

AJUSTE DE CÓPULAS BIVARIADAS VIA MARGINAL NA DIAGONAL

Danielle Gonçalves de Oliveira PRADO¹
Devanil Jaques de SOUZA²
Lucas Monteiro CHAVES³

- RESUMO: É notório o grande número de estudos recentes relacionados à qualidade de ajuste para cópulas e inegável sua relevância em diversas áreas, principalmente, na economia, hidrologia e em dados biológicos. Eleger uma família de cópulas para ajustar um determinado conjunto de dados é uma tarefa importante e complexa. Porém, ainda não existe um método considerado mais adequado para tal. Nos últimos anos, vários métodos têm sido propostos. O objetivo principal deste trabalho é propor dois novos testes para verificar a qualidade de ajuste de cópulas para dados bivariados via marginal na diagonal principal. A fim de verificar a adequabilidade dos testes, foram utilizadas as 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 na diagonal principal através de 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. A conclusão foi feita através da verificação do controle das taxas de erros tipo I e tipo II.
- PALAVRAS-CHAVE: Qualidade de ajuste; cópulas; marginal na diagonal; tau de Kendall.

1 Introdução

É crescente o número de pesquisas relacionadas à qualidade do ajuste de cópulas bem como o estudo de cópulas de valores extremos. Balakrishnan e Lai (2009) mencionam aplicações de cópulas na agricultura, biologia, engenharia, economia e finanças, meio-ambiente, genética, medicina, psicologia, controle de qualidade, confiabilidade e análise de sobrevivência, sociologia, ciências físicas e tecnologia. 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

¹ Universidade Federal Tecnológica Federal do Paraná - UFTPR, Departamento Acadêmico de Matemática, CEP: 86812-460, Apucarana, PR, Brasil, E-mail: danigoprado@hotmail.com

² Universidade Federal de Lavras - UFLA, Departamento de Estatística, Caixa Postal: 3037, CEP: 37200-000, Lavras, MG, Brasil. E-mail: devaniljaques@des.ufla.br

³ Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas, Caixa Postal: 3037, CEP: 37200-000, Lavras, MG, Brasil. E-mail: lucas@dex.ufla.br

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.

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.

O objetivo principal deste trabalho é propor dois novos métodos de verificação da qualidade do ajuste de cópulas para dados bivariados baseados nas marginais obtidas nas diagonais principal e secundária. Um deles utilizando uma estatística com distribuição Qui-quadrado e outro em um teste de hipótese cujo critério de decisão é baseado em intervalos de confiança para assimetria e curtose. Os novos testes foram aplicados para dados provenientes de cópulas das famílias: Clayton, Gumbel, Normal e Frank.

2 Metodologia

A diagonal principal da cópula $C(u, v)$ é denotada por w e a diagonal secundária por z . A matriz de rotação utilizada é dada por:

$$\begin{aligned} \begin{bmatrix} w \\ z \end{bmatrix} &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \\ \Rightarrow \quad u &= \frac{\sqrt{2}}{2}w - \frac{\sqrt{2}}{2}z \\ v &= \frac{\sqrt{2}}{2}w + \frac{\sqrt{2}}{2}z \end{aligned} \quad (1)$$

Na Figura 1 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 .

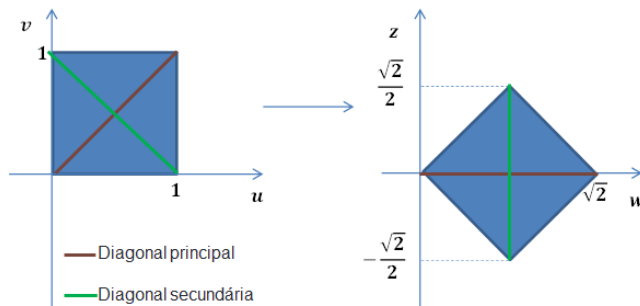


Figura 1 - Rotação do suporte da cópula.

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

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

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

$$\begin{aligned} f_w(w) &= \left[\int_{-w}^w dz \right] I_{\left(0, \frac{\sqrt{2}}{2}\right)}(w) + \left[\int_{-\sqrt{2}+w}^{\sqrt{2}-w} dz \right] I_{\left(\frac{\sqrt{2}}{2}, \sqrt{2}\right)}(w) \\ &= 2w I_{\left(0, \frac{\sqrt{2}}{2}\right)}(w) + 2(\sqrt{2} - w) I_{\left(\frac{\sqrt{2}}{2}, \sqrt{2}\right)}(w) \end{aligned} \quad (3)$$

$$\begin{aligned} f_z(z) &= \left[\int_{-z}^{z+\sqrt{2}} dw \right] I_{\left(-\frac{\sqrt{2}}{2}, 0\right)}(z) + \left[\int_z^{\sqrt{2}-z} dw \right] I_{\left(0, \frac{\sqrt{2}}{2}\right)}(z) \\ &= (\sqrt{2} + 2z) I_{\left(-\frac{\sqrt{2}}{2}, 0\right)}(z) + (\sqrt{2} - 2z) I_{\left(0, \frac{\sqrt{2}}{2}\right)}(z) \end{aligned} \quad (4)$$

Utilizando o *software* R (R CORE TEAM, 2016) 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 2, que são exatamente os gráficos das marginais $f_w(w)$ e $f_z(z)$, respectivamente.

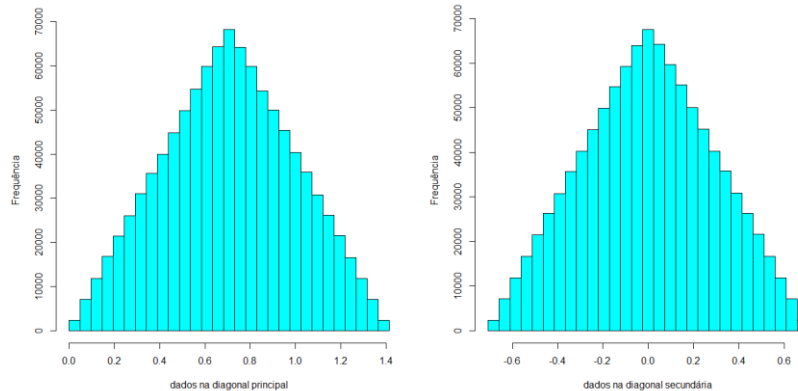


Figura 2 - Gráfico dos dados de uma amostra tamanho $n = 1.000.000$ da cópula da independência plotados nas diagonais principal e secundária.

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 suas 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. Tais famílias foram escolhidas devido à importância dessas cópulas, pelo fato de serem famílias uniparamétricas, o que permite a utilização do Tau de Kendall amostral na estimação do valor do parâmetro, e também pelo fato de serem as mais utilizadas em outros trabalhos similares. A cópula de Clayton pertence à família de cópulas Arquimedianas e se caracteriza por captar dependência na cauda inferior; a cópula Gumbel também é uma Arquimediana de valor extremo, uma vez que capta dependência na cauda superior; A Gaussiana é capaz de reproduzir a mesma estrutura de dependência encontrada em normais bivariadas; e a cópula de Frank, mais uma cópula Arquimediana, apresenta simetria nas caudas superior e inferior. Os cálculos utilizados nos dois testes foram feitos com o auxílio de um pacote do *software* R (R CORE TEAM, 2016) 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 $[0,1] \times [0,1]$.

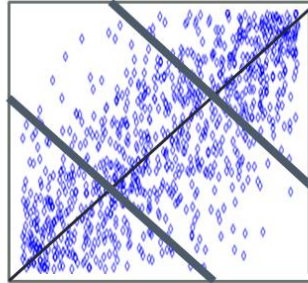


Figura 3 - Divisão das três regiões definidas por retas ortogonais à diagonal.

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.

2.1 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 e 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 calculam-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.

2.2 Teste 2: Intervalo de confiança

Este teste propõe comparar 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 à diagonal principal e os coeficientes de assimetria ($\sqrt{b_1}$) e curtose (b_2):

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

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, rotacionam-se os dados, calculam-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 calculam-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.

3 Resultados

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

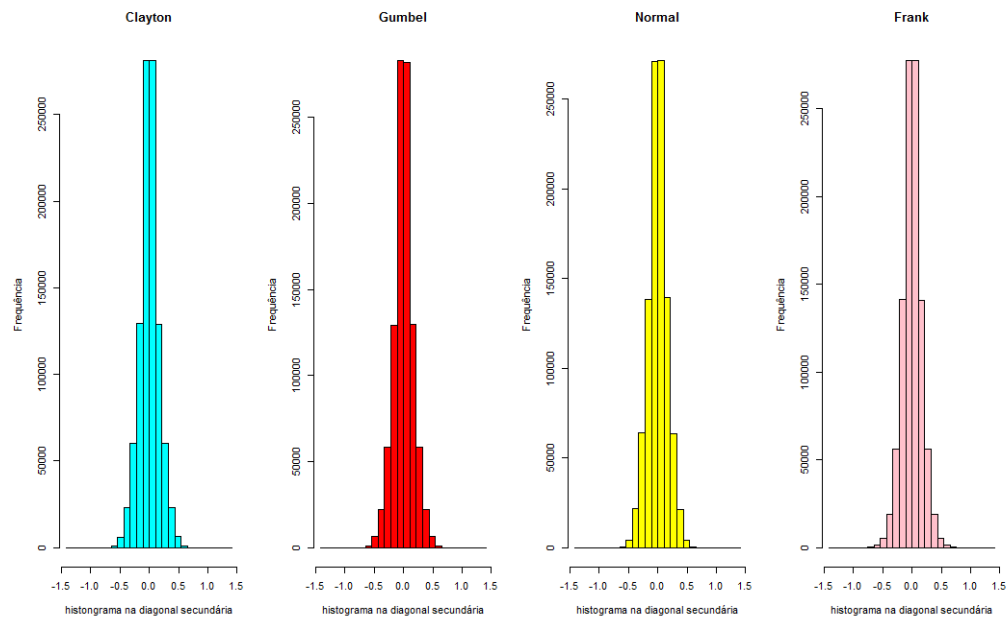


Figura 4 - Família de cópulas na diagonal secundária.

Na Figura 5 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.

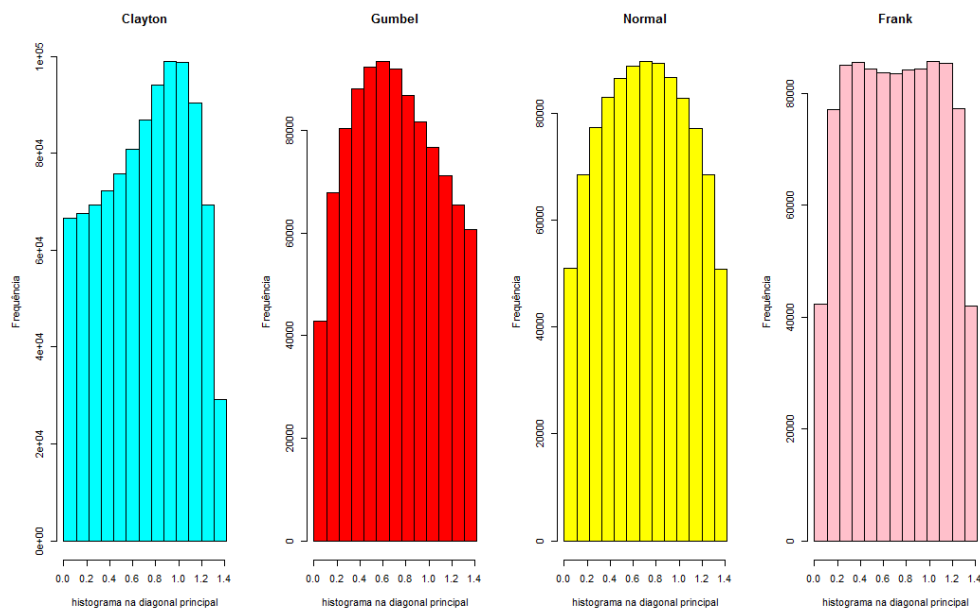


Figura 5 - Família de cópulas na diagonal principal.

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

3.1 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 1.

Tabela 1 - 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 : Clayton	H_a : Gumbel	H_a : Normal	H_a : 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 à 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 contrapartida 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 2.

Tabela 2 - 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 : Clayton	H_0 : Gumbel	H_a : Normal	H_a : 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, os valores apresentados na tabela também indicam um bom controle do teste para o erro tipo I. Já o controle do erro tipo II só possui um bom desempenho quando é comparado à família de cópulas de Clayton e, especificamente, para os valores de $\tau = 0,4$; $\tau = 0,5$; $\tau = 0,6$.

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 3.

Tabela 3 - 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 : Clayton	H_a : Gumbel	H_0 : Normal	H_a : 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
$\tau = 0,5$	Novo	0,0740	0,6910	0,1140	0,5670
	PB	0,0714	0,3151	0,8806	0,0234
	MB	0,0004	0,0004	0,6338	0,0004
$\tau = 0,6$	Novo	0,1320	0,7510	0,1220	0,5870
	PB	0,1053	0,4930	0,6388	0,1243
	MB	0,0004	0,0054	0,4580	0,0044
$\tau = 0,8$	Novo	0,5400	0,8690	0,1120	0,7980
	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 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 Frank a nível de 5% de significância

$n = 600$		H_a : Clayton	H_a : Gumbel	H_a : Normal	H_0 : Frank
$\tau = 0,2$	Novo	0,7630	0,8980	0,9010	0,0480
	PB	0,1763	0,3511	0,4140	0,2312
	MB	0,0004	0,0004	0,0004	0,6788
$\tau = 0,4$	Novo	0,1120	0,4690	0,7790	0,0570
	PB	0,6948	0,8756	0,5369	0,7667
	MB	0,0004	0,0004	0,0484	0,1253
$\tau = 0,5$	Novo	0,1210	0,4620	0,6920	0,0490
	PB	0,6888	0,9395	0,7537	0,5799
	MB	0,0004	0,0034	0,1123	0,4270
$\tau = 0,6$	Novo	0,2570	0,5780	0,7180	0,0510
	PB	0,1933	0,6918	0,5659	0,7577
	MB	0,0004	0,0014	0,0024	0,4440
$\tau = 0,8$	Novo	0,7580	0,8830	0,8840	0,0600
	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.

3.2 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 5.

Tabela 5 - 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 : Clayton	H_a : Gumbel	H_a : Normal	H_a : 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 ao 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 contrapartida 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 6.

Tabela 6 - 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 : Clayton	H_0 : Gumbel	H_a : Normal	H_a : 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
$\tau = 0,8$	Novo	0,6955	0,0978	0,8866	0,7962
	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 7.

Tabela 7 - 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 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_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 testes PB e MP.

Conclusões

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 (os dados se aproximam da independência) ou τ próximo de 1 (os dados

se aproximam da dependência positiva perfeita) as cópulas se aproximam, respectivamente, da cópula da independência $C_i(u, v) = uv$ ou da cópula do mínimo $C_1(u, v) = \min\{u, v\}$. E, portanto, em qualquer dois casos cópulas ajustadas ficam extremamente semelhantes ou à cópula da independência ou à cópula do mínimo, dificultando sua 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.

PRADO, D. G. O.; CHAVES, L. M.; SOUZA, D. J. Fit of bivariate copulas using diagonal marginal. *Rev. Bras. Biom.*, Lavras, v.35, n.3, p.497-514, 2017.

- **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 that best suit 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 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.*

Referências

BALAKRISHNAN, N.; LAI, C. D. *Continuous Bivariate Distributions*, 2.ed., New York: Springer, 2009.

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.

CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. *Copula methods in finance*. New York: John Wiley & Sons, 2004.

COLES, S; HEFFERNAN, J; TAWN, J. Dependence measures for extreme value analyses. *Extremes*, v.2, n.4, p.339-365, 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.; 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*. p.1096-1127, 2008
- 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*. London: CRC Press, 1997.
- KOJADINOVIC, I.; YAN, J.; HOLMES, M. Fast large-sample goodness-of-fit tests for copulas. *Statistic Sinica*, p.841-871, 2011.
- NELSEN, R. B. *An introduction to copulas*. New York: Springer Science & Business Media, 2013.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2016.

Recebido em 22.03.2016

Aprovado após revisão em 06.12.2016