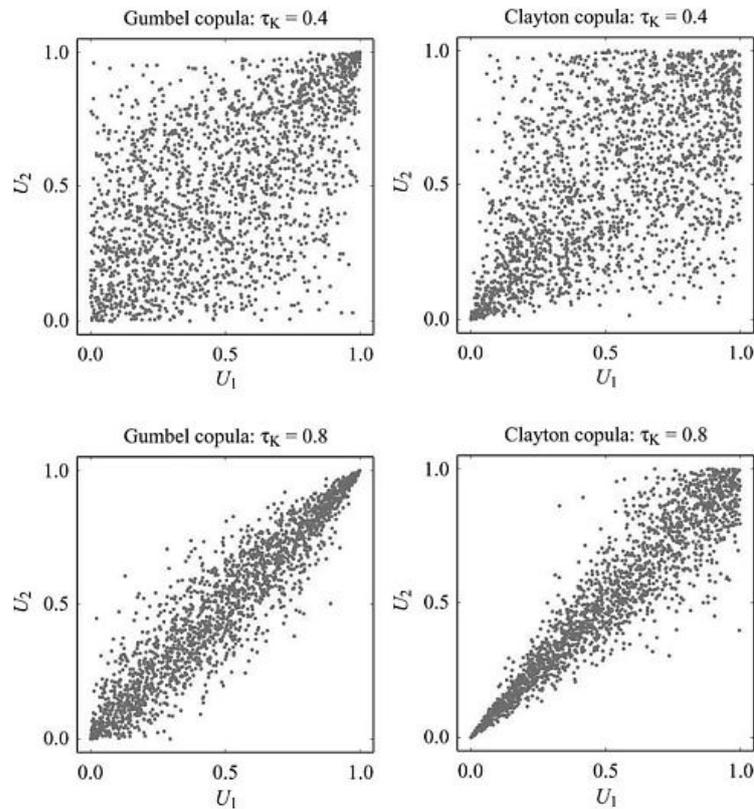


Figura 3.8: Gráficos de dispersão das cópulas de Gumbel e Clayton, para $\tau = 0,4$ e $0,8$.

Fonte: Romeo (2017).

3.6.2 Função *Cross-Ratio*

A função *cross-ratio*, também conhecida como razão cruzada, é uma medida de concordância local comum que, em Análise de Sobrevivência, fornece informações sobre a dependência entre as variáveis de sobrevivência ou os tempos de falha, permitindo uma melhor compreensão dos padrões de sobrevivência e da relação entre os eventos de falha (Romeo, 2017).

Seja (T_1, T_2) o par de tempos de falhas absolutamente contínuos. A função *cross-ratio* de T_1 e T_2 (Clayton 1978; Oakes 1989) é expressa por:

$$X^*(t_1, t_2) = \frac{h_2(t_2|T_1 = t_1)}{h_2(t_2|T_1 > t_1)} = \frac{h_1(t_1|T_2 = t_2)}{h_1(t_1|T_2 > t_2)},$$

em que h_1 e h_2 são funções de risco condicional para (T_1, T_2) , respectivamente.

A partir dessa definição, para cópulas Arquimedianas, a função *cross-ratio* pode ser escrita da seguinte forma (Romeo, 2017):

$$X(S(t_1, t_2)) = X(v) = -v \frac{\phi''(v)}{\phi'(v)}.$$

No caso da cópula PVF, considerando $\phi''(v) = \eta^{\alpha-1}[g(v)]^{\frac{1}{\alpha}-1}v^{-1}$, segue que a função $X(v)$ é decrescente em (t_1, t_2) , e a expressão da *cross-ratio* é dada por:

$$\begin{aligned} X(v) &= -v \frac{\eta^{\alpha-1}v^{-2}[g(v)]^{\frac{1}{\alpha}-1}[(1-\alpha)[g(v)]^{-1}\eta^{\alpha-1} + 1]}{\eta^{\alpha-1}[g(v)]^{\frac{1}{\alpha}-1}v^{-1}} \\ &= (1-\alpha)[g(v)]^{-1}\eta^{\alpha-1} + 1. \end{aligned}$$

Logo,

$$X(v) = 1 - \frac{1-\alpha}{\alpha \log(v) - \eta}.$$

Considerando que $X(v) = X(S(t_1, t_2))$, a Tabela 3.5 apresenta a expressão da função *cross-ratio* $X(v)$ para os casos especiais e limitante da cópula PVF, além de observações sobre o comportamento dessas funções.

Tabela 3.5: Função *cross-ratio* para os casos especiais e limite da cópula PVF.

Cópula	$X(v)$	Observação
Gumbel	$1 - (1-\alpha)[\alpha \log(v)]^{-1}$	Função decrescente em (t_1, t_2)
Gaussiana Inversa	$1 - [\log(v) - 2\eta]^{-1}$	Função decrescente em (t_1, t_2)
Clayton	$\eta^{-1} + 1$	Função constante

Fonte: Autoria própria.

Capítulo 4

Estimação de Cópulas

A função de distribuição populacional pode ser caracterizada por um número finito de termos denominados parâmetros. Neste sentido, a distribuição é dita paramétrica. Para cópulas, o processo de estimação dos parâmetros pode ser feito por meio do método de máxima verossimilhança e também pelo método de inferência para as marginais.

Este capítulo exhibe, de forma resumida, essas principais abordagens paramétricas, presentes na literatura, para se estimar os parâmetros do modelo de cópula.

4.1 Estimação por Máxima Verossimilhança

O método de máxima verossimilhança é empregado para estimar, de forma conjunta, os parâmetros das distribuições marginais e os parâmetros de uma cópula C . A estimação por máxima verossimilhança é baseada na maximização da função de verossimilhança (ou de log-verossimilhança).

Assumindo independência, a função de verossimilhança de $\boldsymbol{\theta}$, associada (ou ainda, condicional) ao vetor de observações $\boldsymbol{x} = (x_1, x_2, \dots, x_p)$, pode ser escrita como: $L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^p f(x_i|\boldsymbol{\theta})$, sendo $f(x_i|\boldsymbol{\theta})$ a FDP de X_i .

No caso das cópulas, pelo teorema de Sklar, a densidade do vetor aleatório \boldsymbol{X} pode ser escrita da seguinte forma:

$$f(x_1, x_2, \dots, x_p) = c(u_1, u_2, \dots, u_p) \prod_{i=1}^p f_i(x_i), \quad (4.1)$$

em que $c(u_1, u_2, \dots, u_p)$ é a densidade associada à cópula, $u_1 = F_1(x_1|\boldsymbol{\theta}_1), \dots, u_p = F_p(x_p|\boldsymbol{\theta}_p)$, sendo $\boldsymbol{\theta}_i$ o vetor n_i -dimensional de parâmetros da i -ésima marginal, $i = 1, 2, \dots, p$.

O logaritmo da função de verossimilhança, resultante da Equação (4.1), é dada

por:

$$\begin{aligned} \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n) &= \sum_{j=1}^n \log(c(F_1(x_{1j} | \boldsymbol{\theta}_1), \dots, F_p(x_{pj} | \boldsymbol{\theta}_p) | \boldsymbol{\theta}_c)) \\ &\quad + \sum_{i=1}^p \sum_{j=1}^n \log(f_i(x_{ij} | \boldsymbol{\theta}_i)), \end{aligned}$$

em que $\boldsymbol{\theta}_c$ é o vetor de parâmetros da cópula e $\boldsymbol{\theta}_i$ é o vetor de parâmetros da i -ésima distribuição marginal, $i = 1, \dots, p$. O estimador de máxima verossimilhança (EMV) de $\boldsymbol{\psi} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c)$, $\hat{\boldsymbol{\psi}}$, é expresso por:

$$\hat{\boldsymbol{\psi}} = \arg \max_{\boldsymbol{\psi} \in \Psi} \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n),$$

em que Ψ é o espaço paramétrico de $\boldsymbol{\psi}$ e $\arg \max_{\boldsymbol{\psi} \in \Psi} \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n)$ é definido como o valor de $\boldsymbol{\psi}$ em que o logaritmo da função de verossimilhança $\ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n)$ assume seu valor máximo.

Dada a forma analítica da função de verossimilhança, o EMV pode ser obtido tomando a primeira derivada do logaritmo da função de verossimilhança em relação a $\boldsymbol{\psi}$, e igualando os valores das derivadas parciais a zero (vetor nulo). Dessa forma, os EMVs são as soluções das equações a seguir:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n)}{\partial \boldsymbol{\theta}_1} &= \mathbf{0}, \\ &\vdots \\ \frac{\partial \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n)}{\partial \boldsymbol{\theta}_p} &= \mathbf{0}, \\ \frac{\partial \ell(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c | \mathbf{x}_1, \dots, \mathbf{x}_n)}{\partial \boldsymbol{\theta}_c} &= \mathbf{0}, \end{aligned}$$

sendo que as segundas derivadas parciais são negativas.

Assumindo que as condições de regularidade são satisfeitas, o EMV é consistente, eficiente e assintoticamente normal, isto é,

$$\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \rightarrow N_m(\mathbf{0}, I^{-1}(\boldsymbol{\psi}_0)), \quad (4.2)$$

em que m é a dimensão do vetor $\boldsymbol{\psi}$, $I(\boldsymbol{\psi}_0)$ é a matriz de informação de Fisher e $\boldsymbol{\psi}_0$ é o verdadeiro valor do vetor de parâmetros $\boldsymbol{\psi}$.

Para modelos paramétricos, o EMV dos parâmetros pode ser obtido analiticamente ou numericamente usando algoritmos iterativos, como o método de Newton-Raphson.

Pode-se também calcular intervalos de confiança (ICs) para cada um dos parâmetros. Os ICs são uma medida da incerteza em torno das estimativas dos parâmetros e fornecem uma faixa de valores em que supõe-se um determinado nível de confiança de que o verdadeiro valor do parâmetro esteja contido. No caso, a construção de ICs para os parâmetros das marginais e da cópula pode se basear na distribuição normal assintótica do EMV, dada em (4.2).

4.2 Método de Inferência para as Marginais (IFM)

Outro método adotado para estimação dos parâmetros do modelo de cópula é o *Inference Function for Margins*, conhecido como método IFM. Nele, é feita, inicialmente, a estimação dos parâmetros das marginais ($\theta_i, i = 1, \dots, p$) e, posteriormente, a estimação dos parâmetros da cópula (θ_c) (Joe e Xu, 1996; Yan, 2007). Dessa maneira, a estimação dos vetores de parâmetros pelo método IFM, nas duas etapas, é expressa por:

1. Estimar os parâmetros das marginais, fazendo a estimação das distribuições marginais univariadas:

$$\tilde{\theta}_i = \arg \max_{\theta_i} \sum_{j=1}^n \log (f_i(x_{ij}|\theta_i)),$$

para $i = 1, \dots, p$;

2. Estimar os parâmetros da função cópula, condicionando nas estimativas obtidas anteriormente para as marginais:

$$\tilde{\theta}_c = \arg \max_{\theta_c} \sum_{j=1}^n \log (c (F_1(x_{1j}|\tilde{\theta}_1), \dots, F_p(x_{pj}|\tilde{\theta}_p)|\theta_c)).$$

Portanto, o estimador obtido pelo método IFM (ou estimador IFM) é descrito como um vetor $\tilde{\psi} = (\tilde{\theta}_1, \dots, \tilde{\theta}_p, \tilde{\theta}_c)$.

Joe (1996) demonstrou que o estimador IFM, sob condições de regularidade, é consistente, eficiente e satisfaz a propriedade de normalidade assintótica.

No que tange à obtenção de ICs para os parâmetros das marginais e da cópula, Joe e Xu (1996) combinaram os métodos IFM e *jackknife* (reamostragem) para o cálculo de estimativas dos erros padrões das estimativas dos parâmetros do modelo multivariado. Com isso, as derivadas analíticas não se tornam mais necessárias para calcular a inversa da matriz de informação de Godambe (veja, por exemplo, Joe, 1997, p. 301-302, para esta forma matricial), que é a matriz de covariância assintótica associada a $\tilde{\psi}$ sob certas condições de regularidade. Por sua vez, Louzada e Ferreira (2015; 2016) propuseram o uso de métodos *bootstrap* (e.g., percentil e básico) para a construção de ICs.

É importante ressaltar que ambos os métodos (máxima verossimilhança e IFM) têm suas vantagens e limitações. Assim, a escolha entre eles depende das características dos dados, das suposições sobre as marginais e das propriedades específicas da cópula em análise.

No contexto da cópula PVF, este estudo optará por realizar a estimação dos parâmetros por meio do método de máxima verossimilhança. Essa escolha é fundamentada na capacidade deste método em proporcionar flexibilidade na modelagem da dependência

entre variáveis aleatórias. A utilização da máxima verossimilhança permitirá explorar diversas famílias de cópulas, oferecendo uma abordagem versátil para capturar e representar diferentes formas de interdependência entre as variáveis consideradas. Essa flexibilidade é essencial para abranger a variedade de estruturas de dependência que podem surgir em diferentes conjuntos de dados, contribuindo assim para uma análise mais abrangente e adaptável no contexto da cópula PVF.

Capítulo 5

Simulação e Aplicação

Neste capítulo serão apresentadas, inicialmente, três abordagens de algoritmos para simular variáveis aleatórias bivariadas a partir da cópula PVF. Tais abordagens foram baseadas em Genest et al. (1986), Embrechts et al. (2003), Nelsen (2006) e Mai et al. (2012). Em seguida, serão exibidos e discutidos os principais resultados do estudo de simulação, realizado com o objetivo de avaliar as propriedades assintóticas do EMV dos parâmetros da cópula PVF, e a aplicação bivariada do modelo PVF (e dos seus casos especiais e limitante), a fim de ilustrar sua utilidade prática e para mensurar a qualidade e simplicidade de um modelo estatístico proposto.

5.1 Abordagem pela Distribuição Condicional

Essa abordagem refere-se à geração de pares de variáveis aleatórias uniformemente distribuídas no intervalo $[0, 1]$, em que uma cópula bivariada C será a distribuição conjunta, com parâmetro(s) conhecido(s). Tal método pode ser aplicado a cópulas cuja função inversa apresenta solução analítica.

No caso da cópula PVF, é possível a geração de dados pela distribuição condicional, de forma fechada, apenas para o seu caso limite: a cópula de Clayton.

Assim, considerando a cópula de Clayton, o algoritmo para a geração de um vetor aleatório (u_1, u_2) , a partir da amostragem condicional, será:

1. Simular um par de variáveis aleatórias independentes (v_1, v_2) de $U(0, 1)$;
2. Fazer $u_1 = v_1$;
3. Fazer $v_2 = C(u_2|v_1)$. Assim, $v_2 = \frac{(\phi^{-1})'(c_2)}{(\phi^{-1})'(c_1)}$, com $c_1 = \phi(u_1) = u^{-\eta} - 1$ e $c_2 = \phi(u_1) + \phi(u_2) = u_1^{-\eta} + u_2^{-\eta} - 2$. Logo, $v_2 = \left(\frac{u_1^{-\theta} + u_2^{-\theta} - 2}{u_1^{-\theta}} \right)^{-\frac{1}{\theta}} - 1$.

Portanto, a segunda variável do par, amostrada da cópula de Clayton, é:

$$u_2 = \left[v_1^{-\theta} \left(v_2^{-\frac{\theta}{\theta+1}} - 1 \right) + 1 \right]^{-\frac{1}{\theta}}.$$

Para a aplicação do método da distribuição condicional para a cópula PVF, é necessário uma abordagem numérica, já que a inversa da condicional não pode ser resolvida explicitamente de forma fechada. Portanto, Romeo (2018) propôs o seguinte algoritmo de geração de dados para a cópula PVF:

1. Simular $u_2 \sim U(0, 1)$ e $q \sim U(0, 1)$;
2. Calcular $F_{U_2}(u_1) := \frac{\partial}{\partial u_2} C(u_1, u_2)$ e $u_1 = F_{U_2}^{-1}(q)$.

Aqui, é necessário encontrar as raízes da equação de forma numérica, isto é, $F_{U_2}^{-1}(q) - q = 0$, uma vez que $u_1 = F_{U_2}^{-1}(q)$ não possui solução em forma fechada;

- 2.1. Definir um valor inicial $u_1^{(0)} = u_0$;
- 2.2. Fazer $u_1^{(j+1)} = u_1^{(j)} - \left(\frac{\partial}{\partial u_2} C(u_1^{(j)}, u_2) - q \right) \left[\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1^{(j)}, u_2) \right]^{-1}$;
- 2.3. Se $\left| u_1^{(j+1)} - u_1^{(j)} \right| < \epsilon$, retornar $u_1 = u_1^{(j+1)}$;
3. Retornar (u_1, u_2) ;
4. Repetir os passos anteriores n vezes para obter uma amostra de n pares da cópula PVF.

O passo 2 pode ser implementado de forma direta usando, por exemplo, a função *uniroot* no *software* estatístico R.

5.2 Algoritmo para Cópulas Arquimedianas em Geral

Como a distribuição condicional da cópula PVF não é diretamente inversível, uma alternativa seria usar um algoritmo para cópulas Arquimedianas em geral, baseado no teorema a seguir.

Teorema (Leal, 2015): Sejam U_1 e U_2 variáveis aleatórias $U(0, 1)$, com distribuição bivariada definida por uma cópula Arquimediana com função geradora ϕ . Então, a função de distribuição de $C(U_1, U_2)$ é dada por $K_c(t) = t - \phi(t)/\phi'(t)$.

O algoritmo de geração para a cópula PVF é então descrito a seguir.

1. Simular duas variáveis aleatórias $U(0, 1)$ independentes, (s, q) ;

2. Fazer $t = K_c^{-1}(q)$, em que $K_c(t) = t - \frac{\phi(t)}{\phi'(t)}$.

Sendo $\phi(s) = [g(s)]^{1/\alpha} - \eta$, então $K_c(s) = t \left(1 + \frac{[g(s)]^{1/\alpha} - \eta}{\eta^{\alpha-1} - [g(s)]^{1/\alpha-1}} \right)$;

3. Repetir os passos anteriores n vezes para obter uma amostra de n pares da cópula PVF.

Para a cópula PVF, a equação $t = K_c^{-1}(q)$ não apresenta solução em forma fechada. Desse modo, uma saída seria resolver numericamente a equação $K_c(t) - q = 0$.

Romeo (2018) aponta que a solução numérica da equação $K_c(t) - q = 0$ seguiria o algoritmo da distribuição condicional apresentado na Seção 5.1. Isto é:

2.1. Definir um valor inicial $t^{(0)} = t_0$;

2.2. $t^{(j+1)} = t^{(j)} - \left[t^{(j)} \left(1 + \frac{[g(s)]^{1/\alpha} - \eta}{\eta^{\alpha-1} - [g(s)]^{1/\alpha-1}} \right) - q \right]$;

2.3. Se $|t^{(j+1)} - t^{(j)}| < \epsilon$, retornar $t = t^{(j+1)}$;

3. Retornar o par (u_1, u_2) , em que $u_1 = \phi^{-1}(s\phi(t))$ e $u_2 = \phi^{-1}((1-s)\phi(t))$.

5.3 Abordagem de Marshall e Olkin

Conforme comentado anteriormente na Seção 3.2.1, a cópula PVF faz parte de uma classe de cópulas conhecida como *LT-Archimedean copulas*, a qual possui uma representação definida a partir da transformada de Laplace. A inversa da função geradora da cópula, ϕ , possui uma representação a partir de uma transformada de Laplace de uma distribuição G , isto é,

$$\phi = \mathcal{L}^{(-1)}(G).$$

O método apresentado por Marshall e Olkin (1988) envolve a transformada de Laplace e sua inversa. O algoritmo baseado na representação a partir de transformadas de Laplace é expresso por:

1. Definir uma variável $V \sim G = \mathcal{L}^{(-1)}(\phi)$, com a transformada de Laplace de G dada por $\psi(t) = \int_0^\infty e^{tx} dG(x)$, para $t \geq 0$;

2. Simular duas variáveis aleatórias $U(0, 1)$ independentes, (v_1, v_2) ;

3. Fazer $u_i = \psi \left(-\frac{\log(v_i)}{V} \right)$, para $i = 1, 2$;

4. O vetor (u_1, u_2) possui distribuição associada à cópula com função geradora $\phi = \psi^{-1}$.

A única desvantagem desse método refere-se à exigência de simular uma variável a mais, V , o que nem sempre é fácil de obter, como no caso da cópula PVF. Porém, para os seus casos especiais, Clayton e Gumbel, é possível aplicar esse método. Assim, seguem os respectivos algoritmos.

• **Cópula de Clayton:**

1. Definir uma variável aleatória $V \sim \text{Gama}(1, 1/\eta)$, cuja transformada de Laplace é dada por $\psi(t) = (1 + t)^{-\frac{1}{\eta}}$;
2. Simular duas variáveis aleatórias $U(0, 1)$ independentes, (v_1, v_2) ;
3. Fazer $u_i = \psi\left(-\frac{\log(v_i)}{V}\right)$, para $i = 1, 2$.

• **Cópula de Gumbel:**

1. Definir uma variável aleatória $V \sim \text{Estável Positiva}(1/\alpha)$, cuja transformada de Laplace é dada por $\psi(t) = \exp\{-t^{\frac{1}{\alpha}}\}$;
2. Simular duas variáveis aleatórias $U(0, 1)$ independentes, (v_1, v_2) ;
3. Fazer $u_i = \psi\left(-\frac{\log(v_i)}{V}\right)$, para $i = 1, 2$.

5.4 Simulação

O estudo de simulação deste trabalho envolve a geração de dados simulados com base no modelo estatístico da cópula PVF bivariada. Para tanto, o modelo da cópula PVF é expresso como:

$$C_{\alpha, \eta}(u_1, u_2) = \exp\left\{-\frac{1}{\alpha} \left[\eta^{1-\alpha} \left([g(u_1)]^{\frac{1}{\alpha}} + [g(u_2)]^{\frac{1}{\alpha}} - \eta\right)^{\alpha} - \eta\right]\right\}.$$

Esta simulação foi realizada com o auxílio do *software* R, onde foram geradas $b = 1.000$ amostras (réplicas) da cópula PVF bivariada, de tamanhos $n = 50, 100$ e 500 . Para avaliar o grau de dependência entre as distribuições marginais, foi feita uma combinação dos parâmetros α e η . Para analisar uma associação fraca ($\tau = 0,25$), os parâmetros escolhidos foram: $\alpha = 0,4$ e $\eta = 0,6$. Já para uma associação moderada ($\tau = 0,5$), a combinação foi: $\alpha = 0,25$ e $\eta = 0,2$. Por fim, para investigar uma associação forte ($\tau = 0,75$), a escolha foi: $\alpha = 0,1$ e $\eta = 0,1$.

5.4.1 Resultados e Discussão

A Tabela 5.1 exibe os resultados das simulações para as diferentes combinações de parâmetros α e η , apresentando o valor médio, o vício e o erro quadrático médio (EQM) de cada estimativa dos parâmetros do modelo PVF, obtidos através do método de máxima verosimilhança, segundo os graus de dependência fraca, moderada e forte, respectivamente.

De acordo com a Tabela 5.1, é possível fazer as seguintes observações:

- Para um mesmo valor de α e diferentes valores de n , o valor médio não apresenta diferenças consideráveis, variando em torno de 0,01;
- Para η igual a 0,2 e 0,1, o valor médio não apresenta diferenças notórias;
- Para $\eta = 0,6$, o valor médio vai crescendo à medida que o valor de n também aumenta, porém, para $n = 500$, ocorre uma mudança brusca, na qual a média cai mais de 5 unidades.

Além disso, verifica-se que o viés e o EQM do parâmetro α , em todas as suas diferentes combinações com η , tendem a se aproximar de zero à medida que n aumenta, o que é desejado/esperado, já que o viés representa a diferença entre o valor estimado e o valor verdadeiro. Assintoticamente, pode-se ver que o EMV do parâmetro α é não viesado e EQM consistente.

Nota-se também, na Tabela 5.1, que para a combinação $\alpha = 0,4$ e $\eta = 0,6$, ocorre uma instabilidade na estimação do parâmetro η quando $n < 500$. Já nos casos de dependência moderada e forte, o vício e o EQM tendem a se aproximar de zero à medida que n aumenta, indicando, assim, que o EMV do parâmetro η é não viesado e eficiente, fornecendo estimativas precisas e confiáveis à medida que o tamanho da amostra aumenta. Ademais, o EQM apresenta-se como consistente.

Tabela 5.1: Média, vício e EQM das estimativas dos parâmetros do modelo de cópula PVF (caso bivariado).

Parâmetro	$n = 50$			$n = 100$			$n = 200$			$n = 500$		
	Média	Vício	EQM	Média	Vício	EQM	Média	Vício	EQM	Média	Vício	EQM
$\alpha = 0,4$	0,3696	-0,0304	0,0790	0,3752	-0,0248	0,0594	0,3728	-0,0272	0,0384	0,3797	-0,0203	0,0191
$\eta = 0,6$	5,4719	4,8719	5,876,60	4,7105	4,1105	10,215,36	6,3606	5,7606	31,856,28	0,6714	0,0714	0,1441
$\alpha = 0,25$	0,2392	-0,0108	0,0162	0,2462	-0,0038	0,0080	0,2465	-0,0035	0,0039	0,2480	-0,0020	0,0016
$\eta = 0,2$	0,2287	0,0287	0,0296	0,2123	0,0123	0,0132	0,2069	0,0069	0,0060	0,2027	0,0027	0,0021
$\alpha = 0,1$	0,0984	-0,0016	0,0021	0,1003	0,0003	0,0010	0,0996	-0,0004	0,0005	0,0997	-0,0003	0,0002
$\eta = 0,1$	0,1026	0,0026	0,0018	0,1002	0,0002	0,0008	0,1003	0,0003	0,0004	0,1001	0,0001	0,0002

Fonte: Autoria própria.

5.5 Aplicação

Nesta seção, o modelo PVF (e seus casos especiais e limitante) é aplicado a um conjunto de dados reais sobre exploração de urânio (Cook e Johnson, 1986), disponível no pacote `copula` (Yan, 2007) do *software* R pela nomenclatura de *uranium*.

Esse conjunto de dados é formado por concentrações logarítmicas de sete elementos químicos em 655 amostras de água coletadas perto de Grand Junction (do quadrângulo de Montrose, no oeste do Colorado, Estados Unidos da América). Tais concentrações foram calculadas para os seguintes elementos químicos: urânio (U), lítio (Li), cobalto (Co), potássio (K), céσιο (Cs), escândio (Sc) e titânio (Ti).

Para selecionar duas variáveis do referido conjunto de dados, visando a aplicação bivariada do modelo PVF e de seus casos especiais e limitante, analisou-se quais variáveis estavam mais correlacionadas/associadas entre si, de acordo com o coeficiente τ de Kendall (versão amostral). Os valores de τ , para todas as combinações duas a duas entre as variáveis, estão apresentados na Tabela 5.2.

Tabela 5.2: Coeficiente τ de Kendall para todas as variáveis do conjunto de dados *uranium*, tomadas duas a duas.

	U	Li	Co	K	Cs	Sc	Ti
U	1,00000000	-	-	-	-	-	-
Li	0,13259055	1,00000000	-	-	-	-	-
Co	0,05958400	0,006097769	1,00000000	-	-	-	-
K	0,19311832	0,111016779	-0,099387712	1,00000000	-	-	-
Cs	0,47026041	0,125019408	0,166577514	0,20741073	1,00000000	-	-
Sc	0,09233127	0,102180954	0,535117949	-0,13684047	0,2328236	1,00000000	-
Ti	0,13451850	0,002773836	0,364747010	0,04060148	0,3039988	0,43551173	1,00000000

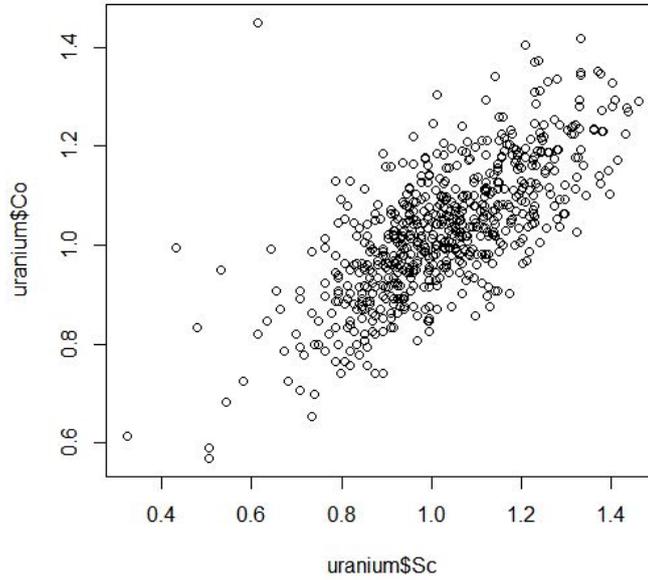
Fonte: Autoria própria.

Analisando a Tabela 5.2, observa-se que as variáveis Co (numérica - concentração logarítmica de cobalto) e Sc (numérica - concentração logarítmica de escândio) são as que apresentaram a melhor (isto é, maior) relação de associação, com coeficiente τ de Kendall igual a 0,535117949, sendo assim um grau de dependência positiva moderada.

Pode-se ainda visualizar como as variáveis Co e Sc estão correlacionadas, por meio do gráfico da Figura 5.1. Nota-se, em geral, um aumento no valor da variável Co quando a outra variável, Sc, também aumenta (e o contrário também). Ou seja, há uma relação positiva entre essas variáveis.

De posse das duas variáveis selecionadas, parte-se para a avaliação da qualidade e obtenção de uma métrica para comparação e seleção entre os modelos PVF e os seus casos especiais e limitante. Pode-se utilizar, por exemplo, o Critério de Informação de Akaike (AIC) (Akaike, 1977), pois este fornece uma medida relativa de *goodness-of-fit* (qualidade

Figura 5.1: Gráfico de dispersão das variáveis Co e Sc.



Fonte: Autoria própria.

do ajuste) do modelo estatístico. Considerando distribuições marginais não paramétricas (acumuladas empíricas) para as variáveis envolvidas, a Tabela 5.3 exibe o valor do AIC dos modelos candidatos: PVF, Clayton, Gumbel e Gaussiana Inversa.

Tabela 5.3: Valores AIC dos modelos PVF, Clayton, Gumbel e Gaussiana Inversa.

PVF	Clayton	Gumbel	Gaussiana Inversa
-501,9077	-399,3757	-455,4856	-455,3962

Fonte: Autoria própria.

Observa-se pela Tabela 5.3 que, dentre os quatro modelos candidatos, o que obteve um melhor desempenho em termos de ajuste aos dados observados, considerando o equilíbrio entre a complexidade do modelo e a qualidade do ajuste, foi o modelo da cópula PVF, visto que apresentou a menor pontuação de AIC ($-501,9077$). Dessa forma, comprova-se que, com o modelo PVF, há menos informações perdidas, portanto, uma maior qualidade do ajuste, ou seja, um bom ajuste do modelo bivariado de cópula.

Para o modelo da cópula PVF, o parâmetro α estimado foi 0,3840549, o qual influencia a forma da cauda da distribuição. Já para o parâmetro η , a estimativa foi de 0,04898758, o que impacta a intensidade da dependência.

Capítulo 6

Considerações Finais

No presente estudo obteve-se novos resultados para a cópula PVF, como a demonstração de que a cópula BB9 é a própria cópula PVF, apenas com diferenciação na sua parametrização. Dessa forma, foi exibida a relação entre os parâmetros da cópula PVF e os da BB9, na forma como eram apresentadas por diferentes autores. Além disso, utilizando o teorema de Ling, provou-se que a cópula PVF é uma cópula Arquimediana.

Ademais, foram apresentados alguns conceitos referentes às funções cópula, com destaque para a cópula PVF. Por meio de uma revisão bibliográfica, conseguiu-se reunir informações disponíveis sobre a cópula PVF e seus casos especiais e limitante, além de apresentar, de forma inédita, outras medidas de concordância para ela (ρ de Spearman, β de Blomqvist e γ de Gini).

Foi realizado um estudo de simulação com o objetivo de expor o bom comportamento das estimativas de máxima verossimilhança dos parâmetros da cópula PVF, com base no vício e EQM. Por meio dessas simulações, também foi verificado que, com diferentes tamanhos amostrais, há existência de uma instabilidade do EMV do parâmetro η (quando a correlação é moderada) no momento em que o tamanho da amostra é menor do que 500.

Por fim, o modelo de cópula PVF, que tem como casos particulares os modelos de cópula de Clayton, Gumbel e Gaussiana Inversa, foi aplicado a um conjunto de dados disponível no pacote `copula` do *software* R. Dentre os modelos candidatos, o modelo PVF foi o que apresentou o melhor ajuste. A comparação/seleção de modelos foi feita utilizando o critério AIC.

Para um trabalho futuro, pretende-se realizar pesquisas mais aprofundadas na literatura, visando confirmar o ineditismo de alguns dos resultados aqui desenvolvidos, bem como estudar/investigar mais detalhadamente o comportamento do EMV do parâmetro η , e realizar uma abordagem multivariada do modelo PVF e seus casos particulares e limite. Intenciona-se também realizar a estimação da cópula PVF utilizando o método

IFM, comparando sua efetividade em relação ao método da máxima verossimilhança. Seria interessante realizar uma abordagem Bayesiana, podendo assumir prioris hierarquicas ou dependencia condicional entre α e η . E por último, outra possibilidade para estudo futuro seria exibir uma generalização da família de cópulas PVF com base em construções quadráticas de cópulas e em algumas formas de convexidade.

Referências

- [1] Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 4:951-972.
- [2] Akaike, H. (1977). On entropy maximization principle. In: Krishnaiah, P.R., Ed., *Applications of Statistics*, North-Holland, Amsterdam.
- [3] Andersen, E.W. (2005). Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis*, 11:333–350.
- [4] Batista, A.C.S. (2019). Controle Estatístico de Processos Multivariados baseado em Funções Cópula. Dissertação de Mestrado em Matemática (área de concentração em Estatística) - Universidade Federal da Bahia.
- [5] Biondo, T.R. (2020). Modelos de Sobrevivência Bivariados baseados na Cópula PVF. Dissertação de Mestrado em Estatística - Universidade de São Paulo.
- [6] Cantoni, M. (2021). Full Bayesian Significance Test para dados de sobrevivência bivariados: seleção de modelos encaixados da cópula PVF. Tese de Doutorado em Estatística - Universidade Federal de São Carlos.
- [7] Cardoso, P.A. (2009). Uma metodologia para estimação do capital econômico: incorporação de dependência entre riscos via cópulas. Dissertação de Mestrado em Atuária - Pontifícia Universidade Católica do Rio de Janeiro.
- [8] Cherubini, U.; Luciano, E.; Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons, Ltd. ISBN: 9780470863442, 310 pages, 1ª Edição.
- [9] Cherubini, U.; Mulinacci, S.; Gobbi, F.; Romagnoli, S. (2011). *Dynamic Copula Methods in Finance*. Wiley Finance. ISBN: 9780470683071, 288 pages, 1ª Edição.
- [10] Clayton DG. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*. 1978; 65:141–151

- [11] Cook, R.D.; Johnson, M.E. (1986). Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics*, 28(2):123-131.
- [12] Dicionário Priberam da Língua Portuguesa (2008-2021). Disponível em: <https://dicionario.priberam.org/chave>. Acesso em: 8 de setembro de 2022.
- [13] Duchateau, L.; Janssen, P. (2008). *The Frailty Model*. Springer, New York.
- [14] Embrechts, P.; McNeil, A; Straumann, D. (2001). *Correlation and Dependency in Risk Management: Properties and Pitfalls*. Cambridge University Press.
- [15] Embrechts, P.; Lindskog, F.; McNeil, A.J. (2003). *Modelling dependence with copulas and applications to risk management*. In: Rachev R (ed.) Handbook of heavy tailed distributions in finance. Elsevier/North-Holland, Amsterdam, pp 329–384.
- [16] Fischer, N. I. Copulas. In: Encyclopedia of Statistical Sciences, Update v. 1, p. 159-163. John Wiley Sons, New York, 1997.
- [17] Genest, C.; MacKay, R.J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician*, 40:280–285.
- [18] Hanagal, D.D. (2009). Weibull extension of bivariate exponential regression model with different frailty distributions. *Statistical Papers*, 50:29–49.
- [19] Hua, L.; Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, 102:1454–71.
- [20] Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396.
- [21] Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer, New York.
- [22] Jaworski, P.; Durante, F.; Hardle, W.K.; Rychlik, T. (2010). *Copula Theory and Its Applications*. Springer, Berlin.
- [23] Joe, H.; Xu, J.J. (1996). The estimation method of inference functions for margins for multivariate models.
- [24] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- [25] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419.
- [26] Joe, H. (2015). *Dependence modeling with copulas*. CRC Press, Boca Raton.

- [27] Jones, M.C.; Noufaily, A.; Burke, K. (2020). A bivariate power generalized Weibull distribution: A flexible parametric model for survival analysis. *Statistical Methods in Medical Research*, 29(8):2295-2306.
- [28] Kolev, N.; Anjos, U.; Mendes, B.V.M. (2006). Copulas: a review and recent developments. *Stochastic Models*, 22(4):617-660.
- [29] Leal, R.B. (2015). Cópulas: Utilização no R e aplicações com dados de sobrevivência. Disponível em: <https://www.ufpe.br/documents/39362/4014760/Copulas+UFPE.pdf/3b30989a-af93-4f36-9c8d-251ba196d4a5>
- [30] Lloréns, L. L. (2017) *Teoría de cópulas Introducción y aplicaciones a Solvencia II*, Fundación MAPFRE, Madrid.
- [31] Louzada, F.; Ferreira, P. H. (2015). On the classical estimation of bivariate copula-based seemingly unrelated Tobit models through the proposed inference function for augmented margins method. *Journal of Data Science*, 13(4):771-794.
- [32] Louzada, F.; Ferreira, P. H. (2016). Modified inference function for margins for the bivariate Clayton copula-based SUN Tobit model. *Journal of Applied Statistics*, 43(16):2956-2976.
- [33] Mai, J.; Scherer, M. (2012). *Simulating copulas: stochastic models, sampling algorithms, and applications*. Imperial College, Boca Raton.
- [34] Mallick, M.; Ravishanker, N. (2006). PVF frailty models with a flexible baseline hazard. *International Journal of Statistics and Systems*, 1:57–80.
- [35] Marshall, A.W.; Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83:834–841.
- [36] Michaelis dicionário escolar língua portuguesa: nova ortografia conforme acordo ortográfico da língua portuguesa. São Paulo: Melhoramentos, 2008. 951 p.
- [37] Nelsen, R.B. (2006). *An introduction to copulas*, 2nd edn. Springer, New York.
- [38] Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84:487–493.
- [39] Romeo, J.S.; Tanaka, N.I.; Pedroso-de-Lima, A.C. (2006). Bivariate survival modeling: a Bayesian approach based on copulas. *Lifetime Data Analysis*, 12:205–222.
- [40] Romeo, J.S.; Meyer, R. (2015). Bayesian semi-parametric analysis of recurrent failure time data using copulas. *Biometrical Journal*, 57:982–1001.

- [41] Romeo, J.S.; Meyer, R. (2017). Bayesian approach for modelling bivariate survival data through the PVF. *Lifetime Data Analysis*, 24(2):355-383.
- [42] Saminger-Platz, S.; Kolesárová, A.; Šeliga, A.; Mesiar, R.; Klement, E.P. (2021). The impact on the properties of the EFGM copulas when extending this family. *Fuzzy Sets and Systems*, 415:1–26.
- [43] Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201-218.
- [44] Simpson, E.S.; Wadsworth, J.L.; Tawn, J.A. (2021). A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis*, 184:104736.
- [45] Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.
- [46] Spreuw, J. (2014). Archimedean copulas derived from utility functions. *Insurance Math. Insurance: Mathematics and Economics*, 59, 235–242.
- [47] Tweedie, M.C.K. (1984). *An index which distinguishes between some important exponential families*. In: Ghosh JK, Roy J (eds). *Statistics: applications and new directions*. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta, pp 579–604.
- [48] Wienke, A. (2010). *Frailty models in survival analysis*. Chapman and Hall/CRC, Boca Raton 123.
- [49] Yan, J. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1–21.