



DANIELLE GONÇALVES DE OLIVEIRA PRADO

**AJUSTE DE CÓPULAS BIVARIADAS VIA
MARGINAL NA DIAGONAL E TESTE PARA
DEPENDÊNCIA DE VALORES EXTREMOS VIA
AJUSTE DE FUNÇÃO DE PICKANDS POR
POLINÔMIOS DE BERNSTEIN**

**LAVRAS – MG
2016**

DANIELLE GONÇALVES DE OLIVEIRA PRADO

**AJUSTE DE CÓPULAS BIVARIADAS VIA MARGINAL NA
DIAGONAL E TESTE PARA DEPENDÊNCIA DE VALORES
EXTREMOS VIA AJUSTE DE FUNÇÃO DE PICKANDS POR
POLINÔMIOS DE BERNSTEIN**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutora.

Orientador
Dr. Lucas Monteiro Chaves

Coorientador
Dr. Devanil Jaques de Souza

**LAVRAS – MG
2016**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Prado, Danielle Gonçalves de Oliveira.

Ajuste de cópulas bivariadas via marginal na diagonal e teste para dependência de valores extremos via ajuste de função de Pickands por polinômios de Bernstein : / Danielle Gonçalves de Oliveira Prado. – Lavras : UFLA, 2016.

126 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2016.

Orientador(a): Lucas Monteiro Chaves.

Bibliografia.

1. qualidade de ajuste. 2. cópulas. 3. função de Pickands. 4. polinômios de Bernstein. 5. dependência de valores extremos. I. Universidade Federal de Lavras. II. Título.

DANIELLE GONÇALVES DE OLIVEIRA PRADO

**AJUSTE DE CÓPULAS BIVARIADAS VIA MARGINAL NA
DIAGONAL E TESTE PARA DEPENDÊNCIA DE VALORES
EXTREMOS VIA AJUSTE DE FUNÇÃO DE PICKANDS POR
POLINÔMIOS DE BERNSTEIN**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutora.

APROVADA em 18 de fevereiro de 2016.

Dr. Marcos de Oliveira Prates	UFMG-MG
Dr. Evelise Román Corbalán Góis Freire	UFLA
Dr. Daniel Ferreira Furtado	UFLA

Dr. Lucas Monteiro Chaves
Orientador

Dr. Devanil Jaques de Souza
Coorientador

**Lavras– MG
2016**

AGRADECIMENTOS

Mais uma etapa sendo concluída e esse é o momento de agradecer.

Em primeiro lugar, agradeço ao bom Deus que me carregou nos braços nessa longa e árdua caminhada, sempre guiando meus caminhos e iluminando meus estudos.

Aos meus pais e heróis, Júlio e Denise, à minha vizinha tão amada, Diná, e à minha super-irmã-amiga Belles, que estiveram ao meu lado o tempo todo vivenciando comigo cada batalha. Agradeço as orações, o apoio, o carinho e incentivo que sempre me deram. Sem sombra de dúvidas Papai do Céu foi bem generoso e me deu a melhor família do planeta Terra. Vocês sabem o quanto me orgulho disso! Sem palavras para descrever o amor que sinto por vocês.

Ao meu co-orientador prof. Devanil Jaques de Souza e ao meu orientador prof. Lucas Monteiro Chaves que são bem mais que educadores e se tornaram, de coração, pais emprestados. Sou grata por tanta dedicação, atenção, carinho, paciência e compreensão durante essa longa e árdua caminhada, em especial, nos momentos em que tanto precisei nos EUA. Obrigada também por não medirem esforços e sempre dividirem o “nosso” ambiente de trabalho.

Aos membros da banca, prof. Daniel, prof. Júlio, prof. Marcos e profa. Evelise. Obrigada por aceitarem o convite e por terem sido tão atenciosos nas correções.

À Universidade Federal de Lavras, pelas oportunidades concedidas.

A todos os professores do Departamento de Ciências Exatas, pelo conhecimento transmitido e pela disposição em ajudar, em especial ao prof. Daniel que sempre se prontificou a tirar dúvidas e dar novas contribuições.

Às funcionárias da secretaria e da limpeza, sempre muito prestativas e atenciosas.

À Larissa, companheira de república e de curso, amiga fiel e presente desde o “bom dia” no café da manhã até o “boa noite” depois da novela. Foram mais de 5 anos de convivência diária. Uma irmã que ganhei de presente da Matemática/Estatística. Obrigada por me aguentar no sofá de casa, na salinha do DEX, nos congressos, nas viagens, na academia, nas madrugadas de muito estudo e até nas farras...

À Jussa e Tábata, que mesmo longe, nunca deixaram de mandar mensagens de apoio e incentivo. Sinto muita falta de vocês.

Mari e Jair, obrigada por sempre me acolherem tão bem em Londrina. Vocês vão fazer muita falta.

À galerinha da EST 2010/ EST 2012 com a qual tive a oportunidade de conviver e aprender muito. Levarei cada um de vocês para o resto da vida. Sucesso a todos! Com certeza nos veremos em muitos e muitos congressos pela vida a fora.

Aos irmãozinhos de orientação: Cris, Leandro, Carlos, Fernandinha e Luzia. Obrigada por dividirem comigo nosso ambiente de estudo e torná-lo sempre mais agradável.

À Universidade Tecnológica Federal do Paraná – campus Apucarana – por confiar no meu potencial e conceder a liberação para a conclusão do doutorado.

Ao prof. Yan, da University of Connecticut (UCONN) nos EUA, por ter me recebido e contribuído tanto para o desenvolvimento desse projeto. Sou imensamente agradecida pelos ensinamentos e sugestões sobre o tema da tese.

À turminha do “Huskies Forever”, em especial Mary e Nely; aos casais Guto/Yiqi e Flávia/Dijalma. Minha estadia nos Estados Unidos não teria sido tão prazerosa e divertida sem a presença de vocês. Obrigada pela companhia nos

passeios, pelas horas de conversa fiada, pelos lanches e almoços, por me incluírem no melhor time de voley... Enfim, por me adotarem e me fazerem sentir em casa mesmo estando tão longe, em outro país.

Aos tios e tias, padrinhos e madrinhas, primos e primas, obrigada pela torcida.

Aos amigos de Itaúna, Viçosa, Apucarana, Lavras e outros vários lugares pelos quais passei, obrigada por se fazerem presentes através de ligações e mensagens. Paulinha, para sempre minha “chaveirinho” e Carol, minha “irmã gêmea”, as melhores médicas do mundo. Larissa “Codinome”, minha “vaquinha” preferida, sinto falta até dos seus sustos planejados. Carine, prima/amiga adotada. Cecela, desde 1992. Eu tenho as melhores amigas do mundo.

À família Ferrazza, que me adotou e cuida de mim com tanto carinho. O que seria de mim sem as peripécias da Julinha?!

Enfim, obrigada a todos que participaram direta e indiretamente dessa conquista. Não consigo conter a alegria por realizar mais esse sonho: DOUTORA com muito orgulho!!!

RESUMO GERAL

Estudos sobre qualidade de ajuste de cópulas e testes para verificar dependência de valores extremos são de grande interesse na estatística e possuem relevância em outras áreas, como por exemplo na economia. Este trabalho trata desses dois temas e é apresentado em três capítulos. O primeiro cuida do referencial teórico que abrange medidas de associação por postos, cópulas e dependência em valores extremos; o segundo capítulo propõe dois novos testes de verificação de qualidade de ajuste para cópulas bivariadas, baseados nas marginais nas diagonais principal e secundária; o terceiro propõe um método para verificar se um determinado conjunto de dados bivariados possui dependência de valor extremo, através da estimação não-paramétrica da função de Pickands, com base na otimização de polinômios de Bernstein.

Palavras-chave: qualidade de ajuste, cópulas, função de Pickands, polinômios de Bernstein, dependência de valores extremos.

GENERAL ABSTRACT

Studies about goodness of fit for copula and tests to verify the dependence of extreme values are of great interest in statistic and it are relevant in other areas such as the economy. This thesis approaches these two issues and it is presented in three chapters. The first chapter takes care of the theoretical framework that it includes measures of association for posts, copula and dependence of extreme values; the second proposes two new tests about verification to goodness of fit for bivariate copula, and it based on marginal in principal and secondary diagonals; the third proposes a method to verify that a given set of bivariate data has dependency of extreme value, by nonparametric estimating of the Pickands function and it based on Bernstein polynomial optimization.

Keywords: goodness of fit, copula, Pickands function, polynomials of Bernstein, dependence of extreme values.

LISTA DE FIGURAS

Figura 1 Regiões do quadrado unitário.....	27
Figura 2. Propriedade 1 da definição de cópula.....	38
Figura 3. Propriedade 2 da definição de cópula.....	38
Figura 4. Cópula da independência.....	39
Figura 5. Cópula do limite inferior da Desigualdade de <i>Fréchet-Hoeffdin</i>	40
Figura 6. Cópula do limite superior da Desigualdade de <i>Fréchet-Hoeffdin</i>	40
Figura 7. Região onde a cópula está delimitada	42
Figura 8. Comportamento gráfico das famílias de cópulas.....	44
Figura 9. Dependência nas caudas.....	51
Figura 10. Função de dependência de Pickands para a cópula Gumbel-Hougaard com $\tau = 0.50$	59
Figura 11. Gráfico dos polinômios $b_{n,k}$	65
Figura 12. Rotação do suporte da cópula.....	84
Figura 13. Gráfico dos dados de uma amostra tamanho $n = 1.000.000$ da cópula da independência plotados nas diagonais principal e secundária.....	86
Figura 14 Divisão das três regiões definidas por retas ortogonais à diagonal	87
Figura 15. Família de cópulas na diagonal secundária.	90
Figura 16. Família de cópulas na diagonal principal.	91
Figura 17. A-plot de uma cópula de valor extremo.	117

LISTA DE TABELAS

Tabela 1. As cópulas e suas expressões e limites paramétricos.....	43
Tabela 2. Relação entre os parâmetros das famílias de cópulas e o TAU de Kendall	45
Tabela 3 Parâmetro das cópulas em relação ao TAU de Kendall.....	72
Tabela 4 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Clayton a nível de 5% de significância.....	92
Tabela 5 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Gumbel a nível de 5% de significância.....	93
Tabela 6 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas da Normal a nível de 5% de significância.....	94
Tabela 7 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Frank a nível de 5% de significância.....	95
Tabela 8 Taxa de erro tipo I e erro tipo II no IC90% da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$	97
Tabela 9 Taxa de erro tipo I e erro tipo II no IC90% da família de cópulas de Gumbel, dado uma amostra de tamanho $n = 600$	98
Tabela 10 Taxa de erro tipo I e erro tipo II no IC90% da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$	99
Tabela 11 Taxa de erro tipo I e erro tipo II no IC90% da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$	100
Tabela 12. Erro quadrático médio (IMSE) da função ajustada de Pickands, sob 1000 repetições, para a cópula Gumbel. As amostras têm tamanho n	

	= 100, n = 200 e n = 500 e os valores assumidos por TAU de Kendall são 0,25; 0,50 e 0,75.	121
Tabela 13	Comparação entre a velocidade de execução (em segundos) do método proposto, utilizando a três combinações de restrições, e o método cobs do R e a relação entre o cobs e a média das três combinações de restrições.	122
Tabela 14.	Proporção de rejeição de H_0 (em %) para alguns testes realizados a nível nominal de 5% utilizando amostras de tamanho 200 a partir de cinco famílias de cópulas.	124

SUMÁRIO

CAPÍTULO 1	INTRODUÇÃO GERAL.....	14
1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO	16
2.1	Medidas de associação por postos	16
2.1.1	TAU de Kendall	17
2.1.2	O RHO de Spearman	22
2.2	Distribuições de valores extremos	29
2.2.1	O caso univariado	29
2.3	Cópulas	37
2.3.1	Teorema de Sklar	42
2.3.2	Algumas Famílias de Cópulas Bivariadas	43
2.3.3	Cópulas e o TAU de Kendall	44
2.3.4	Cópulas e o RHO de Spearman.....	45
2.3.5	Dependência positiva em quadrante.....	45
2.3.6	Dependência nas Caudas	48
2.3.7	Teoria de valor extremo para o caso bivariado	52
2.3.8	Cópulas de valores extremos	55
2.3.9	Polinômios de Bernstein.....	64
2.3.11	Ajuste de Cópulas.....	71
2.3.12	Qualidade de ajuste de Cópulas	72
3	CONSIDERAÇÕES GERAIS.....	75
	REFERÊNCIAS.....	77
CAPÍTULO 2	QUALIDADE DE AJUSTE DE CÓPULAS	
	VIA MARGINAIS NAS DIAGONAIS.....	80
1	INTRODUÇÃO	82
2	OBJETIVOS	83
3	METODOLOGIA	84
3.1	Teste 1: Qui-quadrado	88
3.2	Teste 2: Intervalo de confiança	89
4	RESULTADOS E DISCUSSÃO	90
4.1	Parte 1: Qui-quadrado.....	91
4.2	Parte 2: Intervalo de confiança	96
5	CONCLUSÃO	102
	REFERÊNCIAS.....	103
CAPÍTULO 3	TESTE PARA DEPENDÊNCIA EM	
	VALORES EXTREMOS, UTILIZANDO CÓPULAS	104
1	INTRODUÇÃO	106
2	OBJETIVOS	109

3	METODOLOGIA	110
3.1	Estimação da função de Pickands utilizando polinômios de Bernstein	110
3.2	Implementação do estimador da função de Pickands	113
3.3	Performance do estimador da função de Pickands	115
3.4	Um teste de significância para cópulas de valor extremo	116
3.5	Validação do teste	119
4	RESULTADOS E DISCUSSÃO	120
5	CONCLUSÃO	125
	REFERÊNCIAS.....	126

CAPÍTULO 1 INTRODUÇÃO GERAL

1 INTRODUÇÃO

É crescente o número de pesquisas relacionadas à qualidade do ajuste de cópulas bem como o estudo de cópulas de valores extremos. A teoria geral sobre cópulas pode ser encontrada em Joe (1997), Nelsen (2013) e em Cherubini et al. (2004).

As cópulas são de grande interesse na estatística, principalmente, por se tratar de funções que permitem separar o comportamento marginal das variáveis aleatórias de sua estrutura de dependência. A construção da distribuição conjunta de variáveis por cópulas é razoável, uma vez que não há restrições quanto às distribuições marginais das variáveis envolvidas.

Segundo Cherubini et al. (2004), a forma como as cópulas representam uma distribuição conjunta, captando diversos tipos de dependência entre as variáveis e identificando as relações entre os valores extremos das variáveis analisadas, é uma grande vantagem em relação a outros modelos matemáticos. Aplicações importantes de valores extremos podem ser encontradas em Coles et al. (2013) ou em Cebrian et al. (2003), entre outros.

Este trabalho é apresentado em três capítulos: o primeiro trata do referencial teórico; o segundo capítulo propõe dois novos testes de verificação de qualidade de ajuste para cópulas bivariadas, baseados nas marginais nas diagonais principal e secundária; o terceiro mostra um método para verificar se um determinado conjunto de dados bivariados possui dependência de valor extremo, através da estimação não-paramétrica da função de Pickands, com base na otimização de polinômios de Bernstein.

2 REFERENCIAL TEÓRICO

2.1 Medidas de associação por postos

Sejam as variáveis aleatórias contínuas X e Y , definidas em um mesmo espaço de probabilidades (Ω, \mathcal{F}, P) , com distribuição conjunta F e marginais F_X e F_Y . Suponha uma amostra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Define-se o *posto* de x_j como a quantidade de valores de x menores ou iguais a x_j , ou seja,

$$Posto(x_j) = \sum_{i=1}^n I_{\{x_i \leq x_j\}}(x_i) = \#\{i | x_i \leq x_j\} = p_j$$

e, da mesma maneira,

$$Posto(y_j) = \sum_{i=1}^n I_{\{y_i \leq y_j\}}(y_i) = \#\{i | y_i \leq y_j\} = q_j.$$

Um par de valores (x_i, y_i) e (x_j, y_j) é dito *concordante* se $(x_i - x_j)(y_i - y_j) > 0$. Caso $(x_i - x_j)(y_i - y_j) < 0$ diz-se que o par é *discordante*. É importante observar que, tanto concordância quanto discordância são características simétricas, ou seja, a ordem em que cada par é tomado é irrelevante. Mais explicitamente:

$$(x_i - x_j)(y_i - y_j) > 0 \Leftrightarrow (x_j - x_i)(y_j - y_i) > 0.$$

A forma de associação definida acima é conhecida como *concordância do tipo 1*.

2.1.1 TAU de Kendall

TAU de Kendall amostral

Definição 1: Sejam c e d os números de pares concordantes e discordantes, respectivamente, em uma amostra de tamanho n . Nelsen (2013) define o coeficiente *TAU de Kendall amostral* como:

$$\begin{aligned} \tau &= \\ &= \frac{c - d}{c + d} \end{aligned} \quad (1)$$

Note que, como neste trabalho serão consideradas variáveis aleatórias contínuas ($P[X_j = X_i] = P[Y_j = Y_i] = 0$), então não será considerada a possibilidade de se encontrar $(x_i - x_j)(y_i - y_j) = 0$ e, $c + d$ é o número total de pares possíveis em uma amostra de tamanho n , dado por $c + d = \binom{n}{2} = \frac{n(n-1)}{2}$. Pode-se, então, reescrever (1) como:

$$\begin{aligned} \tau = \frac{c-d}{c+d} &= \frac{c-d}{\frac{n(n-1)}{2}} = \frac{c - \left(\frac{n(n-1)}{2} - c\right)}{\frac{n(n-1)}{2}} = \frac{2c - \frac{n(n-1)}{2}}{\frac{n(n-1)}{2}} = \frac{4c}{n(n-1)} - 1 = 1 - \\ &\frac{4d}{n(n-1)} \end{aligned} \quad (2)$$

Observações:

- i) A expressão (2) mostra que $\tau = 1$, caso haja concordância perfeita, ou seja, caso o número de pares discordantes seja nulo ($d = 0$).
- ii) $\tau = -1$, caso haja discordância perfeita ($c = 0$).
- iii) O valor de τ é crescente com o número de concordâncias.
- iv) O valor de τ é invariante para transformações que preservem a ordem, ou seja, transformações estritamente

crescentes em x e/ou y . Além disto, é também invariante para transformações estritamente decrescentes em x e y .

- v) O valor $\tau = 0$, apesar de ser um indicativo, não deve ser tomado como um determinante de independência, de acordo com Kendall (1962).

As observações de (i) a (iv) acima mostram características extremamente desejáveis em uma medida de associação entre variáveis aleatórias. A quarta observação, além de deixar claro o caráter não paramétrico do TAU de Kendall, permite ainda que seu valor seja computado a partir dos postos das amostras.

Para ilustrar o cálculo e a interpretação do TAU de Kendall, será utilizado o exemplo a seguir, retirado de Kendall (1962).

Suponha que, em uma amostra de 10 alunos, tenham sido observadas as notas em matemática e música, dadas abaixo:

Aluno:	A	B	C	D	E	F	G	H	I	J
Matem.:	7,4	3,5	2,8	9,8	6,3	2,4	9,6	7,6	1,1	5,9
Música:	5,7	6,8	3,2	9,9	1,5	8,7	6,7	2,7	7,8	3,5

Passando dos valores das notas, para os respectivos postos, temos:

Aluno:	A	B	C	D	E	F	G	H	I	J
Matem.:	7	4	3	10	6	2	9	8	1	5
Música:	5	7	3	10	1	9	6	2	8	4

Deseja-se saber se existe alguma relação entre habilidades para matemática e para música. Uma simples observação, nos postos, mostra que estão longe da concordância perfeita: alguns alunos ocupam a mesma posição em ambas as ordenações (C e D) enquanto outros ocupam

O número de pares concordantes é $c = 21$, o de discordantes é $d = 24$ e o TAU de Kendall:

$$\tau = \frac{21-24}{45} = -0.07.$$

Este valor, muito próximo de zero, é levado a supor uma associação muito pequena entre as habilidades dos alunos para matemática e para música.

TAU de Kendall populacional

Considerando que o TAU de Kendall amostral é dado por:

$$\tau = \frac{n^\circ \text{ de pares concordantes}}{n^\circ \text{ total de pares possíveis}} - \frac{n^\circ \text{ de pares discordantes}}{n^\circ \text{ total de pares possíveis}}$$

Sua interpretação populacional é imediata, dada pela seguinte definição:

Definição: Sejam (X_1, Y_1) e (X_2, Y_2) duas cópias independentes de uma população (X, Y) , com distribuição conjunta $F_{X,Y}$ e marginais F_X e F_Y . Define-se a probabilidade de concordância (do tipo 1) como:

$$\begin{aligned} \mathcal{P}_{c1} &= P[(X_2 - X_1)(Y_2 - Y_1) > 0] \\ &= P[(X_2 > X_1) \cap (Y_2 > Y_1)] + P[(X_2 < X_1) \cap (Y_2 < Y_1)] \\ &= P[(X_1 < X_2) \cap (Y_1 < Y_2)] + P[(X_2 < X_1) \cap (Y_2 < Y_1)] \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[(X_1 < x_2) \cap (Y_1 < y_2)] f_{X_2, Y_2}(x_2, y_2) dx_2 dy_2 + \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[(X_2 < x_1) \cap (Y_2 < y_1)] f_{X_1, Y_1}(x_1, y_1) dx_1 dy_1 = \\
&= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X, Y}(x, y) f_{X, Y}(x, y) dx dy
\end{aligned}$$

e a probabilidade de discordância (do tipo 1) como:

$$\begin{aligned}
\mathcal{P}_{d1} &= P[(X_2 - X_1)(Y_2 - Y_1) < 0] \\
&= P[(X_2 > X_1) \cap (Y_2 < Y_1)] + P[(X_2 < X_1) \cap (Y_2 > Y_1)]
\end{aligned}$$

e o TAU de Kendall populacional como:

$$\tau = \mathcal{P}_{c1} - \mathcal{P}_{d1}$$

Levando em consideração populações contínuas tem-se:

$$P[(X_2 - X_1)(Y_2 - Y_1) = 0] = 0$$

e, portanto,

$$\mathcal{P}_{c1} + \mathcal{P}_{d1} = 1 \Rightarrow \tau = 2\mathcal{P}_{c1} - 1 = 1 - 2\mathcal{P}_{d1}$$

Assim como no caso amostral, o TAU de Kendall populacional também tem algumas propriedades desejáveis. São elas:

i) τ assume valores entre -1 e 1 e é crescente com \mathcal{P}_{c1} .

ii) $\tau = 1$, caso haja concordância perfeita, ou seja:

$$X_1 < X_2 \Leftrightarrow Y_1 < Y_2 \text{ e } X_1 > X_2 \Leftrightarrow Y_1 > Y_2$$

iii) $\tau = -1$, caso haja discordância perfeita.

iv) Se X e Y são independentes, então $\tau = 0$.

As provas das propriedades *i*, *ii* e *iii* são triviais. Apenas a propriedade *iv* será demonstrada.

Como X_1 e X_2 são independentes e identicamente distribuídos, então $P[X_1 > X_2] = P[X_1 < X_2]$. Se, além disso, X e Y são independentes,

$$\begin{aligned} P[(X_2 > X_1) \cap (Y_2 > Y_1)] &= P[(X_2 > X_1)]P[(Y_2 > Y_1)] \\ &= P[(X_2 < X_1)]P[(Y_2 > Y_1)] \\ &= P[(X_2 < X_1) \cap (Y_2 > Y_1)]; \end{aligned}$$

$$\begin{aligned} P[(X_2 < X_1) \cap (Y_2 < Y_1)] &= P[(X_2 < X_1)]P[(Y_2 < Y_1)] \\ &= P[(X_2 > X_1)]P[(Y_2 < Y_1)] \\ &= P[(X_2 > X_1) \cap (Y_2 < Y_1)] \end{aligned}$$

Pode-se, então, reescrever a expressão de \mathcal{P}_{c1} como:

$$\begin{aligned} \mathcal{P}_{c1} &= P[(X_2 > X_1) \cap (Y_2 > Y_1)] + P[(X_2 < X_1) \cap (Y_2 < Y_1)] \\ &= P[(X_2 < X_1) \cap (Y_2 > Y_1)] + P[(X_2 > X_1) \cap (Y_2 < Y_1)] \\ &= \mathcal{P}_{d1} \end{aligned}$$

$$\Rightarrow \tau = \mathcal{P}_{c1} - \mathcal{P}_{d1} = 0$$

2.1.2 O RHO de Spearman

O RHO de Spearman amostral

Suponha uma amostra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ de uma população bivariada (X, Y) . Sejam $(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)$ os respectivos postos. Seja S a soma dos quadrados das diferenças $d_i = (p_i - q_i)$, ou seja,

$$S = \sum_{i=1}^n (p_i - q_i)^2.$$

É trivial que o valor de S não se altera se as amostras forem reorganizadas de maneira em que os postos de X apareçam em ordem crescente:

$$\begin{aligned} \text{posto}(X) &= 1 & 2 & \dots & n \\ \text{posto}(Y) &= s_1 & s_2 & \dots & s_n \end{aligned}$$

O menor valor de S acontece se as amostras são perfeitamente concordantes:

$$\begin{aligned} \text{posto}(X) &= 1 & 2 & \dots & n \\ \text{posto}(Y) &= 1 & 2 & \dots & n \end{aligned} \Rightarrow S = 0.$$

■

O maior valor de S acontece se as amostras são perfeitamente discordantes, ou seja:

$$\begin{aligned} \text{posto}(X) &= 1 & 2 & \dots & i & \dots & n \\ \text{posto}(Y) &= n & n-1 & \dots & n-(i-1) & \dots & 1 \\ d &= 1-n & 3-n & \dots & 2i-1-n & \dots & n-1 \end{aligned}$$

$$\begin{aligned} S &= \sum_{i=1}^n d_i^2 = \sum_{i=1}^n [2i - (n+1)]^2 = \\ &= 4 \sum_{i=1}^n i^2 - 4(n+1) \sum_{i=1}^n i + n(n+1)^2 \\ &= 4 \frac{n(n+1)(2n+1)}{6} - 4(n+1) \frac{n(n+1)}{2} + n(n+1)^2 \\ &= n(n+1) \left[\frac{2(2n+1)}{3} - 2(n+1) + (n+1) \right] \\ &= \frac{n(n+1)}{3} [2(2n+1) - 3(n+1)] = \frac{n(n^2-1)}{3} \\ &\Rightarrow S = \frac{n(n^2-1)}{3}. \end{aligned}$$

■

Como $\frac{3S}{n(n^2-1)}$ varia de zero, para concordância perfeita entre X e Y , até um, para discordância perfeita, e define-se o *RHO de Spearman amostral* como:

$$R = 1 - \frac{6S}{n(n^2 - 1)}$$

O valor de R varia entre -1 (discordância perfeita) e 1 (concordância perfeita).

Retomando ao exemplo dos alunos de matemática e música, têm-se:

Aluno:	I	F	C	B	J	E	A	H	G	D
Matem.:	1	2	3	4	5	6	7	8	9	10
Música:	8	9	3	7	4	1	5	2	6	10
$d =$	-7	-7	0	-3	1	5	2	6	3	0
$d^2 =$	49	49	0	9	1	25	4	36	9	0

$$S = 182 \quad R = 1 - \frac{6 * 182}{10(100 - 1)} = -0,103$$

O RHO de Spearman populacional

Sejam (X_i, Y_i) , (X_j, Y_j) e (X_k, Y_k) três cópias independentes de uma população bivariada e contínua (X, Y) . Diz-se que existe *concordância do tipo 2* se $(X_j - X_i)(Y_k - Y_i) > 0$ e que existe *discordância do tipo 2* se $(X_j - X_i)(Y_k - Y_i) < 0$. É conveniente obter expressões para as probabilidades dessa forma de associação entre (X, Y) .

$$\mathcal{P}_{c2} = P[(X_j - X_i)(Y_k - Y_i) > 0]$$

$$\begin{aligned}
&= P[(X_i < X_j) \cap (Y_i < Y_k)] + P[(X_i > X_j) \cap (Y_i > Y_k)] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{P[(X_i < x_j) \cap (Y_i < y_k)] + \\
&\quad + P[(X_i > x_j) \cap (Y_i > y_k)]\} f_{X_j, Y_k}(x_j, y_k) dx_j dy_k \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_{X,Y}(x, y) + 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y)] dF_X(x) dF_Y(y) \\
&= 1 + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x, y) dF_X(x) dF_Y(y) - \int_{-\infty}^{\infty} F_X(x) dF_X(x) + \\
&\quad - \int_{-\infty}^{\infty} F_Y(y) dF_Y(y) \\
&= \\
&= 1 + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x, y) dF_X(x) dF_Y(y) - \left[\frac{1}{2} F_X^2(x) \right]_{-\infty}^{+\infty} + \\
&\quad - \left[\frac{1}{2} F_Y^2(y) \right]_{-\infty}^{+\infty} \\
&= 1 + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x, y) dF_X(x) dF_Y(y) - \frac{1}{2} - \frac{1}{2} \\
&= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x, y) dF_X(x) dF_Y(y).
\end{aligned}$$

Se, em vez de resolver a integral em relação a $dx_j dy_k$, ela fosse resolvida em relação a $dx_i dy_i$ a solução seria a seguinte expressão equivalente:

$$\begin{aligned}
\mathcal{P}_{c2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{P[(X_j > x_i) \cap (Y_k > y_i)] + \\
&\quad + P[(X_j < x_i) \cap (Y_k < y_i)]\} f_{X_i Y_i}(x_i, y_i) dx_i dy_i \\
&= \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [1 - F_X(x) - F_Y(y) + F_X(x)F_Y(y) + \\
&\quad + F_X(x)F_Y(y)] dF_{X,Y}(x, y) \\
&= 1 + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_X(x)F_Y(y) dF_{X,Y}(x, y) - \int_{-\infty}^{\infty} F_X(x) dF_{X,Y}(x, y) + \\
&\quad - \int_{-\infty}^{\infty} F_Y(y) dF_{X,Y}(x, y) \\
&= 1 + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_X(x)F_Y(y) dF_{X,Y}(x, y) - \frac{1}{2} - \frac{1}{2} \\
&= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_X(x)F_Y(y) dF_{X,Y}(x, y)
\end{aligned}$$

Uma outra forma de se obter a concordância (ou a discordância) do tipo 2 é considerar um par (X, Y) da população bivariada $F_{X,Y}$, um valor X' tomado de forma independente na população marginal F_X e um valor Y' tomado também de forma independente em F_Y . Com isto, a probabilidade de concordância do tipo 2 pode ser reescrita como:

$$\mathcal{P}_{c2} = P[(X - X')(Y - Y') > 0]$$

e vista como uma nova forma de generalização da associação por quadrante.

Para obter os valores máximo e mínimo de \mathcal{P}_{c2} , observa-se que seu valor também é invariante para transformações que preservem ordem e aplica-se o teorema da probabilidade integral:

$$(X_1^*, Y_1^*) = (F_X(X), F_Y(Y))$$

$$X_2^* = F_X(X') \quad Y_2^* = F_Y(Y')$$

Com isto, todas as variáveis envolvidas passam a ter distribuição marginal uniforme no intervalo unitário e \mathcal{P}_{c2} pode ser reescrito como:

$$\mathcal{P}_{c2} = P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) > 0]$$

$$= E[P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) > 0] | (X_1^*, Y_1^*)]$$

A figura seguinte mostra as regiões do quadrado unitário em que $P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) > 0]$, dado o par (X_1^*, Y_1^*) :

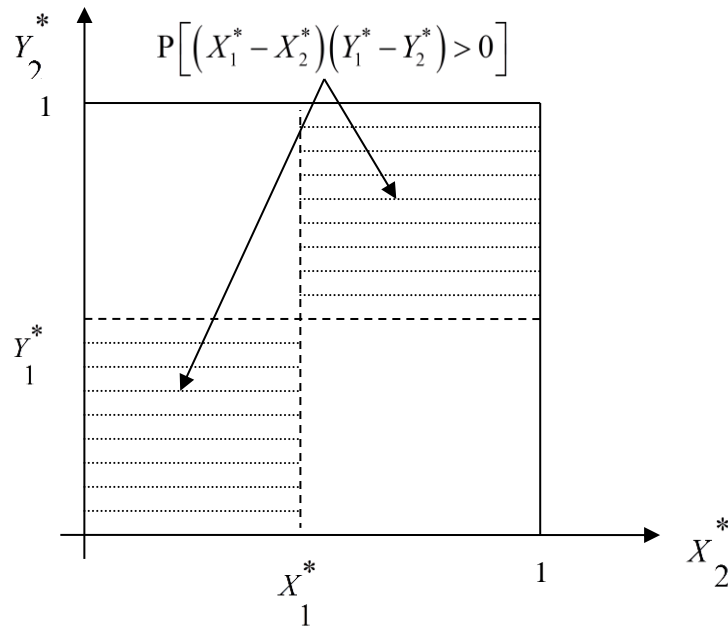


Figura 1 Regiões do quadrado unitário

Portanto:

$$\begin{aligned}
\mathcal{P}_{c2} &= E[(X_1^* Y_1^* + (1 - X_1^*)(1 - Y_1^*))] \\
&= E[X_1^* Y_1^* + 1 - X_1^* - Y_1^* + X_1^* Y_1^*] \\
&= 2E[X_1^* Y_1^*] \\
&= 2\{Cov[X_1^*, Y_1^*] + E[X_1^*]E[Y_1^*]\} \\
&= 2\left\{Cor[X_1^*, Y_1^*]\sqrt{Var[X_1^*]Var[Y_1^*]} + \frac{1}{2} \cdot \frac{1}{2}\right\} \\
&= 2\left\{Cor[X_1^*, Y_1^*]\sqrt{Var[X_1^*]Var[Y_1^*]} + \frac{1}{2} \cdot \frac{1}{2}\right\} \\
&= 2\left\{Cor[X_1^*, Y_1^*][E[X_1^{*2}] - (E[X_1^*])^2] + \frac{1}{4}\right\} \\
&= 2\left\{Cor[X_1^*, Y_1^*]\left[\frac{1}{3} - \left(\frac{1}{2}\right)^2\right] + \frac{1}{4}\right\} \\
&= 2\left\{Cor[X_1^*, Y_1^*]\frac{1}{12} + \frac{1}{4}\right\} \\
&= \frac{1}{6}Cor[X_1^*, Y_1^*] + \frac{1}{2}.
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_{d2} &= P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) < 0] \\
&= E[P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) < 0] | (X_1^*, Y_1^*)] \\
&= E[X_1^*(1 - Y_1^*) + Y_1^*(1 - X_1^*)] \\
&= E[X_1^* + Y_1^* - 2X_1^* Y_1^*] \\
&= 1 - 2E[2X_1^* Y_1^*] \\
&= 1 - \left(\frac{1}{6}Cor[X_1^*, Y_1^*] + \frac{1}{2}\right) \\
&= \frac{1}{2} - \frac{1}{6}Cor[X_1^*, Y_1^*].
\end{aligned}$$

Como o coeficiente de correlação varia entre -1 e 1 , \mathcal{P}_{c2} assume valores entre $\frac{1}{3}$ e $\frac{2}{3}$ e $\mathcal{P}_{c2} - \mathcal{P}_{d2}$ entre $-\frac{1}{3}$ e $\frac{1}{3}$. Define-se então o *RHO de Spearman populacional*, com notação ρ_s , como

$$\rho_s = 3(\mathcal{P}_{c2} - \mathcal{P}_{d2}) = 6\mathcal{P}_{c2} - 3 = 3 - 6\mathcal{P}_{d2}.$$

RHO de Spearman populacional, ρ_s , assume valores entre -1 e 1 e é invariante para transformações que preservem a ordem. Caso haja concordância do tipo 2 perfeita entre X e Y , isto é, caso toda a massa de probabilidade esteja concentrada no gráfico de alguma função crescente têm-se $\rho_s = 1$. Caso a função seja decrescente, perfeita discordância do tipo 2, $\rho_s = -1$.

2.2 Distribuições de valores extremos

2.2.1 O caso univariado

Suponha X_1, X_2, \dots, X_n variáveis aleatórias i.i.d. com distribuição contínua $F_X(x)$, densidade $f_X(x)$ e suporte $\mathbb{X} = \{x \in \mathbb{R} \mid f_X(x) > 0\}$. Considerando o vetor (X_1, X_2, \dots, X_n) , estamos interessados no comportamento assintótico ($n \rightarrow \infty$) de duas estatísticas: a estatística do máximo, definida como $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$, e a estatística do mínimo, $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$. Como $X_{(1)} = \max\{-X_1, -X_2, \dots, -X_n\}$, pode-se restringir o estudo à estatística do máximo. Sabemos que $F_{X_{(n)}}(x) = (F_X(x))^n = F_X^n(x)$. Seja $x_M = \inf \sup(\mathbb{X})$ (possivelmente igual a $+\infty$). Observe que o comportamento assintótico de $X_{(n)}$ simplesmente, é de pouco interesse, pois, se $x < x_M$, então

$\lim_{n \rightarrow \infty} F_{X_{(n)}}(x) = \lim_{n \rightarrow \infty} F_X^n(x) = 0$, e se $x \geq x_M$, então

$\lim_{n \rightarrow \infty} F_{X_{(n)}}(x) = \lim_{n \rightarrow \infty} F_X^n(x) = 1$. Ou seja, se $n \rightarrow \infty$, a estatística do

máximo degenera na singularidade $\lim_{n \rightarrow \infty} P[X_{(n)} = x_M] = 1$.

Observe ainda que singularidade semelhante acontece caso se

tome a distribuição da média amostral. Seja $S_n = \sum_{j=1}^n X_j$ e

$\bar{X}_n = S_n/n$. Nesse caso, se $x < E[X]$, então $\lim_{n \rightarrow \infty} P[\bar{X}_n < x] = 0$ e,

caso $x \geq E[X]$, temos $\lim_{n \rightarrow \infty} P[\bar{X}_n \leq x] = 1$. Porém, um resultado

extremamente útil é obtido quando se toma o limite

$$\lim_{n \rightarrow \infty} P \left[\frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{\text{var}[\bar{X}_n]}} \leq x \right] = \Phi_{(0,1)}(x).$$

Segundo Joe (1997), um resultado semelhante, com contribuições de Fisher, Tippett, von Mises, Gnedenko e de Haas, é obtido quando se

considera o limite da sequência $\lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq x \right]$. Com escolha

apropriada das constantes $\{a_n\}$ e $\{b_n\}$, o chamado **teorema de Fisher-**

Tippett afirma que as únicas distribuições limites possíveis são as famílias de locação e escala baseadas nas distribuições:

(1) $H_0(x) = \exp\{-e^{-x}\}$, $(-\infty < x < +\infty)$ (Distribuição de Gumbel)

(2) $H_1(x; \gamma) = \exp\{-x^{-\gamma}\}$, $(x > 0, \gamma > 0)$ (Distribuição de Fréchet)

(3) $H_{-1}(x; \gamma) = \exp\{-(-x)^\gamma\}$, ($x < 0, \gamma > 0$) (Distribuição de Weibull)

Caso o limite acima exista, diz-se que a distribuição $F_X(x)$ pertence ao **domínio de atração** da distribuição H correspondente.

As condições necessárias e suficientes para a convergência para cada uma dessas três distribuições podem ser encontradas em Galambos (1987), Resnick (1987) ou Mood et al. (1974).

Essas distribuições são casos particulares da chamada **família generalizada de distribuições de valor extremo**,

$$H(x; \theta) = \exp\left\{-\left(1 + \theta x\right)_+^{-1/\theta}\right\}, \quad (-\infty < x < +\infty, -\infty < \theta < +\infty)$$

em que $(w)_+ = \max\{0, w\}$.

É simples verificar que:

- (1) Se $\theta \rightarrow 0$, $H(x; \theta) \rightarrow H_0(x)$.
- (2) Se $\theta > 0$, $H(x; \theta) = H_1(1 + \theta x; 1/\theta)$.
- (3) Se $\theta < 0$, $H(x; \theta) = H_{-1}(-1 - \theta x; -1/\theta)$.

Vejamos exemplos de escolha das sequências $\{a_n\}$ e $\{b_n\}$ e obtenção de uma das distribuições limites (Joe (1997) e Mood et al. (1974)):

Exemplo 1: Considere uma amostra X_1, X_2, \dots, X_n i.i.d. da distribuição logística $F_X(x) = (1 + e^{-x})^{-1}$. Queremos determinar o

limite $\lim_{n \rightarrow \infty} P\left[\frac{X_{(n)} - a_n}{b_n} \leq x\right]$. Parece razoável que se escolha, como

constantes “centralizadoras” $\{a_n\}$, valores próximos de $E[X_{(n)}]$. Sabemos que $U_1 = F_X(X_1)$, $U_2 = F_X(X_2), \dots, U_n = F_X(X_n)$ é uma amostra i.i.d. de uma *Uniforme*(0,1) e que $U_{(n)} = F_X(X_{(n)})$. Então, é razoável supor $E[U_{(n)}] = E[F_X(X_{(n)})] \approx F_X(E[X_{(n)}])$.

Além disso,

$$\begin{aligned} F_X(x) + F_X(-x) &= (1 + e^{-x})^{-1} + (1 + e^x)^{-1} \\ &= \frac{1}{1 + e^{-x}} + \frac{1}{1 + e^x} = \frac{e^x}{1 + e^x} + \frac{1}{1 + e^x} = 1. \\ \Rightarrow E[F_X(X)] + E[F_X(-X)] &= 1. \quad (\text{A}) \end{aligned}$$

Como, $F_U(u) = u$ então $F_{U_{(n)}}(u) = u^n \Rightarrow f_{U_{(n)}}(u) = nu^{n-1}$.

$$\Rightarrow E[U_{(n)}] = \int_0^1 u n u^{n-1} du = \frac{n}{n+1} = E[F_X(X_{(n)})].$$

Segue que:

$$\begin{aligned} F_X(E[X_{(n)}]) &= \left(1 + \exp\{-E[X_{(n)}]\}\right)^{-1} \\ &= \frac{1}{1 + \exp\{-E[X_{(n)}]\}} = \frac{1}{1 + \frac{1}{\exp\{E[X_{(n)}]\}}} \\ &= \frac{\exp\{E[X_{(n)}]\}}{\exp\{E[X_{(n)}]\} + 1} = \frac{\exp\{E[X_{(n)}]\} + 1 - 1}{\exp\{E[X_{(n)}]\} + 1} \\ &= 1 - \left(1 + \exp\{E[X_{(n)}]\}\right)^{-1}. \quad (\text{B}) \end{aligned}$$

$$\begin{aligned}
&= 1 - F_X \left(-E \left[X_{(n)} \right] \right). \\
&= 1 - \left(1 - F_X \left(E \left[X_{(n)} \right] \right) \right). \quad (\text{usando (A)}) \\
&\approx 1 - \left(1 - E \left[F_X \left(X_{(n)} \right) \right] \right) = 1 - \left(1 - \frac{n}{n+1} \right) = 1 - (1+n)^{-1}. \quad (\text{C})
\end{aligned}$$

Comparando (C) com (B):

$$\Rightarrow \exp \left\{ E \left[X_{(n)} \right] \right\} \approx n \quad \Rightarrow E \left[X_{(n)} \right] \approx \ln(n).$$

Tomamos, então $\{a_n\} = \{\ln(n)\}$. Queremos:

$$\begin{aligned}
\lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq x \right] &= \lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - \ln(n)}{b_n} \leq x \right]. \\
&= \lim_{n \rightarrow \infty} P \left[X_{(n)} \leq b_n x + \ln(n) \right]. \\
&= \lim_{n \rightarrow \infty} \left[F_X \left(b_n x + \ln(n) \right) \right]^n. \\
&= \lim_{n \rightarrow \infty} \left[1 + \exp \left\{ -b_n x - \ln(n) \right\} \right]^{-n}. \\
&= \lim_{n \rightarrow \infty} \left[1 + \frac{1}{n} \exp \left\{ -b_n x \right\} \right]^{-n} = \exp \left\{ -e^{-b_n x} \right\}.
\end{aligned}$$

Finalmente, tomando $\{a_n\} = \{\ln(n)\}$ e $\{b_n\} = 1$,

$$\lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq x \right] = H_0(x). \quad (\text{A distr. de Gumbel}) \quad \#$$

Exemplo 2: Considere uma amostra X_1, X_2, \dots, X_n i.i.d. da distribuição de Pareto $F_X(x) = 1 - x^{-1/\gamma}$, ($x > 1$, $\gamma > 0$).

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq x \right] &= \lim_{n \rightarrow \infty} P \left[X_{(n)} \leq b_n x + a_n \right]. \\ &= \lim_{n \rightarrow \infty} \left[1 - (b_n x + a_n)^{-1/\gamma} \right]^n. \\ &= \lim_{n \rightarrow \infty} \left[1 - \frac{1}{n} x^{-1/\gamma} \right]^n \quad (\text{com } \{a_n\} = 0 \text{ e } \{b_n\} = n^\gamma). \\ &= \exp \left\{ -x^{-1/\gamma} \right\} = H_1(x; \gamma). \quad (\text{A distr. de Fréchet}) \quad \# \end{aligned}$$

Exemplo 3: Considere uma amostra X_1, X_2, \dots, X_n i.i.d. da distribuição Beta $F_X(x) = 1 - (1-x)^{-1/\gamma}$, ($0 < x < 1$, $\gamma < 0$).

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq x \right] &= \lim_{n \rightarrow \infty} P \left[X_{(n)} \leq b_n x + a_n \right]. \\ &= \lim_{n \rightarrow \infty} \left[1 - (1 - b_n x - a_n)^{-1/\gamma} \right]^n. \\ &= \lim_{n \rightarrow \infty} \left[1 - (1 - n^\gamma x - 1)^{-1/\gamma} \right]^n, \quad (\{a_n\} = 1 \text{ e } \{b_n\} = n^\gamma). \\ &= \lim_{n \rightarrow \infty} \left[1 - \frac{1}{n} (x)^{-1/\gamma} \right]^n = \exp \left\{ -(-w)^{-1/\gamma} \right\}, \quad (-1 < w < 0). \\ &= H_{-1}(w; \gamma). \quad (\text{a distribuição de Weibull}) \quad \# \end{aligned}$$

Exemplo 4: Uma aplicação. Suponha que $X_{(n)}^1, X_{(n)}^2, \dots, X_{(n)}^k$ sejam as estatísticas do máximo, tomadas em k vetores aleatórios i.i.d., todos de tamanho n e com distribuição $F_X(x)$. Nesse caso, se n é

suficientemente grande, a distribuição generalizada de valor extremo $H((x - \mu)/\sigma; \theta)$ pode ser usada como uma aproximação para $F_X^n(x)$.

Chamando de $\bar{F}_X(x) = P[X > x]$,

$$\bar{F}_X(x) = 1 - F_X(x) \approx 1 - H^{1/n}((x - \mu)/\sigma; \theta).$$

$$= 1 - \exp\left\{-\frac{1}{n}\left(1 + \theta \frac{x - \mu}{\sigma}\right)_+^{-1/\theta}\right\}.$$

$$\approx 1 - \left(1 - \frac{1}{n}\left(1 + \theta \frac{x - \mu}{\sigma}\right)_+^{-1/\theta}\right).$$

Usando $e^w \approx 1 + w$,

$$\text{Assim, } \bar{F}_X(x) = 1 - F_X(x) = \frac{1}{n}\left(1 + \theta \frac{x - \mu}{\sigma}\right)_+^{-1/\theta}.$$

Para $x > 0$ e um valor alto do limiar T , a distribuição condicional $P[X > T + x | X > T] = \bar{F}_X(T + x | T)$, pode ser obtida por:

$$\begin{aligned} \frac{\bar{F}_X(T + x)}{\bar{F}_X(T)} &= \frac{1 - F_X(T + x)}{1 - F_X(T)} \approx \left(\frac{\frac{1}{n}\left(1 + \theta \frac{x + T - \mu}{\sigma}\right)}{\frac{1}{n}\left(1 + \theta \frac{T - \mu}{\sigma}\right)}\right)_+^{-1/\theta} \\ &= \left(\frac{\left(1 + \theta \frac{T - \mu}{\sigma}\right) + \theta \frac{x}{\sigma}}{\left(1 + \theta \frac{T - \mu}{\sigma}\right)}\right)_+^{-1/\theta} \\ &= \left(1 + \frac{\theta x}{\sigma + \theta(T - \mu)}\right)_+^{-1/\theta} = \left(1 + \frac{\theta x}{\sigma^*}\right)_+^{-1/\theta}. \end{aligned}$$

Este resultado é conhecido como a **aproximação de Pareto** para a distribuição condicional na cauda superior de uma distribuição que esteja no domínio de atração de uma distribuição de valor extremo.

Um último resultado univariado, envolvendo valores extremos e transformações de variáveis (Joe, 1997):

Proposição: Sejam X_1, X_2, \dots variáveis aleatórias i.i.d. com distribuição $F_X(x)$ e a transformação estritamente crescente $X_j^* = r(X_j)$, $j=1, 2, \dots$ resulta nas v.a. i.i.d. X_1^*, X_2^*, \dots . Sejam $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ e $X_{(n)}^* = \max(X_1^*, X_2^*, \dots, X_n^*)$.

Suponha que $P\left[\frac{X_{(n)} - a_n}{b_n} \leq z\right] \rightarrow H(z)$ e que

$P\left[\frac{X_{(n)}^* - a_n^*}{b_n^*} \leq z^*\right] \rightarrow H^*(z^*)$, em que $H(z)$ e $H^*(z^*)$ sejam

distribuições de valor extremo. Seja $s(z^*) = \left[\frac{r^{-1}(a_n^* + b_n^* z^*) - a_n}{b_n}\right]$.

Então, $H^*(z^*) = H(s(z^*))$.

Prova:
$$\begin{aligned} H^*(z^*) &= \lim_{n \rightarrow \infty} P\left[\frac{X_{(n)}^* - a_n^*}{b_n^*} \leq z^*\right] \\ &= \lim_{n \rightarrow \infty} P\left[X_{(n)}^* \leq (a_n^* + b_n^* z^*)\right] \\ &= \lim_{n \rightarrow \infty} P\left[X_{(n)} \leq r^{-1}(a_n^* + b_n^* z^*)\right]. \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq \frac{r^{-1}(a_n^* + b_n^* z^*) - a_n}{b_n} \right] \\
&= \lim_{n \rightarrow \infty} P \left[\frac{X_{(n)} - a_n}{b_n} \leq s(z^*) \right] = H(s(z^*)).
\end{aligned}$$

■

2.3 Cópulas

No início deste trabalho, ficou explícito o interesse em medidas de associação invariantes para transformações que preservem ordem. Até este ponto só foram apresentadas medidas de associação baseadas nos postos das amostras. Uma outra transformação importante é aquela dada pela aplicação do teorema da transformação integral: se X é uma variável aleatória com distribuição acumulada $F_X(x)$ então a variável aleatória $W = F_X(X)$ tem distribuição uniforme no intervalo $[0,1]$. Como $F_X(x)$ é uma função crescente, é trivial que a transformação $W = F_X(X)$ preserva ordem.

Considere o intervalo fechado $I = [0,1]$ e o par ordenado $(u, v) \in I^2$. Define-se como cópula bidimensional $C(u, v)$ qualquer função $C: I^2 \rightarrow I$ que satisfaça as duas seguintes propriedades:

$$P1 - C(1, v) = v, C(u, 1) = u \text{ e } C(0, v) = C(u, 0) = 0.$$

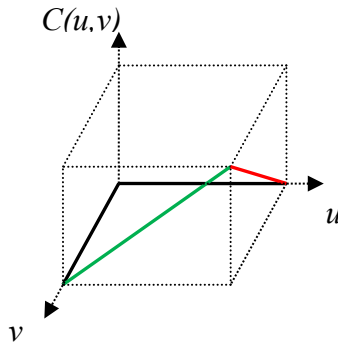


Figura 2. Propriedade 1 da definição de cópula.

$$P2 - V_c([a, b] \times [c, d]) = C(b, d) - C(a, d) - C(b, c) + C(a, c) \geq 0,$$

onde $a < b$ e $c < d$.

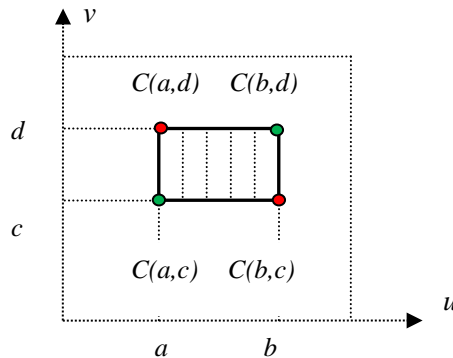


Figura 3. Propriedade 2 da definição de cópula.

Vale ressaltar que a função em P2 define o chamado *C-volume* do retângulo $[a, b] \times [c, d]$ e significa que $C(u, v)$ é uma função bi-crescente. Além disso, a cópula C induz uma medida de probabilidade em I^2 via $V_c([0, u] \times [0, v]) = C(u, v), \forall (u, v) \in I^2$.

Abaixo serão apresentados alguns exemplos importantes de cópulas e neles será verificado se as condições que caracterizam uma cópula são satisfeitas.

Exemplo 1 - $C_i(u, v) = uv$

$$P1 - C(1, v) = 1 \cdot v = v$$

$$C(u, 1) = u \cdot 1 = u$$

$$C(0, v) = 0 \cdot v = u \cdot 0 = C(u, 0) = 0$$

P2 - $C_i(u, v) = uv$ é a área do retângulo $[0, u] \times [0, v]$.

$$\begin{aligned} V_c([a, b] \times [c, d]) &= C(b, d) - C(a, d) - C(b, c) + C(a, c) \\ &= bd - ad - bc + ac = (d - c)(b - \end{aligned}$$

$a) \geq 0$

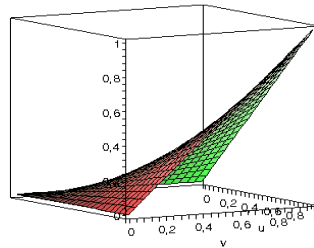


Figura 4. Cópula da independência

Exemplo 2 - $C_0(u, v) = \max\{u + v - 1, 0\}$

$$P1 - C_0(u, 0) = \max\{u + 0 - 1, 0\} = \max\{u - 1, 0\} = 0$$

$$C_0(0, v) = \max\{0 + v - 1, 0\} = \max\{v - 1, 0\} = 0$$

$$C_0(u, 1) = \max\{u + 1 - 1, 0\} = \max\{u, 0\} = u$$

$$C_0(1, v) = \max\{1 + v - 1, 0\} = \max\{v, 0\} = v$$

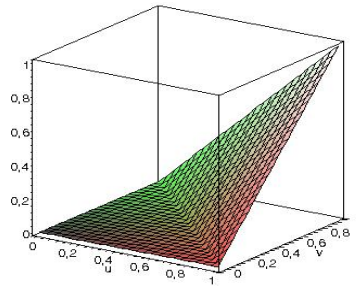


Figura 5. Cópula do limite inferior da Desigualdade de *Fréchet-Hoeffdin*

Exemplo 3 - $C_1(u, v) = \min\{u, v\}$

$$P1 - C_1(u, 0) = \min\{u, 0\} = 0$$

$$C_1(0, v) = \min\{0, v\} = 0$$

$$C_1(u, 1) = \min\{u, 1\} = u$$

$$C_1(1, v) = \min\{1, v\} = v$$

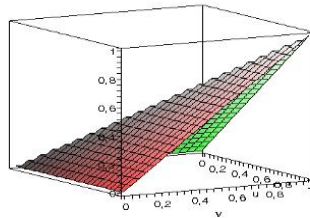


Figura 6. Cópula do limite superior da Desigualdade de *Fréchet-Hoeffdin*

Os dois últimos exemplos compõem a Desigualdade de *Fréchet-Hoeffdin*, que pode ser encontrada em Nelsen (2013) e diz que se $C(u, v)$ é uma cópula, então, para todo $(u, v) \in [0,1] \times [0,1]$ tem-se:

$$M = \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} = W$$

Para verificar esse resultado, tome

$\bar{F}_{X,Y}(x, y) = P[X > x, Y > y]$, segue que:

$$F_X(x) + F_Y(y) - F_{X,Y}(x, y) + \bar{F}_{X,Y}(x, y) = 1.$$

$$\Leftrightarrow F_{X,Y}(x, y) - \bar{F}_{X,Y}(x, y) = F_X(x) + F_Y(y) - 1.$$

$$\Rightarrow F_{X,Y}(x, y) \geq F_X(x) + F_Y(y) - 1.$$

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

Como $C(u, v) \geq 0$, tem-se que $C(u, v) \geq \max(u + v - 1, 0)$.

E ainda,

$$\left. \begin{array}{l} F_{X,Y}(x, y) \leq F_X(x) \Rightarrow C(u, v) \leq u \\ F_{X,Y}(x, y) \leq F_Y(y) \Rightarrow C(u, v) \leq v \end{array} \right\} \Rightarrow C(u, v) \leq \min\{u, v\}$$

Ou seja, toda cópula está contida na região de $[0, 1]^2$ delimitada inferiormente pela cópula $M = \max\{u + v - 1, 0\}$ e superiormente por $W = \min\{u, v\}$ como mostra a Figura 7.

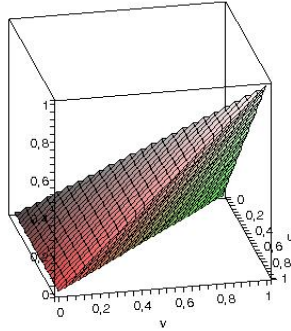


Figura 7. Região onde a cópula está delimitada

Em outras palavras, dada uma distribuição $F_{X,Y}(x, y)$, se os eixos X e Y são “comprimidos” para o intervalo $[0,1]$ via transformações $U = F_X(X)$ e $V = F_Y(Y)$, o resultado é uma nova distribuição de probabilidade $F_{U,V}(u, v)$, de marginais uniformes em $[0,1]$, denominada cópula.

2.3.1 Teorema de Sklar

De acordo com Sklar (1959), seja $F_{X,Y}(x, y)$ a distribuição conjunta de (X, Y) com marginais $F_X(x)$ e $F_Y(y)$. Então, existe uma cópula C tal que $C(F_X(x), F_Y(y)) = F_{X,Y}(x, y)$. Se $F_X(x)$ e $F_Y(y)$ são funções contínuas, a cópula C é única.

E ainda, para quaisquer funções de distribuição de probabilidade $F_X(X)$ e $F_Y(Y)$ e qualquer cópula $C(u, v)$, tem-se que a função H definida implicitamente pela equação $H(F_X^{-1}(u), F_Y^{-1}(v)) = C(u, v)$ é uma conjunta admissível.

2.3.2 Algumas Famílias de Cópulas Bivariadas

A literatura apresenta inúmeras famílias de cópulas, dentre elas, as que serão abordadas nesse trabalho: Clayton, Gumbel, Normal, Frank e t-Student. As expressões de cada uma dessas famílias de cópulas bivariadas e os limites de seus parâmetros podem ser encontrados na Tabela 1.

Tabela 1. As cópulas e suas expressões e limites paramétricos.

Família	Expressão $C(u, v)$	Parâmetro
Clayton	$\max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta}, 0\right)$	$\theta \geq 0$
Gumbel	$\exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$	$\theta \geq 1$
Normal	$N_{X,Y}\left(N_U^{-1}(u), N_V^{-1}(v)\right)$	$-1 \leq \rho \leq 1$
Frank	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right)$	$-1 \leq \theta \leq 1, \theta \neq 0$
t-Student	$t_{2,v}\left(t_v^{-1}(u), t_v^{-1}(v)\right)$	$-1 \leq \rho \leq 1$

Como se pode observar na Figura 8, as famílias de cópulas possuem o comportamento de suas marginais semelhante, apresentando uma pequena diferença apenas em suas extremidades.

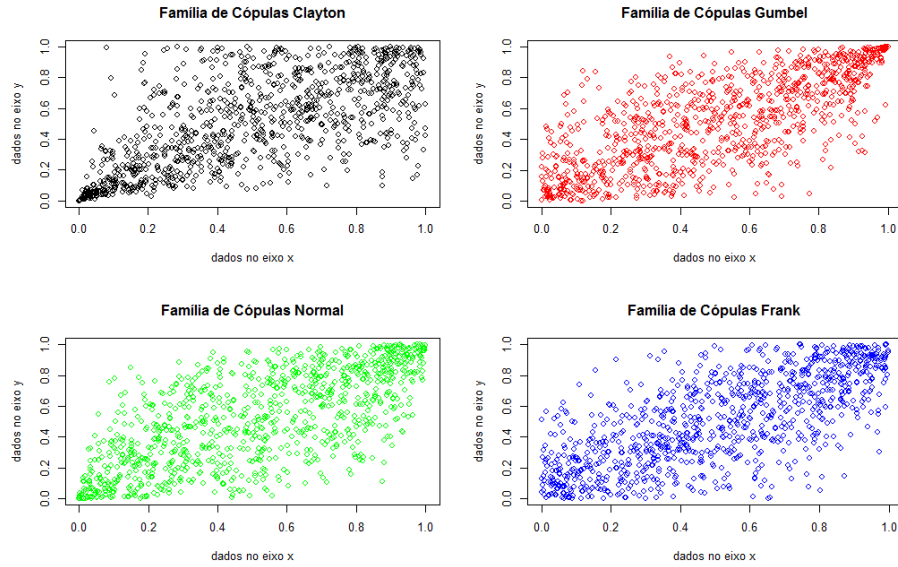


Figura 8. Comportamento gráfico das famílias de cópulas

2.3.3 Cópulas e o TAU de Kendall

Usando a seguinte relação de Tau de Kendall com \mathcal{P}_{c1} , $\tau = 2\mathcal{P}_{c1} - 1$, é imediato que:

$$\begin{aligned} \tau &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x,y) f_{X,Y}(x,y) dx dy - 1. \\ &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(u,v) dC(u,v) - 1. \end{aligned}$$

Abaixo temos a relação entre o parâmetro de algumas famílias de cópulas e o TAU de Kendall.

Tabela 2. Relação entre os parâmetros das famílias de cópulas e o TAU de Kendall

Família	TAU de Kendall	Parâmetro
Clayton	$\tau = \frac{\theta}{(\theta+2)}$	$0 < \tau \leq 1$
Gumbel	$\tau = 1 - \theta^{-1}$	$0 \leq \tau \leq 1$
Normal	$\tau = \frac{2}{\pi} \arcsin(\rho)$	$-1 \leq \tau \leq 1$
Frank	$\tau = 1 - \frac{4}{\theta} \left(1 - \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt \right)$	$-1 \leq \tau \leq 1 - \{0\}$

2.3.4 Cópulas e o RHO de Spearman

Sabe-se que RHO de Spearman é dado por $\rho_S = 6\mathcal{P}_{c1} - 3$.

Portanto:

$$\begin{aligned} \rho_S &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_X(x)F_Y(y) dF_{X,Y}(x, y) - 3. \\ &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv dC(u, v) - 3. \end{aligned}$$

2.3.5 Dependência positiva em quadrante

Diz-se que duas variáveis aleatórias X e Y têm dependência positiva em quadrante se:

$$P[X > x, Y > y] \geq P[X > x]P[Y > y], \quad \forall (x, y) \in \mathbb{R}^2. \quad (1)$$

Significa que o par (X, Y) é mais provável de ocorrer em qualquer região da forma $(x, +\infty) \times (y, +\infty)$ do que seria se X e Y fossem independentes.

Proposição: (Scaillet, 2005)

$$P[X > x, Y > y] \geq P[X > x]P[Y > y] \text{ é equivalente a}$$

$$P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y].$$

$$\text{Prova: } P[X > x, Y > y] \geq P[X > x]P[Y > y] \quad \Leftrightarrow$$

$$1 - P[X \leq x] - P[Y \leq y] + P[X \leq x, Y \leq y]$$

$$\geq (1 - P[X \leq x])(1 - P[Y \leq y]).$$

$$\Leftrightarrow 1 - P[X \leq x] - P[Y \leq y] + P[X \leq x, Y \leq y]$$

$$\geq 1 - P[X \leq x] - P[Y \leq y] + P[X \leq x]P[Y \leq y].$$

$$\Leftrightarrow P[X \leq x, Y \leq Y] \geq P[X \leq x]P[Y \leq Y]. \quad (2)$$

■

Observe que a expressão (2) permite escrever a dependência positiva em quadrante em termos da função acumulada de probabilidade da forma $F_{X,Y}(x, y) \geq F_X(x)F_Y(y)$. Tem-se também uma interpretação equivalente em termos de cópulas. Supondo que $F_X(x)$ e $F_Y(y)$ sejam estritamente crescentes, a cópula definida por $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$, e seja $u = F_X(x)$ e $v = F_Y(y)$, então $x = F_X^{-1}(u)$ e $y = F_Y^{-1}(v)$. Dessa forma,

$$\begin{aligned} C(u, v) &= C(F_X(x), F_Y(y)) = F_{X,Y}(x, y) \\ &= P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y] = uv. \end{aligned}$$

Isto é, $C(u, v) \geq uv$, significando que, se X e Y têm dependência positiva em quadrante, a cópula correspondente é maior ou igual à cópula da independência $C(u, v) = uv$.

Uma aplicação do conceito de dependência positiva em quadrante é em teoria da confiabilidade. O tempo de funcionamento de uma máquina, que possui dois componentes, geralmente é modelado utilizando a independência $P[X > x]P[Y > y]$. No entanto, pode ser mais realista utilizando dependência $P[X > x, Y > y]$ e em certas situações é razoável supor dependência positiva em quadrante.

Exemplo 1: Considere a cópula de Ali-Mikhail-Haq dada por

$$C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)} \text{ em que } \theta > 0 \text{ e } u, v \in (0, 1) \text{ (Nelsen, 2013).}$$

2013).

Supondo $\theta < 1$, então como, $(1-u) < 1$ e $(1-v) < 1$, tem-se que $0 < (1-u)(1-v)\theta < 1$. Assim, $0 < 1 - \theta(1-u)(1-v) < 1$ e

$$\text{portanto } C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)} \geq uv.$$

Se $F_X(x) = (1 - e^{-x})$ e $F_Y(y) = (1 - e^{-y})$, a cópula define a conjunta:

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)) = C(1 - e^{-x}, 1 - e^{-y}).$$

$$= \frac{(1-e^{-x})(1-e^{-y})}{1-\theta(1-1+e^{-x})(1-1+e^{-y})} = \frac{(1-e^{-x})(1-e^{-y})}{1-\frac{\theta}{e^{x+y}}}.$$

Como $e^{x+y} > 1 \Rightarrow \frac{\theta}{e^{x+y}} < 1 \Rightarrow 0 < 1 - \frac{\theta}{e^{x+y}} \leq 1$. Assim,

$F_{X,Y}(x,y) \geq (1-e^{-x})(1-e^{-y}) = F_X(x)F_Y(y)$, e portanto a distribuição apresenta dependência positiva em quadrante.

2.3.6 Dependência nas Caudas

O conceito de dependência nas caudas está relacionado com a razão entre as caudas superiores ou inferiores de uma distribuição bivariada com a cauda da marginal, e é relevante tanto na caracterização quanto na obtenção das distribuições de valores extremos. A definição é dada em termos da cópula correspondente.

Definição: Considere uma cópula bivariada $C(u,v)$ e denote por:

$$\begin{aligned}\bar{C}(u,v) &= P[U > u, V > v] \\ &= 1 - P[U < u] - P[V < v] + P[U < u, V < v] \\ &= 1 - u - v + C(u,v).\end{aligned}$$

Se o limite à esquerda $\lim_{u \rightarrow 1^-} \frac{\bar{C}(u,u)}{1-u} = \lambda_U \in (0,1]$, diz-se que a

cópula C caracteriza dependência na cauda superior. Diz-se que não há dependência na cauda superior se $\lambda_U = 0$. De maneira semelhante, o

limite à direita $\lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} = \lambda_L \in (0, 1]$ caracteriza dependência na cauda inferior. Diz-se que não há dependência na cauda inferior se $\lambda_L = 0$.

Observe que, tomando a cópula do limite inferior da Desigualdade de Fréchet-Hoeffdin $C(u, v) = \max\{u + v - 1, 0\}$, temos:

$$\bar{C}(u, u) = 1 - u - u + \max\{2u - 1, 0\} = 1 - 2u + \max\{2u - 1, 0\}.$$

$$\text{Dessa forma, } \lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - 2u + \max\{2u - 1, 0\}}{1 - u} = 0$$

indicando que essa cópula não apresenta dependência na cauda superior.

Com a cópula do limite superior da Desigualdade de Fréchet-Hoeffdin, $C(u, v) = \min\{u, v\}$, tem-se:

$$\bar{C}(u, u) = 1 - u - u + \min\{u, u\} = 1 - 2u + u = 1 - u.$$

$$\text{Dessa forma, } \lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - u}{1 - u} = 1 \text{ indicando que essa}$$

cópula é caracterizada pela dependência na cauda superior.

Pela Desigualdade de Fréchet-Hoeffdin:

$$\max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} \text{ para toda cópula } C.$$

$$\text{Sendo assim, } \max\{u + v - 1, 0\} \leq \bar{C}(u, v) \leq \min\{u, v\}.$$

Dividindo todos os elementos da desigualdade por $(1 - u)$ e aplicando o limite quando $u \rightarrow 1^-$ em cada, tem-se:

$$0 = \lim_{u \rightarrow 1^-} \frac{\max\{u + v - 1, 0\}}{1 - u} \leq \lim_{u \rightarrow 1^-} \frac{\bar{C}(u, v)}{1 - u} \leq \lim_{u \rightarrow 1^-} \frac{\min\{u, v\}}{1 - u} = 1$$

. De onde segue que os valores de λ_L estão realmente no intervalo $[0, 1]$.

Para a cópula da independência dada por $C(u, v) = uv$, tem-se que $\bar{C}(u, u) = 1 - u - u + uu = 1 - 2u + u^2 = (1 - u)^2$ e então:

$$\lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{(1 - u)^2}{1 - u} = \lim_{u \rightarrow 1^-} (1 - u) = 0, \text{ indicando que a}$$

cópula da independência não apresenta dependência na cauda superior.

A definição de dependência nas caudas significa que $\bar{C}(u, u) = P[U > u, V > u]$. Como $U = F_X(X)$, $V = F_Y(Y)$ e se $x_1 = F_X^{-1}(u)$, $y_1 = F_Y^{-1}(u)$ então:

$$\begin{aligned} \bar{C}(u, u) &= P[U > u, V > u] = P[F_X(X) > u, F_Y(Y) > u]. \\ &= P[X > F_X^{-1}(u), Y > F_Y^{-1}(u)]. \\ &= P[X > x_1, Y > y_1]. \end{aligned}$$

Dessa forma $\bar{C}(u, u)$ representa o volume da cauda da distribuição $F_{X,Y}$ no quadrante superior $\{(x, y), x \geq x_1, y \geq y_1\}$,

$$\begin{aligned} 1 - u &= 1 - P[U \leq u] = 1 - P[F_X(X) \leq u]. \\ &= 1 - P[X \leq F_X^{-1}(u)] = 1 - P[X \leq x_1]. \\ &= P[X > x_1]. \end{aligned}$$

e, portanto, significa o volume da cauda da distribuição $F_{X,Y}$ na faixa $\{(x, y), x \geq x_1\}$ (veja a figura 9).

A dependência nas caudas é o limite da razão entre estes dois volumes quando $u \rightarrow 1$, isto é:

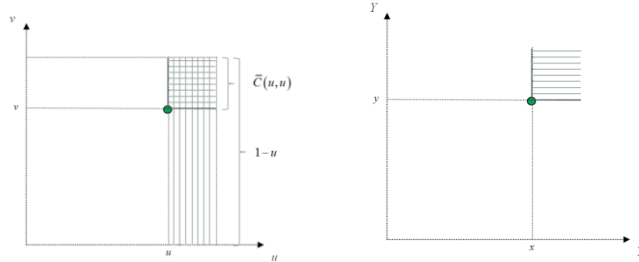


Figura 9. Dependência nas caudas.

Observe, como já demonstrado acima, que para a cópula da independência, $C(u, v) = uv$, tem-se:

$$\begin{aligned} \lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} \frac{P[X > x_1, Y > y_1]}{P[X > x_1]} &= \lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} \frac{P[X > x_1]P[Y > y_1]}{P[X > x_1]} \\ &= \lim_{y_1 \rightarrow \infty} P[Y > y_1] = 0. \end{aligned}$$

Os casos em que o limite $\lambda_1 \neq 0$ têm-se então o conceito que os volumes nas regiões $\{(x, y), x \geq x_1, y \geq y_1\}$ e $\{(x, y), x \geq x_1\}$ estão de alguma forma relacionados, isto é, tem-se uma dependência nas caudas.

Exemplo 1: Considere a cópula Gumbel expressa por:

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

$$\text{Então, } \bar{C}(u, u) = 1 - u - u + \exp\left\{-\left[(-\ln u)^\theta + (-\ln u)^\theta\right]^{1/\theta}\right\}.$$

$$\begin{aligned}
&= 1 - 2u + \exp\left\{-\left[2(-\ln u)^\theta\right]^{1/\theta}\right\}. \\
&= 1 - 2u + \exp\left\{-\left[2^{1/\theta}(-\ln u)\right]\right\}. \\
&= 1 - 2u + \exp\left\{2^{1/\theta} \ln u\right\} \\
&= 1 - 2u + \exp\left\{\ln u^{2^{1/\theta}}\right\}. \\
&= 1 - 2u + u^{2^{1/\theta}}.
\end{aligned}$$

Calculando o limite tem-se:

$$\begin{aligned}
\lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1 - u} &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u} = \lim_{u \rightarrow 1^-} \frac{-2 + 2^{1/\theta} u^{2^{1/\theta} - 1}}{-1}. \\
&= \lim_{u \rightarrow 1^-} \left(2 - 2^{1/\theta} u^{2^{1/\theta} - 1}\right) = 2 - 2^{1/\theta}.
\end{aligned}$$

Como $\theta \geq 1$, então $0 < 2 - 2^{1/\theta} \leq 1$ indicando que a cópula Gumbel é caracterizada pela dependência na cauda superior.

2.3.7 Teoria de valor extremo para o caso bivariado

Se (X, Y) é uma variável aleatória bidimensional, dizemos que X e Y apresentam dependência de valor extremo se

$$\lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} P[X > x_1 | Y > y_1] > 0.$$

Tal propriedade implica que $P[X > x_1 | Y > y_1] > 0$ quaisquer que sejam os valores x_1 e y_1 e, portanto, não importando o quão extremo

eles possam ser. Em termos da cópula,

$C(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$ temos:

$$\begin{aligned} P[X > x_1 | Y > y_1] &= P[F_X(X) > F_X(x_1) | F_Y(Y) > F_Y(y_1)] \\ &= P[U > u | V > v]. \end{aligned}$$

$$\text{Logo } \lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} P[X > x_1 | Y > y_1] = \lim_{\substack{u \rightarrow 1 \\ v \rightarrow 1}} P[U > u | V > v]. \quad \text{Em}$$

particular se y_1 é tal que $F_Y(y_1) = u = F_X(x_1)$ tem-se

$$\lim_{u \rightarrow 1} P[U > u | V > u] > 0.$$

Proposição: A v.a. bidimensional (X, Y) apresenta dependência nas caudas se, e somente se apresenta dependência de valor extremo.

$$\text{Prova: } \lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1-u} > 0 \Leftrightarrow \lim_{u \rightarrow 1} \frac{P[U > u, V > u]}{1-u} > 0.$$

$$\Leftrightarrow \lim_{u \rightarrow 1} \frac{P[U > u | V > u] P[V > u]}{1-u} = \lim_{u \rightarrow 1} P[U > u | V > u] > 0. \quad \blacksquare$$

Proposição: Se (X, Y) é variável aleatória bidimensional que apresenta dependência de valor extremo, isto é,

$$\lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} P[X > x_1 | Y > y_1] > 0$$

então a cópula correspondente:

$$C(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$$

apresenta dependência na cauda superior.

Prova:

$$\lim_{\substack{x_1 \rightarrow \infty \\ y_1 \rightarrow \infty}} P[X > x_1 | Y > y_1] > 0$$

em termos da cópula correspondente é:

$$\lim_{\substack{u_1 \rightarrow 1^- \\ v_1 \rightarrow 1^-}} P[U > u_1 | V > v_1] > 0$$

Logo,

$$\lim_{\substack{u_1 \rightarrow 1^- \\ v_1 \rightarrow 1^-}} P[U > u_1 | V > v_1] = \lim_{\substack{u_1 \rightarrow 1^- \\ v_1 \rightarrow 1^-}} \frac{P[U > u_1, V > v_1]}{P[V > v_1]} = \lim_{\substack{u_1 \rightarrow 1^- \\ v_1 \rightarrow 1^-}} \frac{\bar{C}(u_1, v_1)}{1 - v_1}.$$

Como se está supondo a existência do limite, fazendo-se o limite restrito a $u_1 = v_1$, tem-se:

$$\lim_{u_1 \rightarrow 1^-} \frac{\bar{C}(u_1, v_1)}{1 - v_1} > 0.$$

■

Exemplo 1: A cópula da independência não apresenta dependência de valor extremo pois, se $C(u, v) = uv$, então:

$$\begin{aligned} \Rightarrow \lim_{u \rightarrow 1^-} P[U > u | V > u] &= \lim_{u \rightarrow 1^-} \frac{P[U > u, V > u]}{P[V > u]} \\ &= \lim_{u \rightarrow 1^-} \frac{(1-u)(1-u)}{(1-u)} = \lim_{u \rightarrow 1^-} (1-u) = 0. \end{aligned}$$

A dependência positiva em quadrante não implica em dependência de valor extremo pois

$$P[X > x, Y > y] > P[X > x]P[Y > y] \text{ e,}$$

$$\begin{aligned} \lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} P[X > x | Y > y] &= \lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} \frac{P[X > x, Y > y]}{P[Y > y]} \\ &= \lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} \frac{P[X > x]P[Y > y]}{P[Y > y]} = \lim_{x \rightarrow \infty} P[X > x] = 0. \end{aligned}$$

2.3.8 Cópulas de valores extremos

No sentido de simplificar, serão apresentadas as definições apenas para o caso bivariado, sendo que as definições gerais são similares (Gudendorf e Segers, 2010).

Considere uma v.a. bidimensional (X, Y) com distribuição $F_{X,Y}$ e marginais F_X e F_Y . Suponha uma amostra i.i.d. $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$. Defina agora a v.a. bidimensional $M_n = (X_{(n)}, Y_{(n)})$ em que $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ e $Y_{(n)} = \max\{Y_1, Y_2, \dots, Y_n\}$.

Proposição: A densidade de M_n é dada por $F_{X_{(n)}, Y_{(n)}} = F_{X,Y}^n$ e marginais F_X^n e F_Y^n .

$$\begin{aligned} F_{X_{(n)}, Y_{(n)}}(x, y) &= P[X_{(n)} < x, Y_{(n)} < y] \\ &= P[X_1 < x, X_2 < x, \dots, X_n < x, Y_1 < y, Y_2 < y, \dots, Y_n < y] \\ &= P[(X_1 < x, Y_1 < y), (X_2 < x, Y_2 < y), \dots, (X_n < x, Y_n < y)] \\ &= P[(X_1 < x, Y_1 < y)]P[(X_2 < x, Y_2 < y)] \dots P[(X_n < x, Y_n < y)] \\ &= F_{X,Y}^n(x, y). \end{aligned}$$

As marginais são claramente F_X^n e F_Y^n .

Seja C_n a cópula definida por $F_{X,Y}^n(x, y)$ e as marginais F_X^n e F_Y^n , isto é,

$$\begin{aligned} C_n(u, v) &= F_{X,Y}^n \left((F_X^n)^{-1}(u), (F_Y^n)^{-1}(v) \right) \\ &= \left(F_{X,Y} \left((F_X^n)^{-1}(u), (F_Y^n)^{-1}(v) \right) \right)^n. \end{aligned}$$

$$C_n(u, v) = \left(F_{X,Y} \left(F_X^{-1}(u^{1/n}), F_Y^{-1}(v^{1/n}) \right) \right)^n = \left(C_F(u^{1/n}, v^{1/n}) \right)^n.$$

Como $F_X^n \left((F_X^n)^{-1}(u) \right) = u$, então:

$$\begin{aligned} \left[F_X \left((F_X^n)^{-1}(u) \right) \right]^n &= u \Rightarrow F_X \left((F_X^n)^{-1}(u) \right) = u^{1/n} \\ &\Rightarrow (F_X^n)^{-1}(u) = F_X^{-1}(u^{1/n}). \end{aligned}$$

Portanto,

$$\begin{aligned} C_n(u, v) &= \left(F_{X,Y} \left(F_X^{-1}(u^{1/n}), F_Y^{-1}(v^{1/n}) \right) \right)^n \\ &= \left(F_{X,Y} \left(F_X^{-1}(u^{1/n}), F_Y^{-1}(v^{1/n}) \right) \right)^n = \left(C_F(u^{1/n}, v^{1/n}) \right)^n. \end{aligned}$$

em que $C_F(u, v) = F_{X,Y} \left(F_X^{-1}(u), F_Y^{-1}(v) \right)$ é a cópula definida por $F_{X,Y}$ e marginais F_X e F_Y .

■

Definição: Uma cópula C é dita cópula de valor extremo se existe uma cópula C_F tal que $\lim_{n \rightarrow \infty} C_F(u^{1/n}, v^{1/n})^n = C(u, v)$ para todo $(u, v) \in [0, 1]^2$.

A cópula C_F é dita estar no domínio de atração de C . A definição de cópulas de valor extremo como limite de cópulas obtidas a partir de estatísticas do máximo geralmente leva a dificuldades técnicas para se constatar que uma determinada cópula é ou não de valor extremo.

Felizmente, existe uma propriedade algébrica que caracteriza as cópulas de valor extremo que é de mais fácil constatação.

Definição: Uma cópula C é maximalmente estável (*max-stable*) se satisfaz $C(u, v) = (C(u^{1/m}, v^{1/m}))^m$, para todo $u, v \in [0, 1]$ e para todo m inteiro maior ou igual a 1.

Exemplo: Tome a cópula de Gumbel-Hougaard, dada por

$$\begin{aligned} C(u, v) &= \exp\left\{-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{1/\theta}\right\}. \\ C(u^{1/m}, v^{1/m}) &= \exp\left\{-\left(-\log(u^{1/m})^\theta + (-\log(v^{1/m}))^\theta\right)^{1/\theta}\right\}. \\ &= \exp\left\{-\left(-\frac{1}{m}\log(u)^\theta + \left(-\frac{1}{m}\log(v)\right)^\theta\right)^{1/\theta}\right\}. \\ &= \exp\left\{\left(-\frac{1}{m}\right)^{1/\theta} \left((-\log(u))^\theta + (-\log(v))^\theta\right)\right\}. \\ &= \left[\exp\left\{(-\log(u))^\theta + (-\log(v))^\theta\right\}\right]^{1/m}. \end{aligned}$$

Portanto, $(C(u^{1/m}, v^{1/m}))^m = C(u, v)$ e segue que C é maximalmente estável.

Teorema: Uma cópula é de valor extremo se, e somente se, é maximalmente estável.

Prova: Se C é uma cópula de valor extremo, utilizando a notação anterior, existe uma cópula C_F tal que

$$\left(C_F(u^{1/m}, v^{1/m})\right)^m \xrightarrow{m \rightarrow \infty} C(u, v). \text{ Logo,}$$

$$\begin{aligned} \left(C(u^{1/n}, v^{1/n})\right)^n &= \left(\lim_{m \rightarrow \infty} C_F\left((u^{1/n})^{1/m}, (v^{1/n})^{1/m}\right)\right)^n \\ &= \left(\lim_{m \rightarrow \infty} C_F\left(u^{1/nm}, v^{1/nm}\right)^m\right)^n \\ &= \lim_{m \rightarrow \infty} \left(C_F\left(u^{1/nm}, v^{1/nm}\right)^{nm}\right) \\ &= \lim_{nm \rightarrow \infty} \left(C_F\left(u^{1/nm}, v^{1/nm}\right)^{nm}\right) \\ &= C(u, v). \end{aligned}$$

Suponha agora que C seja maximalmente estável. Se F_X e F_Y são duas acumuladas, a cópula C define uma conjunta admissível $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$.

Dessa forma, $C = C_{F_{X,Y}}$ e

$$\begin{aligned} \lim_{m \rightarrow \infty} \left(C_{F_{X,Y}}\left(u^{1/m}, v^{1/m}\right)^m\right) &= \lim_{m \rightarrow \infty} \left(C\left(u^{1/m}, v^{1/m}\right)^m\right) \\ &= \lim_{m \rightarrow \infty} \left(C(u, v)\right) = C(u, v), \text{ e, portanto, } C \text{ é a cópula de valor extremo.} \end{aligned}$$

■

O resultado central na teoria das cópulas de valor extremo é o resultado de Pickands que obtém parametrizações para todas estas cópulas. A construção é bastante técnica e não será apresentada aqui, mas pode ser encontrada em Gudendorf e Segers (2010).

Teorema de Pickands (1981): Uma cópula C é dita uma cópula de valor extremo se, e somente se, admite uma representação da forma :

$$C(y^{1-t}, y^t) = y^{A(t)}, \quad \forall (y, t) \in [0, 1]^2,$$

em que $A(\cdot)$ é uma função convexa, com domínio em $[0, 1]$, imagem em $[1/2, 1]$, e que satisfaz $\max\{s, 1-s\} \leq A(s) \leq 1$. A função $A(\cdot)$ é denominada função de dependência de Pickands.

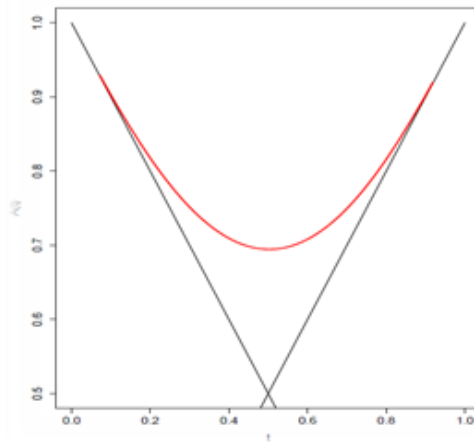


Figura 10. Função de dependência de Pickands para a cópula Gumbel-Hougaard com $\tau = 0.50$.

Outra parametrização na caracterização de cópula de valor extremo pode ser obtida tomando-se $u = y^{1-t}$ e $v = y^t$. Dessa forma, o teorema pode ser enunciado da seguinte forma, tomando-se $u = y^{1-t}$ e $v = y^t$.

Teorema de Pickands (1981): Uma cópula bivariada C é de valor extremo se, e somente se, $C(u, v) = (uv)^{A(\ln(v)/\ln(uv))}$, $(u, v) \in (0, 1]^2 - \{(1, 1)\}$ em que $A(\cdot)$ é uma função convexa, com domínio em $[0, 1]$, imagem em $[1/2, 1]$, e que satisfaz $\max\{t, 1-t\} \leq A(t) \leq 1$.

De fato, temos, nessa nova parametrização:

$$u = \frac{y}{y^t} \quad \Rightarrow \quad y = uv .$$

$$\left. \begin{array}{l} \ln(u) = (1-t)\ln(y) \\ \ln(v) = t\ln(y) \end{array} \right\} \Rightarrow \frac{\ln(u)}{\ln(v)} = \frac{1-t}{t} .$$

$$\Rightarrow t \ln(u) = (1-t) \ln(v) .$$

$$\Rightarrow t \ln(uv) = \ln(v) .$$

$$\Rightarrow t = \frac{\ln(v)}{\ln(uv)} .$$

Sendo assim, $C(y^{1-t}, y^t) = y^{A(t)}$.

$$\Rightarrow C(u, v) = \exp\left\{\ln\left[y^{A(t)}\right]\right\} = \exp\{\ln[y]A(t)\} .$$

$$\Rightarrow C(u, v) = \exp\left\{\ln[uv]A\left(\frac{\ln[v]}{\ln[uv]}\right)\right\} .$$

Este resultado permite reescrever o teorema de Pickands como:

Uma cópula $C(u, v)$ é uma cópula de valor extremo se a transformação

$T_C : [0, 1]^2 \rightarrow [0, 1] \times [0, +\infty)$, definida, para todo $(u, v) \in [0, 1]^2$, como

$(u, v) \rightarrow (t, z) = \left(\frac{\ln(v)}{\ln(uv)}, \frac{\ln(C(u, v))}{\ln(uv)} \right)$, tem como imagem em uma curva convexa, limitada superiormente por $z = 1$ e inferiormente por $z = \max(t, 1-t)$.

■

Exemplo: Uma questão natural é exibir cópulas de valor extremo cuja função de dependência de Pickands assume os valores dos limites superior, $A(t) \equiv 1$, e inferior, $A(t) = \max(t, 1-t)$.

Seja a cópula de valor extremo de Gumbel:

$$C(u, v) = \exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\}, \quad \theta \geq 1.$$

A transformação $(0, 1)^2 \rightarrow (0, 1) \times (0, \infty)$,

$$(u, v) \rightarrow (t, z) = \left(\frac{\ln u}{\ln uv}, \frac{\ln C(u, v)}{\ln uv} \right), \text{ resulta, se a}$$

cópula é a Gumbel, em:

$$z = \frac{\ln \left[\exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\} \right]}{\ln uv}.$$

$$= \frac{1}{\ln uv} \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\}.$$

$$\text{Se } \theta = 1, z = \frac{1}{\ln uv} \left\{ - \left[(-\ln u) + (-\ln v) \right] \right\} = 1.$$

Observe que, fixado um valor de v , $\ln v = k$, em que $k \in (-\infty, 0)$ e

$$t = \frac{\ln u}{\ln uv} = \frac{\ln u}{k + \ln u}.$$

$$\text{Se } u \rightarrow 0 \quad \Rightarrow \quad \ln u \rightarrow -\infty \quad \Rightarrow \quad t \rightarrow 1.$$

$$\text{Se } u \rightarrow 1 \quad \Rightarrow \quad \ln u \rightarrow 0 \quad \Rightarrow \quad t \rightarrow 0.$$

Significa que, se $\theta = 1$, para todo u , o valor de t percorre o intervalo $(0,1)$ e $z = 1 = A(t)$, ou seja, a função de Pickands degenera no seu limite superior.

$$\begin{aligned} \text{Se } \theta \rightarrow +\infty, \quad \lim_{\theta \rightarrow \infty} z &= \lim_{\theta \rightarrow \infty} \frac{1}{\ln uv} \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\}. \\ &= \lim_{\theta \rightarrow \infty} \frac{1}{\ln uv} \left\{ -(-\ln u) \left[1 + \left(\frac{\ln v}{\ln u} \right)^\theta \right]^{1/\theta} \right\}. \\ &= \lim_{\theta \rightarrow \infty} \frac{\ln u}{\ln uv} \left[1 + \left(\frac{\ln v}{\ln u} \right)^\theta \right]^{1/\theta}. \\ &= \begin{cases} \frac{\ln u}{\ln uv} & \text{caso } \frac{\ln v}{\ln u} < 1 \\ \frac{\ln v}{\ln uv} & \text{caso } \frac{\ln v}{\ln u} > 1 \end{cases}. \\ &= \begin{cases} t & \text{caso } \frac{\ln v}{\ln u} < 1 \\ 1-t & \text{caso } \frac{\ln v}{\ln u} > 1 \end{cases}. \end{aligned}$$

Observe agora que:

$$\frac{\ln v}{\ln u} > 1 \Leftrightarrow \ln v > \ln u \Leftrightarrow \ln v + \ln u > 2\ln u \Leftrightarrow \frac{\ln u}{\ln uv} < \frac{1}{2} \Leftrightarrow t < \frac{1}{2}.$$

$$\frac{\ln v}{\ln u} < 1 \Leftrightarrow \ln v < \ln u \Leftrightarrow \ln v + \ln u < 2 \ln u \Leftrightarrow \frac{\ln u}{\ln uv} > \frac{1}{2} \Leftrightarrow t > \frac{1}{2}.$$

Significa que, se $\theta \rightarrow +\infty$, a função de Pickands $z = A(t)$ degenera para seu limite inferior $\max(t, 1-t)$.

Proposição: As cópulas de valor extremo apresentam dependência positiva em quadrante.

Prova: Se A é uma parametrização de Pickands da cópula de valor extremo C então $C(u, v) = (uv)^{A(\log(v)/\log(uv))} \geq (uv)^{1/2} \geq uv$, uma vez que $\frac{1}{2} < A < 1$. Tome $u = F_X^{-1}(x)$ e $v = F_Y^{-1}(y)$, assim:

$$\begin{aligned} P[X > x, Y > y] &= \bar{C}(u, v) \geq 1 - u - v + C(u, v) = 1 - u - v + uv. \\ &\Rightarrow P[X > x, Y > y] \geq P[X > x]P[Y > y]. \end{aligned}$$

■

Proposição: As cópulas de valor extremo apresentam dependência nas caudas.

Prova: Uma cópula de valor extremo pode ser parametrizada por uma função de Pickands, $C(u, v) = \exp\left[\ln(uv)A\left(\frac{\ln u}{\ln(uv)}\right)\right]$. Tem-se então:

$$\begin{aligned}
\lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1-u} &= \lim_{u \rightarrow 1^-} \left[\frac{1-u-u+C(u, u)}{1-u} \right]. \\
&= \lim_{u \rightarrow 1^-} \left[\frac{1-u-u+\exp \left[\ln(u^2) A \left(\frac{\ln u}{\ln(u^2)} \right) \right]}{1-u} \right]. \\
&= \lim_{u \rightarrow 1^-} \left[\frac{1-2u+\exp \left[\ln(u^2) A \left(\frac{1}{2} \right) \right]}{1-u} \right]. \\
&= \lim_{u \rightarrow 1^-} \left[\frac{1-2u+u^{2k}}{1-u} \right] = \lim_{u \rightarrow 1^-} \left[\frac{-2+2ku^{2k-1}}{-1} \right] = 2-2k,
\end{aligned}$$

em que $K = A \left(\frac{1}{2} \right)$.

Como $\frac{1}{2} < k \leq 1$, então $0 < \lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1-u} \leq 1$, e, portanto,

apresenta dependência na cauda superior. ■

2.3.9 Polinômios de Bernstein

Considere os monômios obtidos da expansão do binômio de Newton $(x+y)^n$ com $x+y=1$, isto é, $b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$ em que $k=0,1,\dots,n$.

Proposição: $b_{n,k}(x)$ são polinômios positivos para $0 \leq x \leq 1$, com um único máximo dado por $x = \frac{k}{n}$, se $0 < k < n$.

Prova: Para encontrar o ponto máximo basta calcular a primeira derivada do polinômio de Bernstein e igualar a 0.

$$\binom{n}{k} k x^{k-1} (1-x)^{n-k} - \binom{n}{k} x^k (n-k) (1-x)^{n-k-1} = 0.$$

$$\Rightarrow \binom{n}{k} x^k (1-x)^{n-k} \left[\frac{k}{x} - (n-k) \frac{1}{(1-x)} \right] = 0.$$

$$\Rightarrow \left[\frac{k}{x} - (n-k) \frac{1}{(1-x)} \right] = 0 \Rightarrow \frac{(1-x)k - x(n-k)}{x(1-x)} = 0.$$

$$\Rightarrow k - x - nx + kx = 0 \quad x = \frac{k}{n}.$$

■

Os gráficos dos polinômios $b_{n,k}$ são da forma:

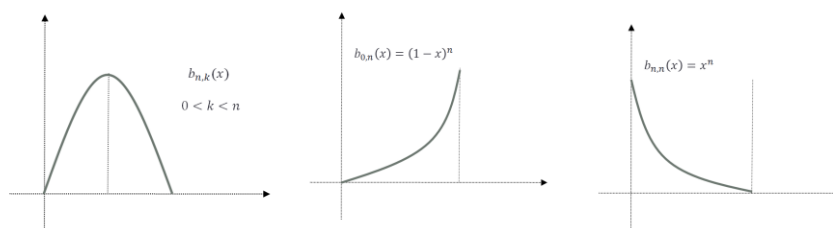


Figura 11. Gráfico dos polinômios $b_{n,k}$.

Proposição: Os polinômios $b_{n,k}$ possuem também a interessante

propriedade da integral ser constante $\int_0^1 b_{n,k}(x) dx = \frac{1}{n+1}$.

Prova: O cálculo da integral é dado por:

$$\begin{aligned} \int_0^1 b_{n,k}(x) dx &= \int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \binom{n}{k} \int_0^1 x^k (1-x)^{n-k} dx. \\ &= \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!}. \\ &= \frac{n!}{(n-k)!k!} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{(n+1)}, \text{ em que } \Gamma \text{ é a função Gama.} \end{aligned}$$

■

Definição: Os polinômios de Bernstein de grau menor ou igual a n são obtidos como combinação linear dos monômios $b_{n,k}$, isto é,

$$B_n(x) = \sum_{k=0}^n \beta_k b_{n,k}(x) = \sum_{k=0}^n \beta_k \binom{n}{k} x^k (1-x)^{n-k}.$$

Neste trabalho será necessário o resultado do cálculo da primeira e da segunda derivada dos polinômios de Bernstein. A primeira derivada é dada por:

$$\begin{aligned} B_n'(x) &= -n\beta_0(1-x)^{n-1} \\ &+ \sum_{k=1}^{n-1} \beta_k \cdot \binom{n}{k} \left[k x^{k-1} (1-x)^{n-k} - (n-k)x^k (1-x)^{n-1-k} \right] \\ &+ n\beta_n x^{n-1}. \end{aligned}$$

$$\begin{aligned}
&= -n\beta_0(1-x)^{n-1} + \sum_{k=1}^{n-1} \beta_k \frac{n}{k} \binom{n-1}{k-1} \left[k x^{k-1} (1-x)^{n-k} \right] \\
&\quad - \sum_{k=1}^{n-1} \beta_k \frac{n}{n-k} \binom{n-1}{n-k-1} (n-k) x^k (1-x)^{n-1-k} + n\beta_n x^{n-1}. \\
&= -n\beta_0(1-x)^{n-1} + n \sum_{k=1}^{n-1} \beta_k \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} \\
&\quad - n \sum_{k=1}^{n-1} \beta_k \binom{n-1}{k} x^k (1-x)^{n-1-k} + n\beta_n x^{n-1}. \\
&= -n\beta_0(1-x)^{n-1} + n\beta_n x^{n-1} + n\beta_1(1-x)^{n-1} + n \sum_{k=2}^{n-1} \beta_k \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} \\
&\quad - n \sum_{k=1}^{n-2} \beta_k \binom{n-1}{k} x^k (1-x)^{n-1-k} - n\beta_{n-1} x^{n-1}. \\
&= -n\beta_0(1-x)^{n-1} + n\beta_n x^{n-1} + n\beta_1(1-x)^{n-1} - n\beta_{n-1} x^{n-1} \\
&\quad + n \sum_{j=1}^{n-2} \beta_{j+1} \binom{n-1}{j} x^j (1-x)^{n-1-j} - n \sum_{k=1}^{n-2} \beta_k \binom{n-1}{k} x^k (1-x)^{n-1-k}. \\
&= n(\beta_n - \beta_{n-1}) x^{n-1} + n(\beta_1 - \beta_0)(1-x)^{n-1} \\
&\quad + n \sum_{k=1}^{n-2} (\beta_{k+1} - \beta_k) \binom{n-1}{k} x^k (1-x)^{n-1-k}.
\end{aligned}$$

Para o cálculo da segunda derivada faremos:

$$B_n''(x) = n \left(\left[\sum_{k=0}^{n-1} (\beta_{k+1} - \beta_k) \cdot b_k(x, n-1) \right] \right)'.$$

Usando $n-1 = M$ e $\beta_{k+1} - \beta_k = \alpha_k$, teremos:

$$\begin{aligned}
B_n''(x) &= n \left(\left[\sum_{k=0}^M \alpha_k \cdot b_k(x, M) \right] \right)' = n B_M'(x). \\
&= n \left(M \sum_{k=0}^{M-1} (\alpha_{k+1} - \alpha_k) \cdot b_k(x, M-1) \right). \\
&= n(n-1) \sum_{k=0}^{n-2} ((\beta_{k+2} - \beta_{k+1}) - (\beta_{k+1} - \beta_k)) \cdot b_k(x, M-1). \\
&= n(n-1) \sum_{k=0}^{n-2} (\beta_{k+2} - 2\beta_{k+1} + \beta_k) \cdot b_k(x, n-2). \\
&= n(n-1) \sum_{k=0}^{n-2} (\beta_{k+2} - 2\beta_{k+1} + \beta_k) \cdot \binom{n-2}{k} x^k (1-x)^{n-2-k}. \\
&= n(n-1) \left[(\beta_2 - 2\beta_1 + \beta_0)(1-x)^{n-2} + (\beta_n - 2\beta_{n-1} + \beta_{n-2})x^{n-1} \right] \\
&\quad + n(n-1) \sum_{k=1}^{n-3} (\beta_{k+2} - 2\beta_{k+1} + \beta_k) \cdot \binom{n-2}{k} x^k (1-x)^{n-2-k}.
\end{aligned}$$

Note que $B_n''(x) \geq 0$ se, e somente se, $\beta_k - 2\beta_{k+1} + \beta_{k+2} \geq 0$ para $k = 0, 1, \dots, n-2$.

Os polinômios de Bernstein permitem aproximações de funções contínuas da forma:

$$f(x) = \lim_{n \rightarrow \infty} \sum_{k=0}^n f\left(\frac{k}{n}\right) b_{n,k}(x).$$

Esta aproximação no intervalo $[0, 1]$ é uniforme, sendo um exemplo construtivo para o teorema de aproximação de Stone-Weierstrass (Lima, 2007).

2.3.10 Regressão não paramétrica com restrição de forma utilizando polinômios de Bernstein

Em razão de suas propriedades, os polinômios de Bernstein são adequados à regressão sujeita a minimização de soma de quadrados. O problema em sua forma mais simples é, dados os pontos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, obter um polinômio de Bernstein

$B_n(x) = \sum_{k=0}^n \beta_k b_{k,n}(x)$ que minimiza a soma de quadrados

$$\sum_{i=1}^N (B_n(x_i) - y_i)^2 = \sum_{i=1}^N \left(\sum_{k=0}^n \beta_k b_{k,n}(x_i) - y_i \right)^2.$$

No caso desse trabalho, a minimização da soma de quadrados está sujeita a um conjunto de restrições lineares.

A solução é obtida facilmente derivando-se em relação aos coeficientes β_k , obtendo-se um sistema linear:

$$\frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^N \left(\sum_{k=0}^n \beta_k b_{k,n}(x_i) - y_i \right)^2 \right] = \sum_{i=1}^N 2 \left(\sum_{k=0}^n \beta_k b_{k,n}(x_i) - y_i \right) b_{j,n}(x_i) = 0$$

$$\Rightarrow \sum_{k=0}^n \left[\beta_k \sum_{i=1}^N b_{k,n} b_{j,n}(x_i) - \sum_{i=1}^N y_i b_{j,n}(x_i) \right] = 0, \quad j = 1, \dots, n.$$

As propriedades dos polinômios de Bernstein permitem resolver problemas bem mais complexos do que acima descrito. Para simplificar a notação e para explicitar as restrições, será usada a notação para um vetor $v' = (v_1, \dots, v_N)$. $v \geq 0$ significa que $v_i \geq 0$, $i = 1, \dots, N$. Se $A_{i,N}$ é uma

matrix $l \times N$, então $A_{l,N}v \geq 0$ significa que as l entradas $A_{l,N}v$ são positivas.

Por exemplo, considere o problema de se obter o polinômio de Bernstein que minimize a soma de quadrados $\sum_{i=1}^N (B_n(x_i) - y_i)^2$ sujeita a restrição que este polinômio seja monótono não decrescente.

Os polinômios de Bernstein monótonos não decrescentes podem ser caracterizados por um sistema de desigualdades lineares. De fato, como $B'_n(x) = n \sum_{k=0}^n (\beta_{k+1} - \beta_k) b_{k,n-1}(x)$, segue que $B'_n(x) \geq 0$ para $x \in [0,1]$ se $\beta_{k+1} - \beta_k \geq 0$, $x = 0, \dots, n-1$, e, portanto, $\beta_0 \leq \beta_1 \leq \dots \leq \beta_n$.

Estas condições, de forma condensada, podem ser expressas por:

$$\begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \beta_1 - \beta_0 \\ \vdots \\ \beta_n - \beta_{n-1} \end{pmatrix} \geq 0, \text{ isto é } A\beta \geq 0.$$

Uma restrição mais complexa é supor que o polinômio de Bernstein satisfaz a restrição de ser convexo. A condição de convexidade é satisfeita se $B''_n(x) \geq 0$. Como

$$B''_n(x) = n(n-1) \sum_{k=0}^{n-2} (\beta_{k+2} - 2\beta_{k+1} + \beta_k) b_{k,n-2}(x), \text{ a restrição é}$$

satisfeita se $\beta_{k+2} - 2\beta_{k+1} + \beta_k \geq 0$, que pode ser expressa por:

$$\begin{pmatrix} 1 & -2 & 1 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & & & \ddots & & \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \geq 0.$$

Problemas envolvendo os polinômios de Bernstein são problemas de programação quadrática com restrições lineares e podem ser resolvidos com algoritmos eficientes já implementados em várias linguagens. Uma restrição a mais necessária é a da limitação dos coeficientes:

$$\sum_{k=0}^n |\beta_k| < M_n.$$

2.3.11 Ajuste de Cópulas

O ajuste de uma família uniparamétrica de cópulas depende do parâmetro que a caracteriza. Anteriormente, foi visto que é possível estabelecer uma relação entre o TAU de Kendall, ou o RHO de Spearman, e o parâmetro das cópulas, de forma que, o parâmetro de uma cópula pode ser escrito como uma função que dependa de TAU de Kendall ou uma função que dependa de RHO de Spearman.

Para ajustar uma determinada família de cópulas a um conjunto de dados, calcula-se o valor do TAU de Kendall, ou do RHO de Spearman, desses dados e em seguida, o parâmetro dessa família é encontrado invertendo-se a função que associa o TAU de Kendall, ou RHO de Spearman, ao valor do parâmetro de cada uma dessas cópulas.

Na Tabela 3 são apresentadas as funções do parâmetro de Clayton, Gumbel e Normal. Devido à complexidade da função do TAU de Kendall para a cópula de Frank, o valor do parâmetro será calculado apenas numericamente.

Tabela 3 Parâmetro das cópulas em relação ao TAU de Kendall

<i>Família</i>	<i>Parâmetro</i>
<i>Clayton</i>	$\theta = \frac{2\tau}{(1-\tau)}$
<i>Gumbel</i>	$\theta = \frac{1}{(1-\tau)}$
<i>Normal</i>	$\rho = \sin\left(\frac{\tau\pi}{2}\right)$

2.3.12 Qualidade de ajuste de Cópulas

Eleger uma família de cópulas para ajustar um determinado conjunto de dados é uma tarefa importante e bem complexa. Porém, ainda não existe um método dito mais adequado para tal. Nos últimos anos, vários métodos têm sido propostos observando-se os diferentes tipos de características dos dados.

Genest e Rivest (1993) propuseram um procedimento não paramétrico para ajustar cópulas Arquimedianas. Ané e Kharoubi (2003) apresentam um método de seleção baseado em minimizar a distância entre uma cópula paramétrica e outra cópula empírica. Berg e Bakken (2005) propuseram uma abordagem baseada na transformação de Rosenblatt aplicada a pseudo-amstras.

Chen e Fan (2005) recorrem a um teste da razão de uma pseudo-verossimilhança com a finalidade de selecionar o melhor modelo para uma família de cópula. Sua abordagem, inspirada por uma adaptação semiparamétrica do critério de informação de Akaike, torna possível a escolha do modelo mais parcimonioso.

Enquanto Huard et al. (2006) propõem um método Bayesiano que se baseia no TAU de Kendall para selecionar uma cópula mais provável dentre um conjunto de nove possíveis famílias de cópulas. Quessy (2009) sugerem uma abordagem de seleção para cópulas bivariadas baseado na dependência de RHO de Spearman.

Genest e Rémillard (2008) mostram dois tipos de testes. Um compara uma cópula empírica por meio da estatística de Cramér-von Mises, com uma cópula paramétrica, o outro mede a distância entre uma cópula paramétrica e uma estimativa não paramétrica.

Genest et al. (2009) apresentam uma revisão de testes propostos na literatura e defendem o uso de “blanket tests”, cuja implementação não requer categorização arbitrária dos dados, nem qualquer escolha estratégica de parâmetro para a cópula.

Kojadinovic et al. (2011) propõem um procedimento de teste para grandes amostras baseado em um multiplicador (teorema do limite central) que é válido quando baseado assintoticamente em estimadores calculados através do método dos momentos. A utilização da abordagem do multiplicador ao invés do bootstrap paramétrico reduz consideravelmente o tempo de processamento do teste.

O software livre R possui um pacote denominado *copula-package* {*copula*}. Esse pacote possui a função *fitCopula* {*copula*} que ajusta um modelo a um conjunto de dados multivariados, para diversas famílias de cópulas. E ainda, possui a função *gofCopula* {*copula*} que

verifica a qualidade de ajuste de dois testes para cópulas. O primeiro teste é baseado em Genest e Rémillard (2008) e Genest et. al. (2009). E o segundo teste é baseado em Kojadinovic et. al. (2011) e Kojadinovic e Yan (2011).

Essa abordagem é considerada ainda recente e muitos estudos a respeito ainda vêm sendo propostos, visto que a maioria dos métodos apresentados na literatura possuem algum tipo de crítica com relação à qualidade de ajuste. As principais críticas e obstáculos observados estão relacionados ao exaustivo tempo computacional gasto à medida que o tamanho da amostra aumenta, e ao baixo poder dos testes para amostras de tamanho pequeno ($n = 100$).

3 CONSIDERAÇÕES GERAIS

Este trabalho consiste em dois tópicos. O primeiro propõe dois novos testes de verificação de qualidade de ajuste para cópulas bivariadas, baseados nas marginais nas diagonais principal e secundária, sendo que um utiliza estatística qui-quadrado e o outro, intervalos de confiança para assimetria e curtose. O segundo desenvolve um novo método para verificar se um determinado conjunto de dados bivariados possui dependência de valor extremo, através da estimação não-paramétrica da função de Pickands utilizando polinômios de Bernstein convexos.

O comportamento das famílias de cópulas é semelhante, portanto, já se espera que dificuldades sejam encontradas em qualquer tipo de teste de qualidade de ajuste. E, realmente, em todos os casos analisados, os dois testes propostos apresentaram um bom controle do erro tipo I, porém, ainda há uma falha quando se trata do controle da taxa de erro tipo II.

O teste que utiliza intervalo de confiança apresenta um melhor desempenho no controle da taxa de erro tipo II quando se utilizam os valores intermediários de TAU de Kendall ($\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$). Uma provável razão para isso é que, para τ próximo de zero (dados se aproximam da independência) ou τ próximo de 1 (dados se aproximam da dependência positiva perfeita) as cópulas se aproximam, respectivamente, da cópula da independência $C_i(\mathbf{u}, \mathbf{v}) = \mathbf{uv}$ ou da cópula do mínimo $C_1(\mathbf{u}, \mathbf{v}) = \min\{\mathbf{u}, \mathbf{v}\}$. E, portanto os casos as cópulas têm comportamentos parecidos, o que dificulta a sua distinção.

Essa tese sugere para que muitos estudos ainda possam ser realizados a fim de que o desempenho dos métodos propostos se torne mais eficientes.

REFERÊNCIAS

- ANÉ, T.; KHAROUBI, C. Dependence Structure and Risk Measure. **The journal of business**, v. 76, n. 3, p. 411-438, 2003.
- CEBRIAN, A., DENUIT, M., LAMBERT, P. Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database. **Belgian Actuarial Journal**, v. 3, n. 1, p. 33-41, 2003.
- CHEN, X.; FAN, Y. Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. **Canadian Journal of Statistics**, v. 33, n. 3, p. 389-414, 2005.
- CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. **Copula methods in finance**. John Wiley & Sons, 2004.
- COLES, S.; HEFFERNAN, J.; TAWN, J. Dependence measures for extreme value analyses. **Extremes**, v. 2, n. 4, p. 339-365, 1999.
- GALAMBOS, J. **The Asymptotic Theory of Extreme Order Statistics**. Krieger. **New York**, 1987.
- GENEST, C.; RIVEST, L-P. Statistical inference procedures for bivariate Archimedean copulas. **Journal of the American statistical Association**, v. 88, n. 423, p. 1034-1043, 1993.
- GENEST, C.; RÉMILLARD, B. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In: **Annales de l'IHP Probabilités et statistiques**. 2008. p. 1096-1127.
- GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. **Insurance: Mathematics and economics**, v. 44, n. 2, p. 199-213, 2009.
- GUDENDORF, Gordon; SEGERS, Johan. Extreme-value copulas. In: **Copula theory and its applications**. Springer Berlin Heidelberg, 2010. p. 127-145.

- HUARD, D.; EVIN, G.; FAVRE, A. C. Bayesian copula selection. **Computational Statistics & Data Analysis**, v. 51, n. 2, p. 809-822, 2006.
- JOE, H.. **Multivariate models and multivariate dependence concepts**. CRC Press, 1997.
- KENDALL, M. G. Rank Correlation Methods. Charles Griffin and Company. 1962.
- KOJADINOVIC, I.; YAN, J. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. **Statistics and Computing**, v. 21, n. 1, p. 17-30, 2011.
- KOJADINOVIC, I.; YAN, J; HOLMES, M. Fast large-sample goodness-of-fit tests for copulas. **Statistica Sinica**, v. 21, n. 2, p. 841, 2011.
- KOJADINOVIC, I.; SEGERS, J.; YAN, J. Large-sample tests of extreme-value dependence for multivariate copulas. **Canadian Journal of Statistics**, v. 39, n. 4, p. 703-720, 2011.
- LIMA, E. L.. **Espaços Métricos**. Rio de Janeiro: IMPA, 2007.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics**. New York: McGraw-Hill, 1974. 564 p.
- NELSEN, R. B. **An introduction to copulas**. Springer Science & Business Media, 2013.
- PICKANDS, J. Multivariate extreme value distributions. In: **Proceedings 43rd Session International Statistical Institute**. 1981. p. 859-878.
- QUESSY, J-F. Theoretical efficiency comparisons of independence tests based on multivariate versions of Spearman's rho. **Metrika**, v. 70, n. 3, p. 315-338, 2009.
- RESNICK, Sidney I. Extreme Values. **Regular Variation, and Point Processes**, New York et al, 1987.
- SCAILLET, O. A Kolmogorov–Smirnov type test for positive quadrant dependence. *Can. J. Statist.* 33, 415–427. (2005).

SKLAR, M. Fonctions de répartition à n dimensions et leurs marges.
Université Paris 8, 1959.

CAPÍTULO 2 QUALIDADE DE AJUSTE DE CÓPULAS VIA MARGINAIS NAS DIAGONAIS

RESUMO

É notório o grande número de estudos recentes relacionados à qualidade de ajuste para cópulas e inegável a sua relevância em diversas áreas, principalmente na economia. Eleger uma família de cópulas para ajustar um determinado conjunto de dados é uma tarefa importante e bem complexa. Porém, ainda não existe um método dito mais adequado para tal. Nos últimos anos, vários métodos têm sido propostos observando-se os diferentes tipos de características dos dados. O objetivo principal deste trabalho é propor dois novos testes para verificar a qualidade de ajuste de cópulas para dados bivariados via marginais nas diagonais principal e secundária. A fim de verificar a adequabilidade dos testes, foram utilizadas as seguintes famílias de cópulas: Clayton, Gumbel, Normal e Frank. Os cálculos utilizados nos dois testes foram feitos com o auxílio do software livre R. No primeiro teste, são abordadas as marginais nas diagonais, principal e secundária, das cópulas em estudo e em seguida realizou-se um teste qui quadrado para distribuições de frequências. O segundo teste proposto verificou se os coeficientes de assimetria e curtose de amostras das famílias de cópulas de interesse pertencem aos intervalos de confiança construídos na diagonal principal e/ou na diagonal secundária, via Monte Carlo, para tal. A conclusão foi feita através da verificação do controle das taxas de erros tipo I e tipo II.

Palavras-chave: qualidade de ajuste, cópulas, marginais nas diagonais, TAU de Kendall.

ABSTRACT

It is clear the large number of recent studies about the goodness of fit for copula and undeniable its relevance in various fields, especially in economy. To elect a family of copula to fit a given data set is an important and very complex task and there is not a known method best suited for this purpose. In recent years, various methods have been proposed observing the different characteristics of the data. The main aim of this paper is to propose two new tests to verify goodness of fit bivariate copulas data via marginal in primary and secondary diagonal. In order to verify the suitability of the tests, the following families of copulas are used: Clayton, Gumbel, Normal and Frank. The calculations use are performed with the help of the free software R. In the first test, the marginal in the diagonal, principal and secondary of the copula in study is approached and then chi-square test is performed. The second proposed test verifies if the coefficients of skewness and kurtosis for samples of copulas' families belongs to the confidence interval constructed on the main diagonal and/or secondary diagonal, via Monte Carlo to such. The conclusion is made by checking the control of the rates of errors type I and type II.

Keywords: goodness of fit, copula, marginal in the diagonal, Kendall's TAU.

1 INTRODUÇÃO

É notório o grande número de estudos recentes relacionados à qualidade de ajuste para cópulas e inegável sua crescente relevância em diversas áreas, principalmente na hidrologia e na economia como pode ser visto em Genest e Favre (2007) e Nelsen (2005).

Eleger uma família de cópulas para ajustar um determinado conjunto de dados é uma tarefa importante e bem complexa. Porém, ainda não existe um método dito mais adequado para tal. Nos últimos anos, vários métodos têm sido propostos observando-se os diferentes tipos de características dos dados.

As cópulas são de grande interesse na estatística, principalmente, por se tratar de funções que permitem separar o comportamento marginal das variáveis aleatórias de sua estrutura de dependência. A construção da distribuição conjunta de variáveis por cópulas é razoável, uma vez que, não há restrições quanto às distribuições marginais das variáveis envolvidas e ainda, a cópula de uma distribuição capta diversos tipos de dependência entre as variáveis. Portanto, a cópula é uma maneira de isolar as estruturas de dependência das variáveis.

2 OBJETIVOS

O objetivo principal deste capítulo é propor um novo método de ajuste de cópulas para dados bivariados baseados nas marginais obtidas nas diagonais principal e secundária.

Os objetivos específicos do estudo foram:

- a) baseado no novo método de ajuste, propor e avaliar um novo teste de aderência para cópulas bivariadas, utilizando uma estatística com distribuição qui-quadrado.
- b) baseado no novo método de ajuste, propor e avaliar um teste de hipótese cujo critério de decisão é baseado em intervalos de confiança para assimetria e curtose.
- c) aplicar os novos testes para dados provenientes de cópulas das famílias: Clayton, Gumbel, Normal e Frank.

3 METODOLOGIA

A proposta deste capítulo é apresentar dois testes para ajuste de cópulas, todos baseados nas marginais sobre as diagonais, principal e secundária, do suporte da cópula.

A diagonal principal da cópula é denotada por w e a diagonal secundária por z . A matriz de rotação utilizada é dada por:

$$\begin{aligned} \begin{bmatrix} w \\ z \end{bmatrix} &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \\ \Rightarrow \quad u &= \frac{\sqrt{2}}{2}w - \frac{\sqrt{2}}{2}z \\ v &= \frac{\sqrt{2}}{2}w + \frac{\sqrt{2}}{2}z \end{aligned}$$

Na Figura 12 é apresentado o gráfico de rotação do suporte da cópula $C(u, v)$, em que os eixos u e v são transformados em w e z .

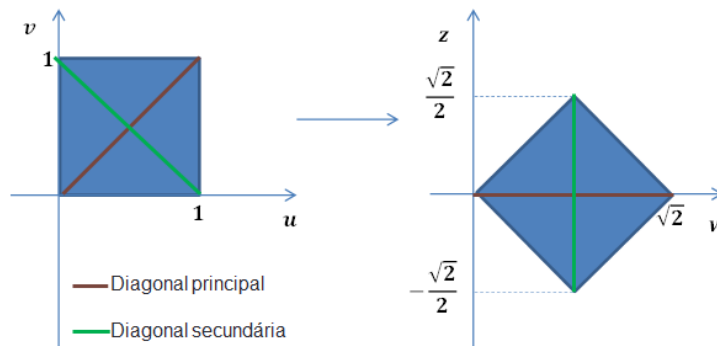


Figura 12. Rotação do suporte da cópula.

Para ilustrar essa rotação, considere a cópula da independência dada por:

$$C(u, v) = uv \Rightarrow c(u, v) = \frac{d^2}{dudv} C(u, v) = 1$$

Os cálculos das marginais $f_W(w)$, diagonal principal, e $f_Z(z)$, diagonal secundária, são apresentados em seguida:

$$\begin{aligned} f_W(w) &= \left[\int_{-w}^w dz \right] I_{\left(0, \frac{\sqrt{2}}{2}\right)}(w) + \left[\int_{-\sqrt{2}+w}^{\sqrt{2}-w} dz \right] I_{\left(\frac{\sqrt{2}}{2}, \sqrt{2}\right)}(w) \\ &= zw I_{\left(0, \frac{\sqrt{2}}{2}\right)}(w) + 2(\sqrt{2} - w) I_{\left(\frac{\sqrt{2}}{2}, \sqrt{2}\right)}(w) \\ f_Z(z) &= \left[\int_{-z}^{z+\sqrt{2}} dw \right] I_{\left(-\frac{\sqrt{2}}{2}, 0\right)}(z) + \left[\int_z^{\sqrt{2}-z} dw \right] I_{\left(0, \frac{\sqrt{2}}{2}\right)}(z) \\ &= (\sqrt{2} + 2z) I_{\left(-\frac{\sqrt{2}}{2}, 0\right)}(z) + (\sqrt{2} - 2z) I_{\left(0, \frac{\sqrt{2}}{2}\right)}(z) \end{aligned}$$

Utilizando o software livre R para rotacionar e plotar em um gráfico os dados de uma amostra tamanho $n = 1.000.000$ da cópula da independência, obtemos a imagem da Figura 13, que são exatamente os gráficos das marginais $f_W(w)$ e $f_Z(z)$, respectivamente.

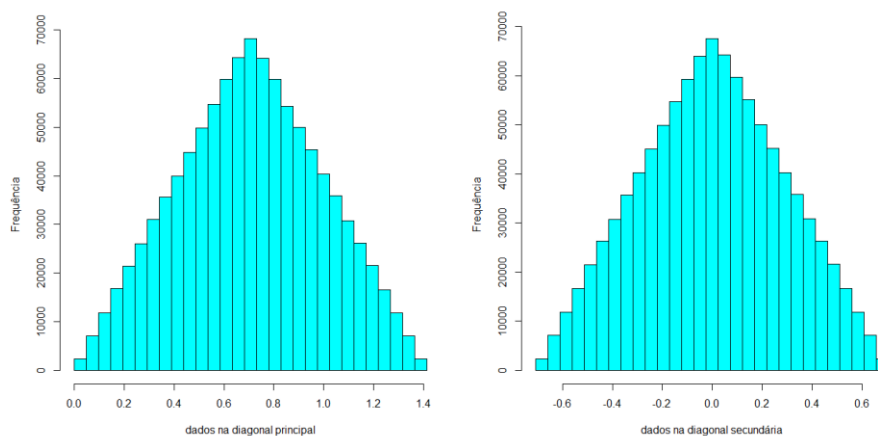


Figura 13. Gráfico dos dados de uma amostra tamanho $n = 1.000.000$ da cópula da independência plotados nas diagonais principal e secundária.

O primeiro passo para obter os dados nas marginais sobre as diagonais é a transformação dos dados pareados originais para dados pertencentes ao intervalo $[0,1] \times [0,1]$ (suporte da cópula). Considere os dados (x_i, y_i) , $i = 1, 2, \dots, n$, e seus respectivos postos (R_i, S_i) . A transformação utilizada foi $(x_i, y_i) \rightarrow \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right)$. Em seguida, a idéia é distinguir as cópulas pelas marginais nas diagonais: a principal e a secundária.

A fim de verificar a adequabilidade dos testes, são utilizadas as seguintes famílias de cópulas: Clayton, Gumbel, Normal e Frank. Os cálculos utilizados nos dois testes foram feitos com o auxílio de um pacote do software livre R denominado *copula-package* {*copula*}.

Como as cópulas se distinguem de forma mais pronunciada nas extremidades, os dados serão divididos em três regiões definidas por retas ortogonais à diagonal principal. Os dados localizados na parte central serão

descartados, ou seja, serão utilizados apenas os dados localizados nas extremidades do quadrado $[0,1] \times [0,1]$.

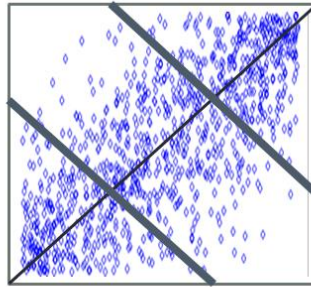


Figura 14 Divisão das três regiões definidas por retas ortogonais à diagonal

Os dois testes propostos neste trabalho foram comparados com outros dois métodos de verificação de qualidade de ajuste disponíveis no software R através do pacote *gofCopula {copula}*. O primeiro método disponível no R (PB) utiliza o bootstrap paramétrico na comparação entre a cópula empírica e uma estimativa paramétrica da cópula, sob a hipótese nula, utilizando os dados na posição usual, como pode ser visto em Genest et al. (2013) e Genest e Rémillard (2008). O segundo (MB) faz uso de um multiplicador λ apresentado em Genest et al. (2009) e Kojadinovic et al. (2011).

Por simulação computacional, ficou constatado que as marginais empíricas obtidas na diagonal secundária, possuem formas muito semelhantes e, portanto, incapazes de distinguir cópulas diferentes. Em razão desse fato, os testes propostos utilizam apenas as marginais empíricas na diagonal principal.

3.1 Teste 1: Qui-quadrado

Esse teste usa a estatística Qui-Quadrado para distribuições de frequência em relação a diagonal principal.

Passo 1) Define-se o valor de tau de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$) e o tamanho da amostra ($n = 600$).

Passo 2) Geram-se n dados para cada uma das cópulas, definidas pelo valor do tau de Kendall, nas famílias: Clayton, Gumbel, Frank, Normal.

Passo 3) Os dados, para cada cópula, são plotados no quadrado $[0,1] \times [0,1]$ e em seguida é realizada uma rotação desse quadrado de 45 graus.

Passo 4) Obtém-se a marginal empírica relativa à diagonal principal.

Passo 5) A diagonal principal é dividida em classes e calcula-se a frequência relativa observada para cada uma das cópulas.

Passo 6) Escolhe-se uma cópula que será definida como hipótese nula: H_0 , gera-se uma nova amostra desta cópula, faz-se a rotação e calcula-se a frequência relativa observada.

Passo 7) Os passos 2, 3, 4, 5 e 6 são repetidos por N vezes.

Passo 8) Calcula-se a estatística Qui-Quadrado usando as frequências relativas observadas de cada cópula e a frequência relativa observada da cópula definida na hipótese nula.

Passo 9) O valor-p é calculado pela função *pchisq* definida no pacote *Chisquare {stats}* no software R e calcula-se as taxas de erro tipo I e tipo II.

Dessa forma, pretende-se verificar se o teste controla de maneira adequada tais erros e, então, se o teste é viável para apontar qual das famílias, dentre as propostas, se ajusta melhor a um determinado conjunto de dados.

3.2 Teste 2: Intervalo de confiança

Este teste propõe compara os coeficientes amostrais de assimetria e curtose, obtidos das marginais na diagonal principal.

Passo 1) Define-se o valor de tau de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$) e o tamanho da amostra ($n = 600$).

Passo 2) Geram-se n dados para cada uma das cópulas, definidas pelo valor do tau de Kendall, nas famílias: Clayton, Gumbel, Frank, Normal.

Passo 3) Os dados, para cada cópula, são plotados no quadrado $[0,1] \times [0,1]$ e em seguida é realizada uma rotação desse quadrado de 45 graus.

Passo 4) Obtém-se a marginal empírica relativa à diagonal principal.

Passo 5) Calcula-se os quatro primeiros momentos amostrais (m_1 , m_2 , m_3 , m_4) relativos a diagonal principal e os coeficientes de assimetria ($\sqrt{b_1}$) e curtose (b_2):

$$\sqrt{b_1} = \frac{m_3}{(m_2)^{3/2}} \quad e \quad b_2 = \frac{m_4}{(m_2)^2}.$$

Passo 5) Repete-se os passos 2, 3, 4 e 5 N vezes.

Passo 6) Constroem-se os intervalos de confiança a nível de 10% para assimetria e curtose para cada uma das cópulas.

Passo 7) Geram-se novamente amostras, rotaciona-se os dados, calcula-se os momentos amostrais e verifica se a assimetria e curtose pertencem ao $IC_{10\%}$ construído no passo 6.

Passo 8) Repete-se por M vezes o passo 7.

Passo 9) Define-se uma cópula como hipótese nula: H_0 e calcula-se as taxas de erro tipo I e tipo II.

Assim, mais uma vez, pretende-se verificar se o teste controla as taxas de erros tipo I e tipo II.

4 RESULTADOS E DISCUSSÃO

Com a finalidade de verificar se existe diferença entre famílias de cópulas, são plotadas as marginais de cada uma das famílias de cópulas (Clayton, Gumbel, Normal e Frank), na diagonal principal e na diagonal secundária.

Na Figura 15 são apresentados os histogramas, na diagonal secundária, de amostras de tamanho $n = 100.000$ de cada uma das família de cópulas. A semelhança entre os histogramas impede qualquer tipo de distinção entre essas famílias.

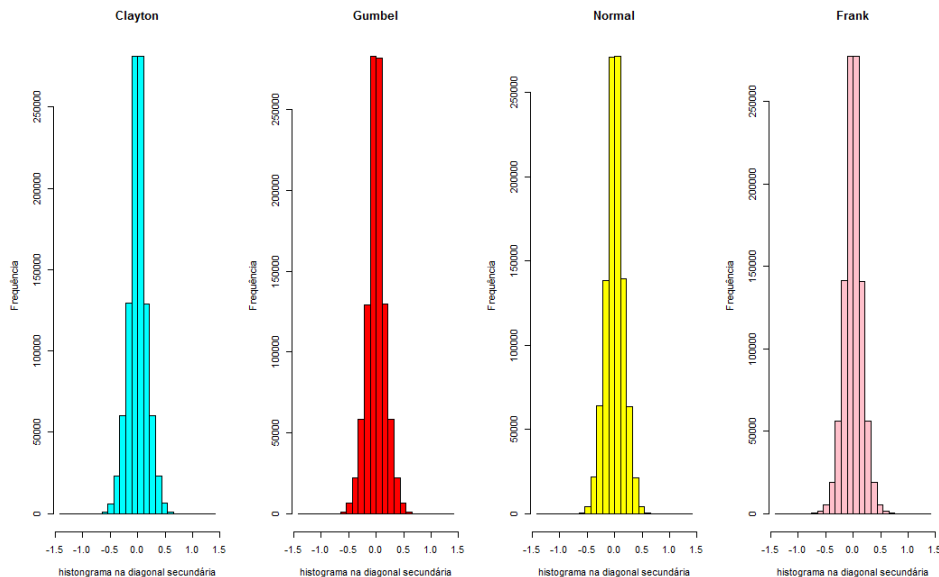


Figura 15. Família de cópulas na diagonal secundária.

Na Figura 16 têm-se os histogramas de amostras ($n = 100.000$) de cada família de cópulas, plotadas na diagonal principal. Os gráficos apresentam uma distinção clara entre as famílias de cópulas apresentadas. Portanto, as análises serão baseadas apenas sobre essa diagonal.

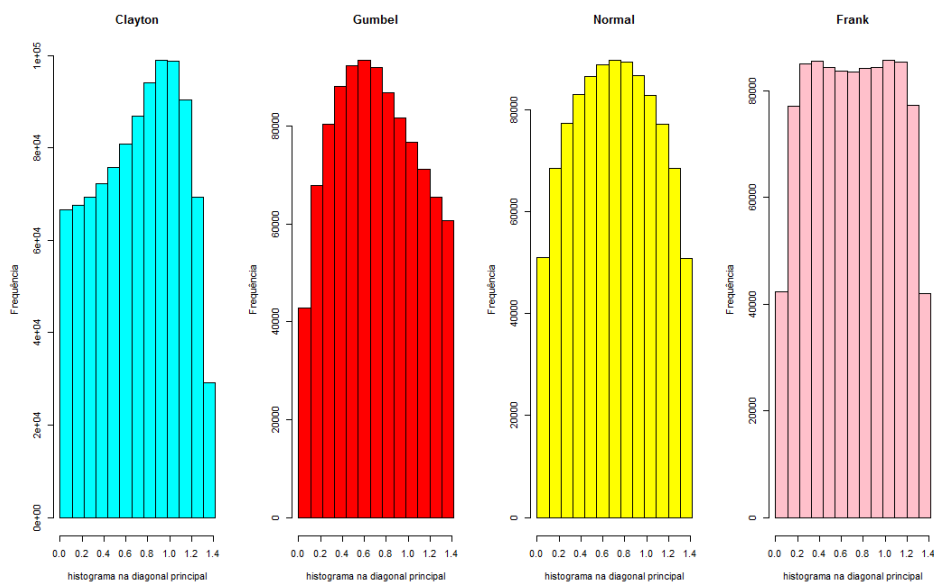


Figura 16. Família de cópulas na diagonal principal.

A fim de uma melhor compreensão, os resultados serão apresentados em duas partes. Cada uma referente a um dos testes propostos na metodologia.

4.1 Parte 1: Qui-quadrado

Realiza-se o teste qui-quadrado para distribuições de frequência na diagonal principal de cada uma das famílias de cópulas e se restringe às caselas localizadas nas extremidades da diagonal, que são as regiões em que as cópulas se distinguem. As frequências esperadas são obtidas tomando-se uma amostra $N = 1.000.000$ da cópula de interesse. O processo é repetido 1.000 vezes no intuito de contabilizar os erros cometidos.

Esse novo método é realizado, supondo a hipótese nula para cada uma das famílias de cópulas de interesse, e comparado com outros dois métodos de

verificação de qualidade de ajuste disponíveis no software R através do pacote *gofCopula {copula}*: PB e MB.

Ao se tomar uma amostra ($n = 600$) da cópula de Clayton, as porcentagens de erro tipo I e erro tipo II, para cada valor de TAU de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$), comparadas com os dois testes disponíveis no R, podem ser observadas na Tabela 4.

Tabela 4 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Clayton a nível de 5% de significância.

$n = 600$		H_0 :	H_a :	H_a :	H_a :
		Clayton	Gumbel	Normal	Frank
$\tau = 0,2$	Novo	0.0550	0.0890	0.4860	0.3390
	PB	0.8696	0.2262	0.0254	0.0134
	MB	0.6178	0.0004	0.0004	0.0004
$\tau = 0,4$	Novo	0.0670	0.0040	0.2360	0.1150
	PB	0.5199	0.7537	0.3511	0.1263
	MB	0.4970	0.0004	0.0004	0.0004
$\tau = 0,5$	Novo	0.0510	0.0050	0.2050	0.1390
	PB	0.2852	0.9605	0.8796	0.5749
	MB	0.6078	0.0004	0.0004	0.0004
$\tau = 0,6$	Novo	0.0680	0.0340	0.2670	0.2170
	PB	0.4150	0.9315	0.8756	0.6388
	MB	0.1563	0.0004	0.0004	0.0004
$\tau = 0,8$	Novo	0.0580	0.5910	0.6920	0.7740
	PB	0.7177	0.5429	0.6158	0.9135
	MB	0.9545	0.0004	0.0004	0.0004

Para a hipótese nula do teste, H_0 : a amostra pertence a família de cópulas de Clayton, a tabela acima indica um bom controle do teste para o erro tipo I a nível de 5% de significância. Em relação à Gumbel, para os valores de TAU de Kendall: $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, a porcentagem de erro tipo II é pequena.

Os testes PB e MB não apresentam bom controle da taxa de erro tipo I, em contra-partida a taxa de erro tipo II, no teste MB é sempre muito baixo.

Para amostra da cópula de Gumbel, as porcentagens de erro tipo I e erro tipo II, para cada valor de TAU de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$) encontram-se na Tabela 5.

Tabela 5 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Gumbel a nível de 5% de significância.

$n = 600$		H_a :	H_0 :	H_a :	H_a :
		Clayton	Gumbel	Normal	Frank
$\tau = 0,2$	Novo	0.2280	0.0580	0.8380	0.7140
	PB	0.3801	0.5000	0.2112	0.1683
	MB	0.0054	0.7337	0.1273	0.0214
$\tau = 0,4$	Novo	0.0030	0.0580	0.7680	0.3900
	PB	0.1953	0.5069	0.1663	0.0274
	MB	0.0004	0.1063	0.0094	0.0004
$\tau = 0,5$	Novo	0.0040	0.0440	0.8200	0.4240
	PB	0.0644	0.7097	0.0966	0.0164
	MB	0.0004	0.5749	0.0004	0.0004
$\tau = 0,6$	Novo	0.0210	0.0440	0.8540	0.5500
	PB	0.0784	0.5489	0.3481	0.1453
	MB	0.0004	0.0194	0.0004	0.0004

$\tau = 0,8$	Novo	0.5320	0.0610	0.9380	0.8870
	PB	0.5079	0.1913	0.8736	0.9855
	MB	0.0004	0.5529	0.0004	0.0004

Se a amostra testada é gerada a partir da família de cópulas de Gumbel, a tabela também indica um bom controle do teste para o erro tipo I. Já o controle do erro tipo II só apresenta um bom desempenho quando a amostra da família Gumbel é comparada à família de Clayton e para os valores $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$.

Em relação aos testes PB e MB, eles continuam não apresentam bom controle da taxa de erro tipo I, exceto em algumas situações. A taxa de erro tipo II, no teste MB continua sempre muito baixa.

Quando é observada uma amostra ($n = 600$) da cópula da Normal, as porcentagens de erro tipo I e erro tipo II, para os valores de TAU de Kendall $\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$ são dados na Tabela 6.

Tabela 6 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas da Normal a nível de 5% de significância.

$n = 600$		$H_a:$	$H_a:$	$H_0:$	$H_a:$
		Clayton	Gumbel	Normal	Frank
$\tau = 0,2$	Novo	0.3860	0.6740	0.1250	0.7980
	PB	0.9525	0.9665	0.2022	0.7077
	MB	0.0004	0.4560	0.2312	0.4270
$\tau = 0,4$	Novo	0.0770	0.6420	0.1200	0.6110
	PB	0.6918	0.8846	0.4270	0.2782
	MB	0.0004	0.4460	0.7067	0.0084
	Novo	0.0740	0.6910	0.1140	0.5670

$\tau = 0,5$	PB	0.0714	0.3151	0.8806	0.0234
	MB	0.0004	0.0004	0.6338	0.0004
	Novo	0.1320	0.7510	0.1220	0.5870
$\tau = 0,6$	PB	0.1053	0.4930	0.6388	0.1243
	MB	0.0004	0.0054	0.4580	0.0044
	Novo	0.5400	0.8690	0.1120	0.7980
$\tau = 0,8$	PB	0.3351	0.5709	0.3101	0.9215
	MB	0.0004	0.0014	0.2022	0.0004

Nesse caso, a amostra testada é gerada a partir família de cópulas Normal e a porcentagem do erro tipo I, apresentada na tabela, é pequena, indicando um bom controle do erro. No entanto, o erro tipo II não é controlado.

Quanto aos testes PB e MB, o controle da taxa de erro tipo I melhora, mas ainda apresenta valores altos. No teste MB a taxa de erro tipo II, continua sempre muito baixa, exceto em algumas situações.

Para amostra de tamanho $n = 600$ da cópula de Frank, as porcentagens de erro tipo I e erro tipo II, para cada valor de TAU de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$) são apresentadas na Tabela 7.

Tabela 7 Porcentagens de erro tipo I e erro tipo II para o teste 1 dado uma amostra de tamanho $n = 600$, da família de cópulas de Frank a nível de 5% de significância.

$n = 600$		$H_a:$	$H_a:$	$H_a:$	$H_0:$
		Clayton	Gumbel	Normal	Frank
	Novo	0.7630	0.8980	0.9010	0.0480
$\tau = 0,2$	PB	0.1763	0.3511	0.4140	0.2312
	MB	0.0004	0.0004	0.0004	0.6788
	Novo	0.1120	0.4690	0.7790	0.0570

$\tau = 0,4$	PB	0.6948	0.8756	0.5369	0.7667
	MB	0.0004	0.0004	0.0484	0.1253
	Novo	0.1210	0.4620	0.6920	0.0490
$\tau = 0,5$	PB	0.6888	0.9395	0.7537	0.5799
	MB	0.0004	0.0034	0.1123	0.4270
	Novo	0.2570	0.5780	0.7180	0.0510
$\tau = 0,6$	PB	0.1933	0.6918	0.5659	0.7577
	MB	0.0004	0.0014	0.0024	0.4440
	Novo	0.7580	0.8830	0.8840	0.0600
$\tau = 0,8$	PB	0.0074	0.0164	0.0084	0.8566
	MB	0.0004	0.0024	0.0024	0.8026

Assim como acontece quando a amostra é gerada a partir da família de cópulas da Normal, quando a amostra é proveniente da Frank, a tabela mostra que apenas o erro tipo I é controlado. A interpretação também é a mesma para os testes PB e MB.

4.2 Parte 2: Intervalo de confiança

Como visto nos resultados apresentados na parte 1, as cópulas não se distinguem pela marginal secundária. Portanto, todas as análises realizadas para o teste 2 foram baseadas apenas na diagonal principal de cada uma das famílias de cópulas em estudo. Esse outro método também é realizado, supondo a hipótese nula para cada uma das famílias de cópulas de interesse, e comparado com outros dois métodos (PB e MB) de verificação de qualidade de ajuste disponíveis no software R.

Construídos os intervalos de confiança para assimetria e curtose para uma amostra de tamanho $n = 600$ da cópula de Clayton, para cada valor de TAU de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$), as porcentagens de erro tipo I e erro tipo II, são apresentadas na Tabela 8.

Tabela 8 Taxa de erro tipo I e erro tipo II no $IC_{90\%}$ da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$.

$n = 600$		H_0 :	H_a :	H_a :	H_a :
		Clayton	Gumbel	Normal	Frank
$\tau = 0,2$	Novo	0.0889	0.0129	0.0111	0.0612
	PB	0.1923	0.8906	0.5459	0.4200
	MB	0.1913	0.0004	0.0414	0.0194
$\tau = 0,4$	Novo	0.0946	0.0045	0.0010	0.0317
	PB	0.1073	0.9685	0.7757	0.4990
	MB	0.1153	0.0004	0.0004	0.0004
$\tau = 0,5$	Novo	0.0923	0.0156	0.0006	0.0399
	PB	0.5169	0.8576	0.6278	0.2782
	MB	0.5109	0.0004	0.0004	0.0004
$\tau = 0,6$	Novo	0.0908	0.6617	0.0016	0.0214
	PB	0.7017	0.6388	0.6848	0.9075
	MB	0.9295	0.0004	0.0004	0.0004
$\tau = 0,8$	Novo	0.0980	0.6736	0.0017	0.0234
	PB	0.4610	0.8616	0.9145	0.9845
	MB	0.1413	0.0004	0.0004	0.0004

Nesse caso, em que a amostra é da cópula de Clayton, a taxa de erro tipo I é bem controlada a nível de 10% de significância, sendo o maior valor assumido

0.094. Além disso, há um controle do erro tipo II, com exceção da família Gumbel e valores de TAU de Kendall, $\tau = 0,6$ e $\tau = 0,8$.

Os testes PB e MB não apresentam bom controle da taxa de erro tipo I, com exceção de alguns casos e em contra-partida a taxa de erro tipo II, no teste MB é sempre muito baixo.

Para o intervalo de confiança de assimetria e curtose da cópula de Gumbel, em uma amostra de tamanho $n = 600$ e para cada valor de TAU de Kendall ($\tau = 0,2$, $\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$, $\tau = 0,8$), as porcentagens de erro tipo I e erro tipo II, podem ser visualizadas na Tabela 9.

Tabela 9 Taxa de erro tipo I e erro tipo II no $IC_{90\%}$ da família de cópulas de Gumbel, dado uma amostra de tamanho $n = 600$.

$n = 600$		$H_a:$	$H_0:$	$H_a:$	$H_a:$
		Clayton	Gumbel	Normal	Frank
$\tau = 0,2$	Novo	0.0125	0.0908	0.6058	0.4217
	PB	0.0144	0.9465	0.0014	0.0014
	MB	0.0004	0.1883	0.0414	0.0074
$\tau = 0,4$	Novo	0.0048	0.0945	0.6405	0.3190
	PB	0.0594	0.8026	0.0324	0.0044
	MB	0.0004	0.9035	0.4200	0.0104
$\tau = 0,5$	Novo	0.0180	0.0878	0.7262	0.3370
	PB	0.5269	0.1683	0.5179	0.1833
	MB	0.0004	0.2342	0.0074	0.0004
$\tau = 0,6$	Novo	0.7148	0.0915	0.8999	0.8204
	PB	0.4020	0.2782	0.8006	0.9495
	MB	0.0004	0.5909	0.0004	0.0004
	Novo	0.6955	0.0978	0.8866	0.7962

$\tau = 0,8$	PB	0.1883	0.6118	0.4860	0.8026
	MB	0.0004	0.0244	0.0044	0.0004

É possível observar que a taxa de erro tipo I também está controlada quando a hipótese nula envolve a família de cópulas da Gumbel. Porém, o teste não controla a taxa de erro tipo II.

O teste MB apresenta uma melhora no controle da taxa de erro tipo I, com exceção de alguns casos e ainda a taxa de erro tipo II, no teste MB é sempre muito baixa. Os resultados para o teste PB oscilam muito.

No caso do intervalo de confiança de assimetria e curtose da cópula da Normal, para os mesmos valores de TAU de Kendall, as porcentagens de erro tipo I e erro tipo II, encontram-se descritas na Tabela 10.

Tabela 10 Taxa de erro tipo I e erro tipo II no $IC_{90\%}$ da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$.

$n = 600$		$H_a:$ Clayton	$H_a:$ Gumbel	$H_0:$ Normal	$H_a:$ Frank
$\tau = 0,2$	Novo	0.2430	0.5868	0.0929	0.8034
	PB	0.9945	0.9945	0.0394	0.9155
	MB	0.0004	0.0414	0.5029	0.5159
$\tau = 0,4$	Novo	0.1460	0.6079	0.0931	0.5998
	PB	0.2482	0.5559	0.8296	0.0494
	MB	0.0004	0.0004	0.8236	0.0044
$\tau = 0,5$	Novo	0.2137	0.7040	0.0838	0.5206
	PB	0.3611	0.7597	0.5319	0.1803
	MB	0.0004	0.0004	0.0944	0.0034
$\tau = 0,6$	Novo	0.7727	0.8991	0.0942	0.8259
	PB	0.3091	0.6008	0.3161	0.8826

	MB	0.0004	0.1193	0.2382	0.0004
$\tau = 0,8$	Novo	0.7518	0.8944	0.1007	0.8003
	PB	0.2892	0.5639	0.3521	0.9305
	MB	0.0004	0.1273	0.1143	0.0014

Quando a amostra é proveniente da cópula Normal, o teste controla apenas o erro tipo I. A taxa de erro tipo II não é controlada para nenhum valor de TAU de Kendall.

Os testes PB e MB não apresentam bom controle da taxa de erro tipo I, com exceção de alguns casos e novamente a taxa de erro tipo II, no teste MB, é sempre muito baixa.

Para os mesmos valores de TAU de Kendall, as porcentagens de erro tipo I e erro tipo II, para o intervalo de confiança de assimetria e curtose da cópula de Frank são apresentados na Tabela 11.

Tabela 11 Taxa de erro tipo I e erro tipo II no $IC_{90\%}$ da família de cópulas de Clayton, dado uma amostra de tamanho $n = 600$.

$n = 600$		$H_a:$ Clayton	$H_a:$ Gumbel	$H_a:$ Normal	$H_0:$ Frank
$\tau = 0,2$	Novo	0.1903	0.3861	0.7972	0.0964
	PB	0.9225	0.9615	0.6898	0.4730
	MB	0.0044	0.0004	0.0674	0.6668
$\tau = 0,4$	Novo	0.0989	0.2634	0.5611	0.0997
	PB	0.9555	0.9855	0.8716	0.3771
	MB	0.0004	0.0004	0.0854	0.6018
$\tau = 0,5$	Novo	0.1740	0.3126	0.5189	0.0938
	PB	0.5909	0.8906	0.7397	0.6378
	MB	0.0004	0.0004	0.0094	0.4730

$\tau = 0,6$	Novo	0.8345	0.7931	0.7962	0.0920
	PB	0.0434	0.1053	0.1173	0.5759
	MB	0.0004	0.0004	0.0014	0.5169
$\tau = 0,8$	Novo	0.8231	0.7959	0.7880	0.1008
	PB	0.1373	0.2782	0.3281	0.2592
	MB	0.0004	0.0004	0.0014	0.6018

Com a hipótese nula envolvendo a Frank, temos o mesmo resultado obtido no caso anterior e, portanto, o teste controla apenas a taxa de erro tipo I. As análises são as mesmas do caso anterior também, para os teste PB e MP.

O comportamento das famílias de cópulas em estudo é muito parecido, portanto, já se espera que dificuldades sejam encontradas em qualquer tipo de teste de qualidade de ajuste.

Em todos os casos analisados os dois testes propostos nesse trabalho apresentaram um bom controle do erro tipo I. Porém, ainda há uma falha quando se trata do controle da taxa de erro tipo II.

O teste 2 apresenta um melhor desempenho do controle da taxa de erro tipo II quando se utilizam os valores intermediários de TAU de Kendall ($\tau = 0,4$, $\tau = 0,5$, $\tau = 0,6$). Uma provável razão para isso é que, para τ próximo de zero (dados se aproximam da independência) ou τ próximo de 1 (dados se aproximam da dependência positiva perfeita) as cópulas se aproximam, respectivamente, da cópula da independência ($C_i(\mathbf{u}, \mathbf{v}) = \mathbf{uv}$) da cópula do mínimo ($C_1(\mathbf{u}, \mathbf{v}) = \min\{\mathbf{u}, \mathbf{v}\}$). Em ambos os casos as cópulas tem o comportamento parecido, o que dificulta a distinção.

Apesar de não apresentarem valores satisfatórios para todas as famílias de cópulas analisadas: Clayton, Gumbel, Normal e Frank, os dois testes propostos apresentam um bom potencial para diferenciar as cópulas das famílias Clayton e Gumbel.

5 CONCLUSÃO

- O teste de ajuste baseado nas marginais pela diagonal principal utilizando uma estatística qui-quadrado apresenta problemas, pois não controla de forma adequada a taxa de erro tipo II.
- O teste de ajuste baseado nas marginais pela diagonal principal utilizando intervalos de confiança também apresenta problemas no controle da taxa de erro tipo II, porém com desempenho melhor.

REFERÊNCIAS

GENEST, C.; FAVRE, A-C. Everything you always wanted to know about copula modeling but were afraid to ask. **Journal of hydrologic engineering**, v. 12, n. 4, p. 347-368, 2007.

GENEST, C.; HUANG, W.; DUFOUR, J-M. A regularized goodness-of-fit test for copulas. **Journal de la Société Française de Statistique**, v. 154, n. 1, p. 64-77, 2013.

GENEST, C.; RÉMILLARD, B. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In: **Annales de l'IHP Probabilités et statistiques**. 2008. p. 1096-1127.

GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. **Insurance: Mathematics and economics**, v. 44, n. 2, p. 199-213, 2009.

JOE, H. **Multivariate models and multivariate dependence concepts**. CRC Press, 1997.

KOJADINOVIC, I.; YAN, J.; HOLMES, M. Fast large-sample goodness-of-fit tests for copulas. **Statistica Sinica**, p. 841-871, 2011.

NELSEN, R. B. Dependence modeling with archimedean copulas. 2005.

CAPÍTULO 3 TESTE PARA DEPENDÊNCIA EM VALORES EXTREMOS, UTILIZANDO CÓPULAS

RESUMO

É grande o número de estudos recentes relacionados a dados bivariados que possuem dependência de valores extremos, e é crescente sua relevância em diversas áreas. Duas variáveis aleatórias têm “dependência de valor extremo” se é alta a probabilidade de ocorrer valor extremo em uma, dado que ocorreu valor extremo na outra. Uma ferramenta no estudo dessa dependência são as cópulas de valores extremos, que se caracterizam por funções de Pickands convexas. Vários estimadores não paramétricos das funções de Pickands têm sido propostos como, por exemplo, o uso de B-splines. Neste trabalho é proposto um novo estimador utilizando polinômios de Bernstein, sujeitos a três tipos de restrições previamente estabelecidas. O desempenho deste novo estimador em detectar dependência de valor extremo é estudado via simulação utilizando-se as cópulas Gumbel, Clayton, Frank, Gaussiana e t com 4 graus de liberdade. Para o ajuste da função de Pickands foi utilizado o *IMSE (Integrated Mean Square Error)*. Um teste de hipótese é desenvolvido apresentando um bom controle da taxa de erro tipo I, porém, o mesmo, não controla a taxa de erro tipo II. Para o ajuste via polinômios de Bernstein restrito e o teste de hipótese foram implementadas rotinas no software livre R que apresentam velocidades de execução substancialmente inferiores às de outras rotinas já existentes.

Palavras-chave: função de Pickands, polinômios de Bernstein, dependência de valores extremos.

ABSTRACT

There is a great number of recent studies about bivariate data with dependence on extreme values, and is increasing its relevance in many areas. Two random variables have "extreme value dependence" if there is a high probability to occur extreme value in one, given that extreme value has occurred in the other. A tool in the study of this dependence are the copulas of extreme values, which are characterized by Pickands convex functions. Several non-parametric estimators of Pickands functions have been proposed, for example, the use of B-splines. This paper proposes a new estimator using Bernstein polynomials, subject to three types of previously established restrictions. The performance of this new estimator to detect extreme dependence value is studied via simulation using data from the copulas: Gumbel, Clayton, Frank, Gaussian and t with 4 degrees of freedom. The adjust the Pickands function uses IMSE (Integrated Mean Square Error). A hypothesis test is developed showing good control of Type I error rate. However, it does not control the type II error rate. The fit via restricted Bernstein polynomials and the hypothesis test were implemented routines in software R, which have execution substantially faster than other existing routines.

Keywords: Pickands function, polynomial Bernstein, dependence on extreme values.

1 INTRODUÇÃO

É grande o número de estudos recentes relacionados às cópulas de valores extremos e inegável sua crescente relevância em diversas áreas tais como: matemática financeira, economia, ciências atuariais e hidrologia como pode ser visto em Genest e Favre (2007) e Nelsen (2003). A teoria geral sobre cópulas pode ser encontrada em Joe (1997), Nelsen (2013) e em Cherubini et al. (2004).

Duas variáveis aleatórias têm “dependência de valor extremo” (ou dependência em eventos raros) se é alta a probabilidade de ocorrer valor extremo em um, dado que o outro mostrou um valor extremo.

Atualmente, várias propostas de métodos de verificação de dependência extrema são encontradas na literatura. Kotz e Nadarajah (2000) e Padoan (2013) mostram que muitos modelos paramétricos foram introduzidos para modelar este tipo de dependência.

Segundo Cherubini et al. (2004), a forma como as cópulas representam uma distribuição conjunta, captando diversos tipos de dependência e identificando as relações entre os valores extremos é uma grande vantagem em relação a outros modelos matemáticos. Aplicações importantes de valores extremos podem ser encontradas em Coles et al. (1999) ou em Cebrian et al. (2003), entre outros.

Dizemos que, se o par (X, Y) tem dependência de valor extremo, então, a cópula C correspondente pertence à classe de cópulas de valor extremo. Pickands (1981) mostrou que isso ocorre se, e somente se, para todos os valores de $u, v \in (0,1)$, C puder ser expressa da seguinte forma:

$$C(u, v) = \exp \left[\log(uv) A \left\{ \frac{\log(u)}{\log(uv)} \right\} \right]$$

em que a chamada função de Pickands $A: [0,1] \rightarrow \left[\frac{1}{2}, 1\right]$ é convexa e satisfaz $\max(t, 1 - t) \leq A(t) \leq 1$, para todos os valores de $t \in [0,1]$. Qualquer função A que satisfaça essas duas restrições, corresponde a uma cópula de valor extremo.

Vários estimadores não paramétricos da função de Pickands têm sido propostos para modelar a dependência de valores extremos em diversas áreas. Alguns exemplos podem ser vistos em Pickands (1981), Capèra et al. (1997), Hall e Tajvidi (2000), Zhang et al. (2008), Genest e Segers (2009), Gudendorf e Segers (2011, 2012) e Marcon et al. (2014). Em particular, os polinômios de Bernstein podem ser usados para estimar função de Pickands.

Esses polinômios foram utilizados como um tipo de estimador não paramétrico para ajustar uma curva monótona, como pode ser visto em Chak et al. (2005), Chang et al. (2005), Chang et al. (2007), McKay e Ghosh (2011), Petrone (1999) e Stadtmuller (1986), entre outros. Wang e Ghosh (2012) propuseram um estimador com determinadas restrições (por exemplo, monotonicidade e/ou convexidade) que são mantidas e satisfeitas para todo e qualquer tamanho de amostra.

Além disso, Wang e Ghosh (2012) mostraram também que a base do estimador de regressão do polinômio de Bernstein pode ser obtida a partir da solução de um problema de mínimos quadrados. Por conseguinte, o estimador pode ser calculado eficientemente através de um algoritmo de programação quadrática.

Este capítulo está preocupado com a estimação não-paramétrica da função de Pickands, com base na otimização de polinômios de Bernstein, verificando a combinação entre três tipos de restrições previamente estabelecidas. Esse tema ainda não foi claramente demonstrado na literatura para a versão bivariada. Sendo assim, há uma novidade neste trabalho com resultado considerável, já que

fornece um estimador que atende as condições descritas acima, que são de aplicação prática para diversas áreas como, por exemplo, a economia e hidrologia.

2 OBJETIVOS

O objetivo principal deste capítulo é propor um novo método de estimação da função de dependência de Pickands utilizando como função regressora polinômios de Bernstein convexos. Esse método permite então a obtenção de cópulas de valores extremos que podem ser modelo de um conjunto de dados.

Os objetivos específicos do estudo são:

- a) comparar ajuste de cópulas utilizando a soma de quadrados relativa a estimação de Pickands utilizando polinômios de Bernstein convexos com o método implementado no pacote *Cobs*, já existente no software R.
- b) propor e avaliar o desempenho de um novo teste de significância para a hipótese nula dos dados serem provenientes de uma cópula de valor extremo.
- c) investigar o efeito na regressão utilizando polinômios de Bernstein com restrições nos valores e derivadas nos pontos extremo.
- d) comparar o desempenho do novo teste com outros existentes na literatura para as famílias de cópulas Clayton, Gumbel, Gaussiana, Frank e t com 4 graus de liberdade (t_4), desenvolvidos em Cormier et al. (2014).

3 METODOLOGIA

Para facilitar a compreensão, dividimos a metodologia em partes. A primeira está relacionada à construção do estimador da função de Pickands utilizando regressão linear em polinômios de Bernstein, e a segunda relativa a um teste de hipótese para detectar cópulas de valor extremo.

3.1 Estimação da função de Pickands utilizando polinômios de Bernstein

Cormier et al. (2014) utilizam B-splines para estimar funções de dependência de Pickands e obter um teste para detectar cópulas de valor extremo. O método é bastante complexo e computacionalmente oneroso.

Neste trabalho, seguindo a sugestão do professor Dr Yan, da University of Connecticut (UCONN), nos Estados Unidos, serão utilizados polinômios de Bernstein para a estimação da função de Pickands. Tal fato é natural pois, para os polinômios de Bernstein, regressão quadrática com restrição de convexidade transforma-se simplesmente em um problema de regressão quadrática com restrições lineares e pode ser resolvido por vários algoritmos implementados. Este novo método será comparado com outros métodos, dentre eles, o método de B-splines.

A regressão com restrição de convexidade para os polinômios de Bernstein é obtida simplesmente analisando a desigualdade da derivada segunda $B_n''(x) \geq 0$. A função de Pickands satisfaz $A: [0,1] \rightarrow [1/2,1]$ e além da convexidade possui algumas propriedades.

Nos extremos tem-se $A(0) = A(1) = 1$. Considerando a curva parametrizada $(t, A(t))$ e pela propriedade de $A(t) \geq \max\{t, 1-t\}$ tem-se que

o vetor derivada $(1, A'(t))$ para todo $t=0$ é um vetor na metade superior do quarto quadrante e, portanto, $A'(0) \geq -1$ e para $t=1$ é um vetor na metade superior do primeiro quadrante e portanto $0 < A'(0) < 1$.

Dessa forma, a regressão deve ser restrita para os polinômios de Bernstein que satisfazem:

R1) A primeira restrição especifica os valores assumidos pela função de Pickands em seus extremos:

$$B_n(0) = \sum_{k=0}^n \beta_k b_{n,k}(0) = \sum_{k=0}^n \beta_k \binom{n}{k} 0^k (1-0)^{n-k} = \beta_0 = 1.$$

$$B_n(1) = \sum_{k=0}^n \beta_k b_{n,k}(1) = \sum_{k=0}^n \beta_k \binom{n}{k} 1^k (1-1)^{n-k} = \beta_n = 1.$$

R2) A segunda restrição refere-se às primeiras derivadas da função aplicadas em seus extremos:

$$-1 \leq B_n'(0) = n(\beta_n - \beta_{n-1})0^{n-1} + n(\beta_1 - \beta_0)(1-0)^{n-1} + n \sum_{k=1}^{n-2} (\beta_{k+1} - \beta_k) \binom{n-1}{k} 0^k (1-0)^{n-1-k} \leq 0$$

$$-1 \leq B_n'(0) = n(\beta_1 - \beta_0) \leq 0 \Rightarrow (\beta_1 - \beta_0) \geq \frac{-1}{n} \text{ e,}$$

$$0 \leq B_n'(1) = n(\beta_n - \beta_{n-1})1^{n-1} + n(\beta_1 - \beta_0)(1-1)^{n-1} + n \sum_{k=1}^{n-2} (\beta_{k+1} - \beta_k) \binom{n-1}{k} 1^k (1-1)^{n-1-k} \leq 1.$$

$$0 \leq B_n'(1) = n(\beta_n - \beta_{n-1}) \leq 1 \Rightarrow (\beta_{n-1} - \beta_n) \geq \frac{-1}{n}.$$

R3) A terceira restrição é $B_n''(x) \geq 0$, que garante a convexidade. Esta restrição resulta em $N - 1$ restrições.

$$\beta_k - 2\beta_{k+1} + \beta_{k+2} \geq 0 \text{ para } k = 0, 1, \dots, n-2.$$

Portanto, o problema de estimação da função de Pickands utilizando polinômios de Bernstein é um problema de regressão linear que minimiza a

soma de quadrados $\sum_{i=1}^N (B_n(x_i) - y_i)^2 = \sum_{i=1}^N \left(\sum_{k=0}^n \beta_k b_{k,n}(x_i) - y_i \right)^2$, com o

conjunto de restrições que podem ser expressas na forma matricial por:

$$\begin{bmatrix} 1 & 0 & 0 & & \dots & 0 \\ -1 & 1 & 0 & & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots & \vdots \\ 0 & \dots & & 0 & 1 & -2 & 1 \\ 0 & \dots & & & 0 & 1 & -1 \\ 0 & \dots & & & & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n-1} \\ \beta_n \end{bmatrix} \geq \begin{bmatrix} 1 \\ -1/n \\ 0 \\ \vdots \\ 0 \\ -1/n \\ 1 \end{bmatrix}$$

em que, na primeira e última desigualdades vale a igualdade.

E a desigualdade que limita o tamanho dos parâmetros é:

$$\sum_{k=0}^n |\beta_k| < M_n.$$

Esse problema pode ser resolvido utilizando a rotina “*solve.QP*” do pacote *quadprog*. Essa rotina é a implementação do método de Gudendorf e Segers (2011, 2012) para a resolução de problemas de programação quadrática sujeito um conjunto de restrições de desigualdades lineares.

3.2 Implementação do estimador da função de Pickands

O estimador da função de Pickands é implementado utilizando uma transformação T_C . Dada a cópula $C(u, v)$ considere a transformação $T_C : [0, 1]^2 \rightarrow [0, 1] \times [0, +\infty)$, definida para todo $(u, v) \in [0, 1]^2$, como

$$(u, v) \rightarrow (t, z) = \left(t = \frac{\ln(v)}{\ln(uv)}, z = \frac{\ln(C(u, v))}{\ln(uv)} \right).$$

Se a cópula é de valor extremo esta função se degenera e a imagem de T_C deixa de ser toda a faixa $[0, 1] \times [0, +\infty)$ e se torna o gráfico de uma função convexa, a função de Pickands. De fato, $C(u, v) = uv^{A\left(\frac{\ln v}{\ln uv}\right)}$ e

$$T_C(u, v) = \left(t = \frac{\ln(v)}{\ln(uv)}, z = \frac{\ln\left(uv^{A\left(\frac{\ln v}{\ln uv}\right)}\right)}{\ln(uv)} \right) = (t, A(t)).$$

A ideia então é que, se $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$ é uma amostra de uma cópula de valor extremo com parametrização de Pickands $A(t)$, $(t_1, z_1) = T_C(u_1, v_1), \dots, (t_n, z_n) = T_C(u_n, v_n)$ são pontos sobre o gráfico $(t, A(t))$.

Desta forma, ao se realizar a regressão linear com as restrições descritas acima, o ajuste é perfeito, isto é, a soma de quadrados dos pontos observados e dos ajustados é nula. Se por outro lado a cópula não é de valor extremo, os pontos $(t_1, z_1), (t_2, z_2), \dots, (t_n, z_n)$ não estão mais sobre uma curva convexa e a

solução da regressão linear terá soma de quadrados entre os pontos observados e os ajustados positivo e é razoável supor que, quanto maior for essa soma, mais a cópula se afasta de ser uma cópula de valor extremo.

Para o processo de simulação, temos os passos:

Passo 1) Tome uma amostra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ de uma bivariada (X, Y) .

Passo 2) Ordene em ordem crescente x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n . Obtenha os postos correspondentes (n_i, m_i) em que n_i é o posto de x_i e m_i é o posto de y_i .

Passo 3) Obtenha agora a amostra no quadrado $(0,1) \times (0,1)$ através da transformação dada por $\left(\frac{n_1}{n+1}, \frac{m_1}{n+1}\right), \dots, \left(\frac{n_n}{n+1}, \frac{m_n}{n+1}\right)$. A divisão por $n+1$ e não por n é para que não se tenha, em toda a amostra, o ponto $(1,1)$.

Passo 4) Tome para a amostra $\left(\frac{n_1}{n+1}, \frac{m_1}{n+1}\right), \dots, \left(\frac{n_n}{n+1}, \frac{m_n}{n+1}\right)$ os pontos obtidos pela transformação T_C .

Passo 5) A regressão linear, com as respectivas restrições é obtida com o uso da rotina em R.

Como será necessário simular dados com dependência de valor extremo, será adotada a seguinte estratégia: no passo 1, a amostra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ será gerada pela cópula de valor extremo da família Gumbel e em seguida, aplica-se os passos 2, 3, 4 e 5.

3.3 Performance do estimador da função de Pickands

Uma vez que o desempenho do estimador da função de Pickands depende do número de graus de liberdade df , isto é, do grau do polinômio de Bernstein utilizado, têm-se $df = n$. Como o grau do polinômio impacta diretamente na qualidade de ajuste, é razoável utilizar um critério de seleção penalizado. Cormier et al. (2014) sugere como melhor critério de seleção, o “*Schwarz Information Criterion*” (SIC) com penalização, dado por:

$$SIC(n) = \ln \left(\frac{1}{N} \sum_{i=1}^N \left(\hat{B}_n(x_i) - y_i \right)^2 \right) + \frac{\ln(N)}{2N} n,$$

em que B_n o polinômio de Bernstein que estima a função de Pickands e n é o grau do polinômio.

Amostras com tamanhos $n = 100$, $n = 200$ e $n = 500$ da cópula Gumbel são geradas para estimar a função de Pickands.

O processo de estimação será realizado considerando as restrições (R_1, R_2, R_3) . Considerando duas restrições (R_1, R_3) e apenas uma restrição (R_3) . Para cada uma das configurações de simulação, o procedimento será repetido por 1000 vezes. Esses métodos, posteriormente, serão comparados com o método de estimação de *cobs*, que é um pacote implementado no R e utilizado para ajustar uma curva de forma suave usando restrição qualitativa via programação linear (He e Ng, 1999).

O desempenho dos estimadores é avaliado pelo “*Integrated Mean Squared Error*” (IMSE) que é o erro quadrático médio, dado por:

$$IMSE = \frac{1}{N} \sum_{i=1}^N \left(\hat{B}_n(x_i) - y_i \right)^2$$

3.4 Um teste de significância para cópulas de valor extremo

O novo método de estimação da função de Pickands utilizando polinômios de Bernstein permite a obtenção de um teste de significância para avaliar se um conjunto de dados pode ser representado de forma apropriada por um modelo com dependência de valor extremo.

Esse teste segue os mesmos procedimentos utilizados por Cormier et al. (2014) e baseado no conceito de A-plot, descrito de maneira simplificada como:

Dados realizações de uma variável aleatória bidimensional (X, Y) , $\{x_1, x_2, \dots, x_N\}$ e $\{y_1, y_2, \dots, y_N\}$. As distribuições marginais de X e Y serão estimadas pelas acumuladas empíricas:

$$F_N(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(X_i \leq t) \quad G_N(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(Y_i \leq t).$$

Definindo as variáveis aleatórias $\hat{U}_i = F_N(X_i)$, $\hat{V}_i = G_N(Y_i)$. Observe que pelo teorema da transformação da probabilidade integral, as distribuições de \hat{U}_i e \hat{V}_i são próximas da distribuição uniforme. A idéia agora é construir a cópula empírica N como a distribuição empírica de $(\hat{U}_1, \hat{V}_1), \dots, (\hat{U}_N, \hat{V}_N)$. Se $(u_1, v_1), \dots, (u_N, v_N)$ é uma realização dessas variáveis aleatórias então:

$$\hat{C}_N(u, v) = \frac{1}{N+1} \#\{(u_i, v_i), u_i \leq u, v_i \leq v\}.$$

Divide-se por $N+1$ para que \hat{C}_N não assumam, necessariamente, o valor 1.

Constrói-se os pares $(t_1, z_1), \dots, (t_N, z_N)$, em que:

$$t_i = \frac{\ln(v_i)}{\ln(u_i v_i)}, \quad z_i = \frac{\ln(\hat{C}_N(u_i, v_i))}{\ln(u_i v_i)}.$$

A idéia aqui é que, se a cópula C é de valor extremo, os pontos (t_i, z_i) devem estar próximos ao gráfico da função de Pickands da cópula C . Esse procedimento é denominado de A-plot.

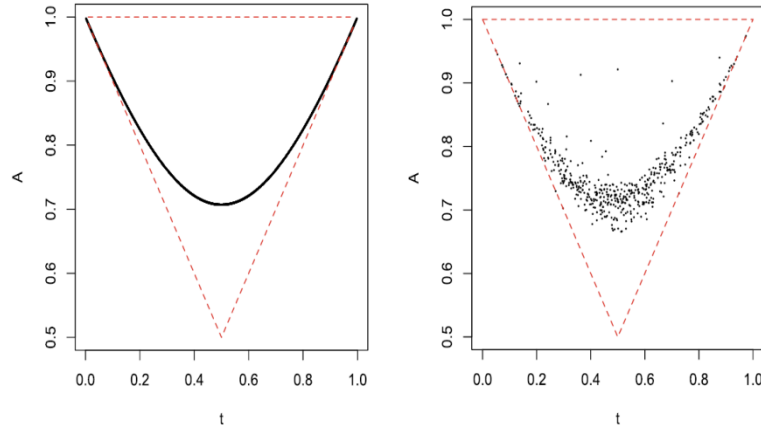


Figura 17. A-plot de uma cópula de valor extremo.

A hipótese nula H_0 é que os pontos observados (u_i, v_i) são provenientes de uma cópula de valor extremo C . Com os pontos (t_i, z_i) , minimizando a média da soma dos quadrados, um polinômio de Bernstein $B_n(t)$ é ajustado e denominado $\hat{A}(t)$ como uma estimativa da função de Pickands.

A estatística do teste é a soma dos desvios entre os pontos observados e os pontos ajustados:

$$T_N = \frac{1}{N} \sum_{i=1}^N \{z_i - \hat{A}(t_i)\}^2.$$

O teste consiste em, rejeitar H_0 se T_N for suficientemente grande. A distribuição nula da estatística T_N não é conhecida e é calculada de forma aproximada utilizando reamostragem bootstrap paramétrica da seguinte forma:

1 – Utilizando os dados $(u_1, v_1), \dots, (u_N, v_N)$, calcula-se o τ amostral que será utilizado como uma estimativa para o TAU populacional.

2 – Um representante da família de cópula de valor extremo é escolhido. Nesse trabalho, a cópula escolhida foi a cópula de Gumbel.

3 – Com a cópula de Gumbel definida pelo TAU amostral, gera-se um conjunto de dados $(u_{11}, v_{11}), \dots, (u_{N1}, v_{N1})$.

4 – Utilizando a transformação da função de Pickands obtêm-se os pontos $(t_1, z_1), \dots, (t_N, z_N)$.

5 – Um polinômio de Bernstein $B_{n,1}(t)$ é obtido como estimativa $\hat{A}_1(t)$ da função função de Pickands.

6 – O valor da estatística $T_{N,1}$ é calculado por:

$$T_{N,1} = \frac{1}{N} \sum_{i=1}^N \left\{ z_{i,1} - \hat{A}_1(t_{i,1}) \right\}^2 .$$

7 – Os passos 3, 4, 5 e 6 são repetidos N_b vezes.

8 – Com os valores de $T_{n,j}$, $j = 1, 2, \dots, N_b$, a distribuição empírica de T_N é obtida.

O valor-p é obtido dividindo-se por N_b o número de vezes que $T_{N,j}$ é maior que T_n , isto é:

$$valor - p = \frac{1}{N_b} \sum_{j=1}^{N_b} \mathbf{1}(T_{N,j} \geq T_n).$$

O valor-p é confrontado com o nível de significância α adotado.

- Se $valor - p \leq \alpha$, rejeita-se H_0 .
- Se $valor - p > \alpha$, não rejeita-se H_0 .

3.5 Validação do teste

Afim de verificar a taxa de erro tipo I, são geradas 1.000 repetições de amostras de tamanho 200 da cópula Gumbel (família de cópula de valor extremo), usando os valores de TAU de Kendall dados por $\tau = 0,25$, $\tau = 0,50$, $\tau = 0,75$. Toma-se o nível nominal de 5%.

Para avaliar o poder do teste proposto, isso é, a taxa de erro tipo II, dados são gerados para cada uma das cópulas: Clayton, Frank, Gaussiana, t_4 , que não são cópulas de valor extremo, usando as mesmas configurações de repetição, tamanho de amostra e valores de TAU de Kendall ($\tau = 0,25$, $\tau = 0,50$, $\tau = 0,75$).

4 RESULTADOS E DISCUSSÃO

Este capítulo propõe um estimador não paramétrico para ajustar a função de Pickands através de polinômios de Bernstein com uma certa variedade de restrições que podem ser obtidos como a solução de um problema de programação quadrática para um valor fixo de $df = n$.

Utilizamos o método *SIC* como critério para selecionar o valor ideal de df . Esses valores diferem de acordo com o valor assumido por TAU de Kendall. Quando maior o valor de TAU de Kendall, mais elevado é o valor assumido por df .

A Tabela 12 mostra, o resumo dos resultados de “*Integrated Mean Squared Error*” (IMSE) para a cópula Gumbel. A coluna $(R_1 + R_2 + R_3)$ apresenta o resultado quando são utilizadas todas as restrições: os valores da função de Pickands no extremos (0,1) e (1,1), as primeiras derivadas nos extremos e as $N - 1$ segundas derivadas correspondentes às restrições de convexidade. A coluna $(R_1 + R_3)$ mostra os resultados quando são usadas as restrições correspondentes a $f(0) = f(1) = 1$ e a restrição relacionada à convexidade. A coluna (R_3) apresenta os resultados quando apenas a restrição de convexidade é utilizada. E a coluna cobs apresenta os resultados quando se utiliza o pacote “cobs” do R.

Os resultados são comparáveis (inferiores) e satisfatórios para a cópula Gumbel em relação aos obtidos quando se usa o método de “cobs”. Tais resultados não evidenciam vantagens quando são usadas todas as restrições $(R_1 + R_2 + R_3)$, duas restrições, $(R_1 + R_3)$ ou apenas uma restrição (R_3) , o que evidencia que o uso das três restrições, como proposto, não diminui a qualidade de ajuste.

Tabela 12. Erro quadrático médio (IMSE) da função ajustada de Pickands, sob 1000 repetições, para a cópula Gumbel. As amostras têm tamanho $n = 100$, $n = 200$ e $n = 500$ e os valores assumidos por TAU de Kendall são 0,25; 0,50 e 0,75.

tamanho da amostra	valor de τ	df	$R_1 + R_2 + R_3$	$R_1 + R_3$	R_3	<i>cobs</i>
$n = 100$	0.25	4	0.2091	0.2095	0.2082	0.2126
	0.50	7	0.1406	0.1405	0.1405	0.1414
	0.75	11	0.1046	0.0979	0.0977	0.1114
$n = 200$	0.25	4	0.2169	0.2143	0.2149	0.2191
	0.50	7	0.1424	0.1425	0.1427	0.1556
	0.75	11	0.1050	0.0979	0.0987	0.1085
$n = 500$	0.25	4	0.2193	0.2195	0.2196	0.2307
	0.50	7	0.1441	0.1440	0.1446	0.1506
	0.75	11	0.1046	0.0984	0.0991	0.1062

A principal qualidade do método proposto em relação ao método já existente na literatura é sua velocidade de execução. A Tabela 13 apresenta, para três tamanhos diferentes de amostra ($n = 100$, $n = 200$, $n = 500$), o tempo (em segundos) da execução do método proposto para cada uma das três combinações de restrições: todas as restrições ($R_1 + R_2 + R_3$), duas restrições, ($R_1 + R_3$) e apenas uma restrição (R_3); para o *cobs*; e ainda a relação média entre *cobs* e as combinações de restrições.

O tempo não varia quando se utiliza as três, duas, ou apenas uma das restrições. Em todos os casos, esse tempo é substancialmente inferior ao tempo de execução do método *cobs*. Para amostras de tamanho $n = 100$, a grandeza é de 21 vezes mais veloz. À medida que o tamanho da amostra aumenta, essa grandeza de velocidade diminui, no entanto, continua sendo 7 vezes mais eficaz quando a amostra é de tamanho $n = 500$.

Tabela 13 Comparação entre a velocidade de execução (em segundos) do método proposto, utilizando a três combinações de restrições, e o método cobs do R e a relação entre o cobs e a média das três combinações de restrições.

tamanho da amostra	$R_1 + R_2 + R_3$	$R_1 + R_3$	R_3	<i>cobs</i>	relação média
<i>n</i> = 100	3.95	4.19	3.96	85.57	21.18
	4.05	3.94	4.10	80.25	19.91
	4.01	4.06	4.42	81.81	19.62
	4.02	4.19	4.08	79.75	19.45
	4.14	4.13	3.97	79.32	19.44
	4.17	3.91	3.88	79.79	20.04
	4.39	4.06	4.05	90.90	21.79
	4.79	4.82	5.06	90.39	18.48
	5.05	4.94	4.90	91.43	18.43
	4.82	4.84	4.96	93.61	19.18
<i>n</i> =200	8.32	8.16	8.52	113.21	13.57
	8.30	9.50	11.50	140.14	14.33
	7.21	7.11	7.07	104.48	14.79
	7.32	7.25	7.23	103.83	14.28
	7.35	6.98	7.10	104.87	14.66
	7.33	7.14	7.13	105.10	14.46
	7.68	7.53	8.23	114.33	14.64
	8.30	8.84	8.12	116.79	13.87
	8.46	8.31	8.19	113.28	13.62
	8.29	8.39	8.14	116.03	14.03
<i>n</i> =500	25.62	26.30	23.56	177.95	7.07
	23.58	25.08	24.30	178.15	7.32
	31.02	32.02	27.08	221.31	7.31
	28.02	28.98	29.55	198.03	6.59
	27.96	29.07	27.21	192.87	6.87
	27.54	27.92	27.79	199.93	7.20
	27.67	27.71	27.86	204.09	7.35
	27.34	27.55	27.64	200.48	7.29
	27.99	28.04	27.54	199.89	7.17
	27.67	27.86	27.84	198.90	7.16

A Tabela 14 é reproduzida da mesma forma que a tabela 2 da página 649 de Cormier et al. (2014) e nela, são apresentadas as taxas de erro tipo I e o poder do método proposto nesse trabalho (usando as três combinações de restrições), observadas, quando os dados são originados das famílias de cópulas Gumbel, Clayton, Frank, Gaussian ou t com 4 graus de liberdade para $\tau = 0,25$, $\tau = 0,50$, $\tau = 0,75$. Esses valores são comparados respectivamente com outros testes de valores extremos propostos em Kojadinovic e Yan (2010), Du e Nešlehová (2013), Ben Ghorbal et al. (2009), Bucher et al. (2011), Kojadinovic et al. (2011) e Cormier et al. (2014).

Os resultados obtidos na Tabela 14 estão baseados nos valores obtidos para T_N e sugerem que o teste proposto neste trabalho tem nível nominal razoavelmente bom (próximo a 0,05) para todos os valores de TAU de Kendall em estudo. Em contraste, Cormier et al. (2014) mostram que o teste BDV (Bucher et al. 2011), o teste KY (Kojadinovic e Yan, 2010) e o teste KSY (Kojadinovic et al. 2011) são conservadores, especialmente quando $\tau = 0,50$ e $\tau = 0,75$.

Todas as três combinações de restrições utilizadas para verificar se o conjunto de dados bivariados possui dependência de valor extremo apresentaram um bom controle do erro tipo I para todos os valores de TAU de Kendall analisados ($\tau = 0.25$, $\tau = 0.50$, $\tau = 0.75$). Porém, ainda há uma falha quando se trata do controle da taxa de erro tipo II, já que os valores da taxa de rejeição não são satisfatórios para todas as famílias de cópulas analisadas, exceto quando os dados tem origem na família Clayton.

Tabela 14. Proporção de rejeição de H_0 (em %) para alguns testes realizados a nível nominal de 5% utilizando amostras de tamanho 200 a partir de cinco famílias de cópulas.

τ	df	Modelo	$R_1 + R_2 + R_3$	$R_1 + R_3$	R_3	KY	DN	BGN	BDV	KSY	CGN
0.25	4	Gumbel	5.5	7.0	3.5	3.8	5.2	5.4	4.5	5.0	4.7
		Clayton	92.5	92.5	94.5	98.4	96.7	98.0	87.4	94.6	97.7
		Frank	6.5	6.0	6.5	58.3	57.0	38.4	29.1	66.1	18.7
		Gaussiana	12.5	8.5	9.5	36.5	40.3	37.3	16.8	38.7	25.5
		t_4	21.0	19.5	24.5	23.9	19.6	26.2	10.5	26.6	37.7
0.50	7	Gumbel	6.0	4.5	5.0	3.9	5.0	5.1	2.9	4.0	5.4
		Clayton	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Frank	63.0	63.0	62.0	95.7	84.8	59.4	73.0	96.5	87.8
		Gaussiana	38.0	33.0	34.0	61.8	61.7	62.6	23.7	51.0	59.4
		t_4	39.5	35.0	48.0	50.1	45.3	56.0	15.8	52.7	58.6
0.75	11	Gumbel	5.0	5.0	4.5	3.2	5.3	4.9	2.5	2.3	6.2
		Clayton	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Frank	73.5	99.0	100.0	99.9	92.9	58.5	78.3	99.0	98.3
		Gaussiana	13.5	56.5	50.0	66.5	71.1	75.2	8.4	46.7	56.5
		t_4	25.5	51.0	54.0	50.6	55.8	67.8	4.6	69.2	45.8

* As seis últimas colunas da tabela são reproduzidas da mesma forma que a tabela 2 da página 649 de Cormier et al. (2014).

Quanto ao poder, nenhum dos procedimentos isoladamente destaca-se como o mais poderoso e, para $\tau = 0,75$, o teste BDV é, sem dúvidas, o menos poderoso. Os resultados encontrados para o teste proposto são comparáveis com os resultados dos outros testes, encontrados na tabela (KY, DN, BGN, BDV, KSY e CGN).

5 CONCLUSÃO

- A utilização de polinômios de Bernstein convexos com restrições nos extremos se revelou um método eficiente com desempenho comparável ao de outros métodos propostos na literatura.
- A velocidade de execução do método de ajuste baseada em polinômios de Bernstein é substancialmente superior à velocidade de execução do pacote *Cobs*, do software livre R.
- A introdução de duas novas restrições, além da convexidade, para os polinômios de Bernstein se revelaram promissoras.

REFERÊNCIAS

- BEN GHORBAL, N.; GENEST, C.; NEŠLEHOVÁ, J. On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. **Canadian Journal of Statistics**, v. 37, n. 4, p. 534-552, 2009.
- BÜCHER, Axel et al. New estimators of the Pickands dependence function and a test for extreme-value dependence. **The Annals of Statistics**, v. 39, n. 4, p. 1963-2006, 2011.
- CAPÉRAÀ, P.; FOUGÈRES, A. L.; GENEST, C. A nonparametric estimation procedure for bivariate extreme value copulas. **Biometrika**, v. 84, n. 3, p. 567-577, 1997.
- CEBRIAN, A., DENUIT, M., LAMBERT, P. Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database. **Belgian Actuarial Journal**, v. 3, n. 1, p. 33-41, 2003.
- CHAK, P. M.; MADRAS, N.; SMITH, B. Semi-nonparametric estimation with Bernstein polynomials. **Economics Letters**, v. 89, n. 2, p. 153-156, 2005.
- CHANG, I.; HSIUNG, C. A., WU, Y. J.; YANG, C. C. Bayesian survival analysis using Bernstein polynomials. **Scandinavian journal of statistics**, v. 32, n. 3, p. 447-466, 2005.
- CHANG, I. S.; CHIEN, L. C., HSIUNG, C. A, WEN, C.C., WU, Y. J. Shape restricted regression with random Bernstein polynomials. In: **Complex datasets and inverse problems**. Institute of Mathematical Statistics, 2007. p. 187-202.
- CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. **Copula methods in finance**. John Wiley & Sons, 2004.
- COLES, S; HEFFERNAN, J; TAWN, J. Dependence measures for extreme value analyses. **Extremes**, v. 2, n. 4, p. 339-365, 1999.
- CORMIER, E.; GENEST, C.; NEŠLEHOVÁ, J. G. Using B-splines for nonparametric inference on bivariate extreme-value copulas. **Extremes**, v. 17, n. 4, p. 633-659, 2014.
- DU, Y.; NEŠLEHOVÁ, J. A moment-based test for extreme-value dependence. **Metrika**, v. 76, n. 5, p. 673-695, 2013.

GENEST, C.; FAVRE, A-C. Everything you always wanted to know about copula modeling but were afraid to ask. **Journal of hydrologic engineering**, v. 12, n. 4, p. 347-368, 2007.

GENEST, C.; SEGERS, J. Rank-based inference for bivariate extreme-value copulas. **The Annals of Statistics**, v. 37, n. 5B, p. 2990-3022, 2009.

GUDENDORF, G.; SEGERS, J. Nonparametric estimation of an extreme-value copula in arbitrary dimensions. **Journal of multivariate analysis**, v. 102, n. 1, p. 37-47, 2011.

GUDENDORF, G.; SEGERS, J. Nonparametric estimation of multivariate extreme-value copulas. **Journal of Statistical Planning and Inference**, v. 142, n. 12, p. 3073-3085, 2012.

HALL, P; TAJVIDI, N. Distribution and dependence-function estimation for bivariate extreme-value distributions. **Bernoulli**, v. 6, n. 5, p. 835-844, 2000.

HE, X.; NG, P. COBS: qualitatively constrained smoothing via linear programming. **Computational Statistics**, v. 14, n. 3, p. 315-338, 1999.

JOE, H. **Multivariate models and multivariate dependence concepts**. CRC Press, 1997.

KOJADINOVIC, I.; YAN, J. Comparison of three semiparametric methods for estimating dependence parameters in copula models. **Insurance: Mathematics and Economics**, v. 47, n. 1, p. 52-63, 2010.

KOJADINOVIC, I.; SEGERS, J.; YAN, J. Large-sample tests of extreme-value dependence for multivariate copulas. **Canadian Journal of Statistics**, v. 39, n. 4, p. 703-720, 2011.

KOTZ, S.; NADARAJAH, S. **Extreme value distributions**. Imperial College Press, London, 2000.

MARCON, G.; PADOAN, S.; NAVEAU, P., MULIERE, P. Multivariate Nonparametric Estimation of the Pickands Dependence Function using Bernstein Polynomials. **arXiv preprint arXiv:1405.5228**, 2014.

MCKAY, C. S.; GHOSH, S. K. A variable selection approach to monotonic regression with Bernstein polynomials. **Journal of Applied Statistics**, v. 38, n. 5, p. 961-976, 2011.

NELSEN, Roger B. Properties and applications of copulas: A brief survey.
In: **Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance**, (Dhaene, J., Kolev, N., Morettin, PA (Eds.)),
University Press USP: Sao Paulo. 2003. p. 10-28.

NELSEN, R. B. **An introduction to copulas**. Springer Science & Business Media, 2013.

PADOAN, Simone A. Extreme value analysis. **Encyclopedia of Environmetrics**, 2013.

PETRONE, S. Random bernstein polynomials. **Scandinavian Journal of Statistics**, v. 26, n. 3, p. 373-393, 1999.

PICKANDS, J. Multivariate extreme value distributions. In: **Proceedings 43rd Session International Statistical Institute**. 1981. p. 859-878.

STADTMÜLLER, U. Asymptotic properties of nonparametric curve estimates. **Periodica Mathematica Hungarica**, v. 17, n. 2, p. 83-108, 1986.

WANG, J.; GHOSH, S. K. Shape restricted nonparametric regression with Bernstein polynomials. **Computational Statistics & Data Analysis**, v. 56, n. 9, p. 2729-2741, 2012.

ZHANG, D; WELLS, M. T.; PENG, L. Nonparametric estimation of the dependence function for a multivariate extreme value distribution. **Journal of Multivariate Analysis**, v. 99, n. 4, p. 577-588, 2008.