

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTATISTICA MATEMATICA E COMPUTAÇÃO
CIENTIFICA
DEPARTAMENTO DE ESTATISTICA

Dissertação de Mestrado

Teste Gráfico para o Ajuste de Cópulas Arquimedianas usando
variáveis BIPIT: Um Estudo de Simulação.

Marta Cristina Colozza Bianchi

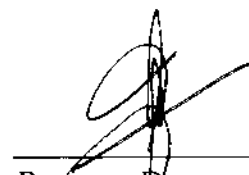
Orientadora: Verónica Andrea González-López

Campinas - SP 2008

TESTE GRÁFICO PARA O AJUSTE DE CÓPULAS ARQUIMEDIANAS USANDO
VARIÁVEIS BIPIT: UM ESTUDO DE SIMULAÇÃO

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Marta Cristina Colozza Bianchi e aprovada pela comissão julgadora.

Campinas, 29 de Julho de 2008



Profa. Dra.: Verónica Andrea
González-Lopez
Orientadora

Banca Examinadora:

1. Profa. Dra. Verónica Andrea González-López (IMECC-UNICAMP)
2. Prof. Dr. Sergio Wechsler (IME-USP)
3. Prof. Dr. Jesus Enrique Garcia (IMECC-UNICAMP)

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em ESTATÍSTICA.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues**

Bianchi, Marta Cristina Colozza

B47t Teste gráfico para o ajuste de cópulas Arquimedianas usando
variáveis BIPIT: um estudo de simulação / Marta Cristina Colozza
Bianchi -- Campinas, [S.P. :s.n.], 2008.

Orientador : Verónica Andrea González-López

Dissertação (mestrado) - Universidade Estadual de Campinas,
Instituto de Matemática, Estatística e Computação Científica.

1. Cópulas. 2. BIPIT. 3. Q-Qplot. 4. Kendall Plot. 5. Teste de
hipótese. 6. Tau de Kendall. I. González-López, Verónica Andrea. II.
Universidade Estadual de Campinas. Instituto de Matemática, Estatística
e Computação Científica. III. Título.

Título em inglês: Test chart for the adjustment Archimedean copulas using variables BIPIT: a study of simulation.

Palavras-chave em inglês (Keywords): 1. Copulas. 2. BIPIT. 3. Q-Qplot. 4. Kendall Plot. 5. Statistical hypothesis testing. 6. Kendall's Tau.

Área de concentração: Teoria de Cópulas

Titulação: Mestre em Estatística

Banca examinadora:

Profa. Dra. Verónica Andrea González-López (IMECC-UNICAMP)

Prof. Dr. Sergio Wechsler (IME-USP)

Prof. Dr. Jesus Enrique Garcia (IMECC-UNICAMP)

Data da defesa: 07/07/2008

Programa de pós-graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 07 de julho de 2008 e aprovada

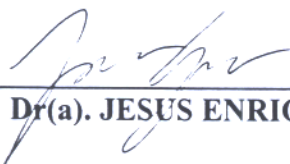
Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). VERÓNICA ANDREA GONZÁLEZ-LÓPEZ



Prof(a). Dr(a). SÉRGIO WECHSLER



Prof(a). Dr(a). JESUS ENRIQUE GARCIA

Agradecimentos

Agradeço a Deus e a todos que contribuíram, direta ou indiretamente para a realização deste trabalho.

À minha orientadora Verónica Andrea González-López pelas muitas horas de trabalho e sua extrema dedicação e compreensão.

À minha família, especialmente, meus pais.

À todos os meus amigos, especialmente, José Clelto Barros Gomes e Márcio Luis Lanfredi Viola, grandes companheiros, pelas grandes conversas que tivemos e ainda temos sobre nossos estudos e sobre a vida.

À CAPES pelo suporte financeiro.

Resumo

A crescente utilização de cópulas para modelagem de dependência em dados multivariados leva ao estudo de metodologias para o ajuste de cópulas. Este estudo é recente, assim como a plena utilização da teoria de cópulas para modelagem padrão. Grande parte das metodologias existentes ainda encontra-se em fase de estudo e somente alguns métodos foram validados recentemente. Há a necessidade de mecanismos de fácil acesso a detecção de estruturas de dependência ainda escassos na literatura. Nesta dissertação, é apresentado um método gráfico para o ajuste de cópulas, adaptado do QQplot, denominado Kendall Plot. Este método torna-se mais completo que o QQplot ao se postular a adição de bandas de confiança ao gráfico Kendall Plot, que permitem tomar uma decisão em relação a uma estrutura de dependência fixa, expressa por uma cópula, a ser testada para a amostra disponível. A redução de dimensão dos dados a uma variável unidimensional denominada BIPIT, que carrega informação a respeito da estrutura de dependência dos dados, permite a utilização da adaptação do QQplot com o fim de se testar estruturas de dependência.

Palavras-Chave: Cópulas, BIPIT, Q-Qplot, Kendall Plot, Teste de hipótese, Tau de Kendall.

Abstract

The growing utilization of copulas to the dependency fitting of multi-variated data leads to the study of methodologies for copulas fitting. This study is recent, such as the complete utilization of the theory of copulas to standard fitting. Many of the existing methodologies are still in studies and only some have been recently validated. There is a need for easy-access mechanisms to detect dependency structures still missing in the statistical literature. It is presented in this dissertation a graphic method to the copulas fitting adapted from QQplot denominated Kendall Plot. This method is more complete than the QQplot due to the addition of confidence bands to the Kendall Plot graphic that allows the researcher to make a decision related to a fixed dependency structure, expressed by a copula, to be tested to the available sample. The reduction of the data dimension to a one-dimensional random variable, called BIPIT, which carries information about dependency data structure, allows the utilization of the QQplot adaptation for testing dependency structures.

Key Words: Copulas, BIPIT, Q-Qplot, Kendall Plot, Statistical hypothesis testing, Kendall's Tau.

Sumário

1	Introdução	1
1.1	Introdução a Teoria de Cópulas	3
1.1.1	Cópula e Teorema de Sklar	4
1.1.2	Cópulas Arquimedianas	14
2	Variável BIPIT	21
2.1	Propriedades da função K	28
3	Tau de Kendall	33
3.1	Concordância	33
4	Kendall Plot	45
4.1	Construção do Kendall Plot	46
4.1.1	QQplot	46
4.1.2	Kendall Plot	47
4.2	Resultados e Fundamentos	50

5	Testes de Aderência para Cópulas	61
5.1	Teste Gráfico - Intervalos de Confiança para Quantis	62
5.1.1	Construção de Intervalos de Confiança para Quantis	66
5.1.2	Teste gráfico para o Kendall Plot: uma aplicação para intervalos de confiança de quantis	67
5.2	Poder do Teste	75
5.2.1	Análise dos resultados	75
5.2.2	Intersecção de $n - 2$ testes	76
5.2.3	Intersecção de 20 testes	87
5.2.4	Conclusão	89
.1	Resultado sobre Esperança de uma v.a. aleatória	91
.2	Convergência do quantil amostral ao quantil teórico de uma dis- tribuição	92
.3	Expressão para $W_{i:n,K_0} \equiv E(H_{(i)}^0)$	94
.4	Integração Numérica	95
.5	Teste bootstrap paramétrico	98
.6	Correção de Bonferroni	101
.7	Teste gráfico Kendall Plot para o parâmetro θ de uma cópula Arquimediana	102
.8	Algoritmos	108
.8.1	Cálculo de $W_{i:n,K_0} \equiv E(H_{(i)}^0)$	109

.8.2 Teste gráfico Kendall Plot com bandas de confiança 115

Lista de Figuras

1.1	<i>Scatterplots</i> : Dados exponenciais independentes (primeira figura); dados exponenciais independentes transformados pelas acumuladas (segunda figura).	5
1.2	Gerador ϕ não estrito e pseudo-inversa $\phi^{[-1]}$ para a cópula Arquimediana W .	18
1.3	Gerador ϕ estrito e pseudo-inversa $\phi^{[-1]}$ para a cópula Arquimediana Π .	19
2.1	QQplot entre a amostra $H_i = \mathbb{H}(x_i, y_i)$ da BIPIT H associada distribuição normal bivariada do Exemplo 2.1 e uma amostra $\{U_1, \dots, U_{100}\}$ com distribuição $U(0, 1)$.	22
2.2	QQplot das amostras das BIPITS H e H_1 : $\{H_1, \dots, H_n\}$ e $\{H_{11}, \dots, H_{1n}\}$ respectivamente, $n = 100$, onde $H_i = \mathbb{H}(u_i, v_i) = u_i v_i$ sendo u_i, v_i amostras de $U(0, 1)$ e $H_{1i} = \mathbb{H}(x_{1i}, y_{1i}) = F_1(x_{1i})G_1(y_{1i})$ sendo x_{1i} com distribuição $exp(2)$ e y_{1i} com distribuição $exp(10)$.	23
2.3	Funções distribuição de Kendall das cópulas M (linha cheia preta), W (linha cheia cinza) e família de Gumbel para $\theta = \{1, 3, 7\}$ (primeira figura); família de Clayton para $\theta = \{1, 3, 4\}$ (segunda figura).	31
2.4	Funções $\lambda(v)$, $v \in \mathbf{I}$, das cópulas M (linha cheia preta), W (linha cheia cinza) e família de Gumbel para $\theta = \{1, 3, 7\}$ (primeira figura); família de Clayton para $\theta = \{1, 3, 4\}$.	32

3.1	QQplots referentes Exemplo 3.5: $H_{0i} = \Pi(u_i, v_i)$, $H_{1i} = C_{-0.5}(u_i, v_i)$, $H_{2i} = C_2(u_i, v_i)$ com C_θ cópula de Clayton com parâmetro θ (A linha com ponto e tracejado é a diagonal principal do gráfico).	43
4.1	Gráficos da inversa da função distribuição $K^{-1}(p)$ vs. $\widehat{H}_{(\lceil np \rceil)} \equiv K_n^{-1}(p)$ proveniente de $n = 100$ pseudo-observações da cópula 4.2.12 (primeira figura); e $W_{\lceil np \rceil:n,K} = E(H_{(\lceil np \rceil)})$, $n = 100$, sob hipótese da mesma cópula (segunda figura).	55
4.2	Kendall Plot sob hipótese nula de independência para a amostra $\{(x_1, y_1), \dots, (x_{100}, y_{100})\}$ sendo x_i com distribuição $exp(2)$ e y_i com distribuição $exp(10)$ com $\{x_i\}$ gerados de forma independente de $\{y_i\}$	56
4.3	Kendall Plot sob hipótese nula de independência de $n = 100$ pares de observações do vetor aleatório (X, Y) com $Y = 1 - X$	57
4.4	Kendall Plot sob hipótese nula de independência, de $n = 100$ pares de observações do vetor aleatório (X, Y) com $Y = X$	58
4.5	Kendall Plot sob hipótese nula de cópula de Clayton $\theta = 2$ ($W_{i:n,K_0}$ referente a Clayton com $\theta = 2$) de amostras aleatórias de tamanho $n = 100$ associadas a cópula de Clayton com os respectivos parâmetros $\theta = \{0.5, 2, 5\}$	59
5.1	Kendall Plots sob hipótese nula de cópula 4.2.12 com $\theta = 3$, nível de significância $\alpha = 0.05$, de amostra aleatória $n = 100$ associada a mesma cópula $\theta = 3$. Primeira figura: $(\xi_p^0, \widehat{H}_{(\lceil np \rceil)})$. Segunda figura: $(\widehat{\xi}_p^0, \widehat{H}_{(\lceil np \rceil)})$	71
5.2	Kendall Plot $(\widehat{\xi}_p^0, \widehat{H}_{(\lceil np \rceil)})$ sob hipótese nula de cópula de independência, nível de significância $\alpha = 0.05$, de amostra aleatória $n = 100$ associada a cópula 4.2.12 com $\theta = 3$	72
5.3	Kendall Plot $(\widehat{\xi}_p^0, \widehat{H}_{(\lceil np \rceil)})$ sob hipótese nula de cópula de Clayton $\theta = 20$, nível de significância $\alpha = 0.05$, de amostras aleatórias $n = 100$ (primeira figura) e $n = 500$ (segunda figura) associadas a cópula de Clayton $\theta = 6$	73

5.4	Gráficos de amostras tamanho $n = 100$ com cópula de Clayton com $\theta = 6$ e $\theta = 20$ (gráficos acima da esquerda a direita). Gráficos de amostras tamanho $n = 500$ com cópula de Clayton com $\theta = 6$ e $\theta = 20$ (gráficos abaixo da esquerda a direita).	74
5.5	Amostras $n = 500$ da copula 4.2.12 para $\theta = \{1, 1.5, 2, 2.5\}$ associado respectivamente a $\tau \approx \{0.3, 0.5, 0.6, 0.7\}$ (da esquerda para direita).	77
5.6	Amostras $n = 500$ das copulas de Clayton e Frank com nível de associação $\tau \approx 0.7$ (da esquerda para direita).	78
5.7	Amostras $n = 500$ das copulas de Gumbel e 4.2.12 com nível de associação $\tau \approx 0.7$ (da esquerda para direita).	79
5.8	Amostras $n = 500$ das copulas de Clayton e 4.2.12 com nível de associação $\tau \approx 0.3$ (da esquerda para direita).	80
5.9	Amostras $n = 500$ das copulas de Frank e Gumbel com nível de associação $\tau \approx 0.3$ (da esquerda para direita).	81
5.10	Amostras $n = 500$ de copula de Gumbel $\theta = \{3.5, 5, 6.66, 10, 20\}$ (de cima para baixo da esquerda para direita).	85
5.11	Amostras $n = 500$ de copula de Clayton $\theta = \{4.66, 6, 8, 12, 18\}$ (de cima para baixo da esquerda para direita).	86
12	Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 30$	96
13	Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 70$	96
14	Integrando $f(x)$ para cópula da independência com $n = 400$ e $i = 70$	97
15	Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 300$	97
16	Função poder sob \mathbf{H}_0 : cópula de Gumbel $\theta = 1.5$ ($\tau \approx 0.33$).	104
17	Função poder sob \mathbf{H}_0 : cópula de Gumbel $\theta = 3.5$ ($\tau \approx 0.7$).	105
18	Função poder sob \mathbf{H}_0 : cópula de Clayton $\theta = 1$ ($\tau \approx 0.3$).	106

19 Função poder sob \mathbf{H}_0 : cópula de Clayton $\theta = 4.66$ ($\tau \approx 0.7$). 107

Lista de Tabelas

1.1	Algumas famílias paramétricas de cópulas Arquimedianas com seus geradores e espaços paramétricos (*na cópula de Clayton o gerador é estrito se $\theta \geq 0$, caso contrário é não estrito).	20
2.1	Cópulas Arquimedianas da Tabela 1.1 com os respectivos geradores ϕ e funções distribuição K .	28
3.1	Medida de associação τ de Kendall expressa em função do parâmetro θ para algumas famílias de cópulas Arquimedianas.	36
5.1	Algumas características das estruturas de dependência de algumas cópulas Arquimedianas (*a manifestação do afunilamento na cauda superior depende do nível de associação τ , veja Figura 5.5)	76
5.2	Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 cópula de Gumbel com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes cópulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.	82
5.3	Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 cópula de Gumbel com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes cópulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$.	82
5.4	Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 cópula de Clayton com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes cópulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.	82

5.5	Intersecção de $n - 2$ testes: percentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$	83
5.6	Intersecção de 20 testes: percentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$	87
5.7	Intersecção de 20 testes: percentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$	87
5.8	Intersecção de 20 testes: percentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$	88
5.9	Intersecção de 20 testes: percentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$	88
10	Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Gumbel $\theta = 1.5$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Gumbel com diferentes níveis de associação.	103
11	Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Gumbel $\theta = 3.5$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Gumbel com diferentes níveis de associação.	103
12	Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Clayton $\theta = 1$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Clayton com diferentes níveis de associação.	108
13	Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Clayton $\theta = 4.66$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Clayton com diferentes níveis de associação.	108

Terminologia

v.a. - variável aleatória

f.d.a. - função distribuição acumulada

IC - intervalo de confiança

BIPIT - bivariate probability integral transformation

PIT - probability integral transformation

$U(0, 1)$ - distribuição uniforme em $[0, 1]$

\mathbb{I} - função indicadora

\mathbf{I} - intervalo $[0, 1]$

\prec_{st} - ordem estocástica ordinária

\prec_k - ordem estocástica de Kendall

i.i.d. - independente e indenticamente distribuído

$\bar{\mathbb{R}}$ - intervalo $[-\infty, \infty]$

Im - conjunto imagem de uma função

ϕ, ϕ_θ - gerador de uma cópula Arquimediana

Ω - conjunto das funções $\phi : \mathbf{I} \rightarrow [0, \infty]$ contínuas, estritamente decrescentes, convexas com $\phi(1) = 0$

K, K_C, K_θ - função distribuição acumulada associada a cópula C com parâmetro θ .

$\tau, \tau_{X,Y}, \tau_C$ - tau de Kendall associado ao vetor aleatório (X, Y) com cópula C .

λ, λ_C - razão entre o gerador ϕ da cópula Arquimediana C e sua derivada.

$\kappa, \kappa_{X,Y}, \kappa_C$ - medida de concordância associada ao vetor aleatório (X, Y) com cópula C .

\mathbb{C} - família paramétrica de cópulas

Θ - espaço paramétrico de uma família paramétrica \mathbb{C} de cópulas.

F, G - função distribuição acumulada unidimensional.

F_n - função distribuição empírica unidimensional

\mathbb{H} - função distribuição acumulada multidimensional (em particular bidimensional).

\mathbb{H}_n - função distribuição empírica multidimensional (em particular bidimensional).

H - v.a. BIPIT

\hat{H} - v.a. BIPIT empírica

$X_{(i)}$ - estatística de ordem

$\lceil n \rceil$ - menor inteiro maior ou igual a n .

$O(g(x))$ - Sejam f e g funções, $f(x) = O(g(x))$ se $\lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} \leq c$ com c constante.

$o(g(x))$ - Sejam f e g funções, $f(x) = o(g(x))$ se $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0$.

W - cópula minimal ou limite inferior de Frechét: representa a perfeita dependência negativa entre variáveis.

M - cópula maximal ou limite superior de Frechét: representa a perfeita dependência positiva entre variáveis.

$|\mathcal{I}|$ - quantidade de elementos do conjunto \mathcal{I} .

Capítulo 1

Introdução

Quando se fala em modelagem de dependência, hoje em dia um dos primeiros temas a ser levado em consideração é a recente teoria de cópulas discutida por Joe (1997) [28] e Nelsen (1999) [30]. Esta teoria se torna atrativa devido às cópulas abrangerem um grande leque de estruturas de dependência e conseguirem modelar completamente a estrutura de dependência dos dados. No Capítulo 1 as primeiras seções são dedicadas à exposição de uma introdução para a teoria de cópulas. Por se tratar de uma teoria recente, o estudo de metodologias formais para o ajuste de cópulas ainda é considerado um problema desafiador. Autores como Genest, Quessy & Rémillard (2006) [20], Dobrić & Schmid (2007) [7], Genest, Rémillard & Beaudoin (2008) [23] e Genest & Rémillard (2008) [22] propoem estatísticas para realização de um procedimento bootstrap paramétrico para modelar os dados a famílias paramétricas de cópulas. A validade deste procedimento é provada por Genest & Rémillard (2008) [22] para duas das estatísticas presentes na literatura. Há a necessidade de mecanismos de fácil acesso a detecção de estruturas de dependência ainda escassos na literatura.

O objetivo central desta dissertação é postular um mecanismo gráfico para modelagem de dependência, em particular, modelagem de cópulas munido de bandas de confiança. O gráfico Kendall Plot, uma adaptação do QQplot, proposto inicialmente por Genest & Boies (2003) [19] como mecanismo para detectar independência é usado para este fim sendo apresentado no Capítulo 4. A estrutura de dependência dos dados multivariados é condensada em uma v.a. unidimensional denominada BIPIT H , detalhada no Capítulo 2, que nada mais é que os dados transformados pela função distribuição acumulada. Sklar(1959) [38] mostra que se as marginais dos dados multivariados são contínuas, então

existe uma única cópula C relacionada a função distribuição dos dados por

$$\mathbb{H}(X_1, \dots, X_n) = C(F_1(X_1), \dots, F_n(X_n)) = C(U_1, \dots, U_n)$$

sendo U_1, \dots, U_n com distribuição $U(0, 1)$. Logo a v.a. BIPIT é a própria cópula. Como a v.a. unidimensional BIPIT guarda informação a respeito da estrutura de dependência dos dados ela é usada para construção do gráfico Kendall Plot que permite testar esta estrutura de dependência.

Como o QQplot, o Kendall Plot constitui um meio auxiliar para a realização do teste, necessitando de uma metodologia formal de ajuste para validar seu resultado. Porém, conforme apresenta o Capítulo 5, o Kendall Plot torna-se uma ferramenta completa para ajuste dos dados visto que permite aproveitar a estrutura da v.a. BIPIT para definir bandas de confiança que permitem tomar uma decisão em relação a uma estrutura de dependência fixa, expressada por uma cópula, a ser testada para a amostra disponível.

O Kendall Plot juntamente com a banda de confiança testam se a BIPIT empírica H dos dados possui distribuição K_0 sob hipótese nula. Esta hipótese não é equivalente a hipótese de testar se os dados estão relacionados a cópula C_0 , que compõem o objetivo da dissertação. De modo geral a cópula C não é unicamente determinada por sua distribuição K . Por exemplo, este fato segue para a classe de cópulas do valor extremo. Portanto nesta dissertação utiliza-se a classe das cópulas Arquimedianas, apresentada no Capítulo 1, que satisfaz C ser unicamente determinada por K , fato que favorece o estudo do Kendall Plot. A grande variedade de estruturas de dependências que esta classe abrange - como por exemplo ser adequada para modelagem de dados financeiros - a facilidade de sua construção e a facilidade da construção de sua função distribuição K (objeto do teste realizado pelo Kendall Plot) também são motivos para ser contemplada nesta dissertação.

Dada a existência de uma relação entre a medida de concordância τ de Kendall e a função distribuição K da v.a. BIPIT, o gráfico Kendall Plot possibilita verificar o quão dependente são os dados em relação a hipótese nula, dependência esta definida pelo τ de Kendall, apenas observando o gráfico Kendall Plot. A medida τ de Kendall e sua relação com o Kendall Plot é explorada no Capítulo 3.

1.1 Introdução a Teoria de Cópulas

A modelagem através da distribuição normal é amplamente utilizada por sua simplicidade analítica e fácil estimação da matrix de correlação, seu único parâmetro de dependência, e por conta da vasta gama de estruturas que podem ser modeladas por esta distribuição. Porém algumas de suas características, de simetria e curtose, por exemplo, limitam sua utilização. Como Embrechts et. al. (2002) [9] mostram, há muitos obstáculos à suposição de normalidade. Para os mercados de crédito e financeiro, o principal é a característica de pequena probabilidade em eventos extremos conjuntos. Evidências empíricas sugerem que no comportamento destes mercados verificam-se eventos extremos mais prováveis que os previstos pela distribuição normal, não só nas marginais, mas também em dimensões superiores. Neste contexto, a modelagem através das cópulas torna-se atraente devido a sua maior variedade de estruturas de dependência.

Desde que Joe (1997) [28] e Nelsen (1999) [30] pela primeira vez introduziram o conceito de cópulas para uso em modelagem padrão, tem havido um interesse crescente nesta abordagem. Cópulas tornaram-se uma ferramenta popular de modelagem multivariada em muitos domínios onde a dependência multivariada é de interesse e o uso da habitual normalidade multivariada está em questão. Em ciências atuárias cópulas são usadas na modelagem de mortalidade e perdas (Frees et. al. (1996) [11]; Frees & Valdez (1998) [12]; Frees & Wang (2005) [13]). Em finanças, cópulas são usadas na classificação de crédito e modelagem de risco (Bouyé et. al. (2000) [2]; Embrechts et. al. (2003) [10]; Cherubini et. al. (2004) [6]). Em estudos biomédicos, cópulas são utilizadas na modelagem de eventos correlacionados e riscos competitivos (Wang & Welss (2000) [40]). Em engenharia, cópulas são utilizadas no controle de processo multivariado e modelagem hidrológica (Genest & Favre (2007) [21]).

A cópula é uma distribuição multivariada cujas marginais são $U(0, 1)$. Seja o vetor aleatório $\mathbf{U} = (U_1, \dots, U_d) \in \mathbf{I}^d$ com cópula d -dimensional C , temos

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d), \quad (u_1, \dots, u_d) \in \mathbf{I}^d$$

e combinado com o fato de que qualquer v.a. contínua pode ser transformada por sua acumulada para uma v.a. com distribuição $U(0, 1)$, cópulas podem ser usadas para fornecer uma estrutura de dependência multivariada separadamente das distribuições marginais. Seja \mathbb{H} uma f.d.a. de $\mathbf{X} = (X_1, \dots, X_d)$ com marginais F_1, \dots, F_d , Sklar (1959)

[38] mostrou que existe uma cópula C d -dimensional tal que para todo $x_i \in \text{Dom } F_i$,

$$\mathbb{H}(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1.1)$$

Como visto, com cópulas pode-se trabalhar a estrutura de dependência em um contexto multivariado, porém esta dissertação trata apenas do contexto bidimensional, apresentando todos os resultados para esta dimensão. Também não há impedimento para se trabalhar com o Kendall Plot, objeto central da tese, em um contexto multivariado, por uma questão de exemplificação trabalha-se em um contexto bivariado.

A função cópula C associada ao vetor aleatório (X, Y) também é conhecida como *função dependência*, já que a cópula contém toda informação de dependência entre as variáveis X e Y . Desta maneira a utilização da equação (1.1) possibilita trabalhar com a estrutura de dependência de (X, Y) de forma livre de medida de escala e locação, restringindo-se apenas ao intervalo $[0, 1]$. A característica da informação de dependência que a cópula contém pode ser vista através do seguinte exemplo.

Exemplo 1.1. São geradas duas amostras, de forma independente. Sejam $(x_1 \dots x_{100})$ e $(y_1 \dots y_{100})$ pseudo observações independentes com distribuição $F(x) = 1 - \exp(-2x)$ e $G(y) = 1 - \exp(-10y)$ respectivamente. Observando o *scatterplot* de X e Y na Figura 1.1 vê-se que não fica evidente a independência entre essas variáveis. Porém, ao plotar $(F(x_i), G(y_i))$, as amostras transformadas pelas respectivas acumuladas F e G , fica evidente a independência entre X e Y . Visto que a cópula é a distribuição conjunta entre $F(X)$ e $G(Y)$. Mostra-se que a cópula evidencia a verdadeira independência existente entre X e Y e como a forma das marginais pode produzir uma falsa impressão de dependência.

Neste capítulo o conceito de cópula, sua relação com o Teorema de Sklar e conceitos sobre a classe das cópulas Arquimedianas, necessários ao desenvolvimento da dissertação, são apresentadas formalmente.

1.1.1 Cópula e Teorema de Sklar

Definição 1.1. Sejam S_1 e S_2 subconjuntos não vazios de $\bar{\mathbb{R}}$, função $\mathbb{H} : S_1 \times S_2 \rightarrow \mathbb{R}$ e o retângulo $B = [x_1, x_2] \times [y_1, y_2]$, $(x_i, y_i) \in \text{Dom } H$ com $i, j = 1, 2, \dots$. Defina-se \mathbb{H} -volume de B como

$$V_{\mathbb{H}}(B) = \mathbb{H}(x_2, y_2) + \mathbb{H}(x_2, y_1) - \mathbb{H}(x_1, y_2) + \mathbb{H}(x_1, y_1) \quad (1.2)$$

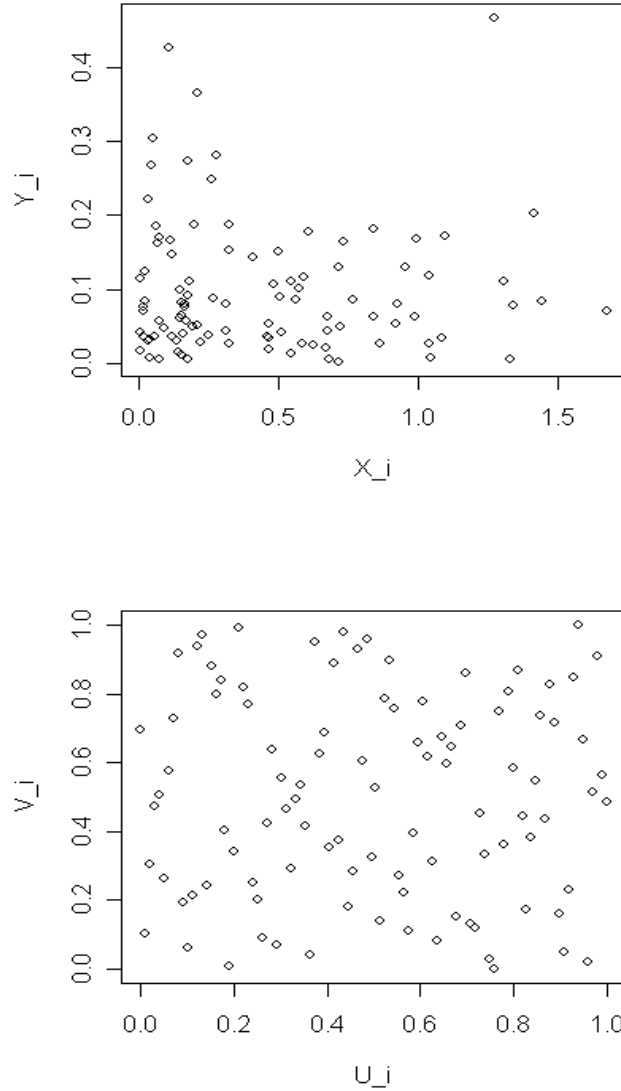


Figura 1.1: *Scatterplots*: Dados exponenciais independentes (primeira figura); dados exponenciais independentes transformados pelas acumuladas (segunda figura).

Definição 1.2. *Uma função \mathbb{H} bidimensional é bicrescente se $V_{\mathbb{H}}(B) \geq 0, \forall B$, com $V_{\mathbb{H}}(B)$ dado pela Definição 1.1.*

Exemplo 1.2. Uma função não decrescente marginalmente pode não ser bicrescente. Seja $\mathbb{H} : \mathbf{I}^2 \rightarrow \mathbf{I}$ definida por $\mathbb{H}(x, y) = \max(x, y)$, temos $V_{\mathbb{H}}(\mathbf{I}^2) = -1$, ou seja, \mathbb{H} não é bicrescente.

Exemplo 1.3. Uma função pode ser bicrescente e decrescente em alguns de seus argumentos. Seja $\mathbb{H} : \mathbf{I}^2 \rightarrow \mathbf{I}$ definida por $\mathbb{H}(x, y) = (2x - 1)(2y - 1)$. Note que $V_{\mathbb{H}}(B)$ pode ser reescrito como $V_{\mathbb{H}}(B) = (y_2 - y_1)(4x_2 - 4x_1) \geq 0$, pois $x_1 \leq x_2$ e $y_1 \leq y_2$ no retângulo B , logo \mathbb{H} é bicrescente. Porém \mathbb{H} é uma função decrescente de x para cada $y \in (0, 1/2)$ e função decrescente de y para cada $x \in (0, 1/2)$.

Definição 1.3. Uma aplicação $\mathbb{H} : S_1 \times S_2 \rightarrow \mathbb{R}$ é aplanada se $\mathbb{H}(x, a_1) = 0 = \mathbb{H}(a_1, y)$, $\forall (x, y) \in S_1 \times S_2$, onde $a_i = \min\{z : z \in S_i\}$.

Definição 1.4. Uma subcopula é uma função C' com as seguintes propriedades:

1. $\text{Dom } C' = S_1 \times S_2$, onde S_1 e S_2 são subconjuntos de \mathbf{I} contendo 0 e 1;
2. C' é aplanada e bicrescente;
3. C' é marginalmente uniforme, ou seja, para todo $u \in S_1$ e $v \in S_2$,

$$C'(u, 1) = u \quad \text{e} \quad C'(1, v) = v \tag{1.3}$$

Exemplo 1.4. Sejam $S_1 = S_2 = \{0, 1\}$. A função $C' : S_1 \times S_2 \rightarrow \mathbb{R}$ definida por

$$C'(0, 0) = C'(0, 1) = C'(1, 0) = 0 \quad \text{e} \quad C'(1, 1) = 1$$

é a mais simples subcopula.

Definição 1.5. Uma cópula é uma subcopula cujo domínio é \mathbf{I}^2 .

Exemplo 1.5. As funções $W, M : \mathbf{I}^2 \rightarrow \mathbf{I}$ dadas por $W(u, v) = \max(u + v - 1, 0)$ e $M(u, v) = \min(u, v)$ são exemplos particulares de cópulas denotadas respectivamente por M e W . É trivial mostrar que a função $M(u, v)$ é aplanada e marginalmente uniforme. Para verificar que esta função é bicrescente, considere o retângulo $B = [x_1, x_2] \times [y_1, y_2]$, x_1, x_2, y_1 e $y_2 \in \mathbf{I}$, para o caso $y_2 \geq x_2 \geq x_1$ e $x_1 \leq y_1 \leq x_2$ e desta maneira

$$V_M(B) = M(x_2, y_2) - M(x_2, y_1) - M(x_1, y_2) + M(x_1, y_1) = x_2 - y_1 \geq 0$$

Para os demais casos a demonstração é análoga. A cópula W pertence a classe das cópulas Arquimedianas e será vista na próxima seção. Estas cópulas apresentam importantes propriedades trabalhadas ao longo desta dissertação. Uma destas propriedades é mencionada no Teorema 1.7.

Como toda cópula é uma subcópula, os próximos conceitos são apresentados para subcópulas, logo válidos paracópulas.

Teorema 1.6. *Uma subcópula bidimensional C' é marginalmente não decrescente.*

Demonstração. Primeiro, provemos que a aplicação $t \rightarrow C'(t, y_2) - C'(t, y_1)$, $y_1, y_2 \in \mathbf{I}$, $y_1 \leq y_2$ é não decrescente. Consideremos $T(t) = C'(t, y_2) - C'(t, y_1)$, se $t_1 \leq t_2$, $t_i \in \mathbf{I}$, temos $T(t_2) - T(t_1) \geq 0$, pois

$$T(t_2) - T(t_1) = C'(t_2, y_2) - C'(t_2, y_1) - C'(t_1, y_2) + C'(t_1, y_1) = V'_C(B)$$

onde $B = [t_1, t_2] \times [y_2, y_1]$. Como C' é bicrescente, então $V'_C(B) \geq 0$, logo a aplicação é não decrescente. Analogamente verifica-se que a aplicação $T^*(t) = C'(x_2, t) - C'(x_1, t)$, $x_1, x_2 \in \mathbf{I}$, $x_1 \leq x_2$ é não decrescente. Agora, seja $x_1 = y_1 = 0$ nas aplicações T e T^* , como C' é aplanada, segue que C' é marginalmente não decrescente. \square

Teorema 1.7. *Seja C' uma subcópula. Então para todo $(u, v) \in \text{Dom } C'$,*

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v) \quad (1.4)$$

Demonstração. Como C' é aplanada, marginalmente uniforme e não decrescente temos

$$0 = C'(0, v) \leq C'(u, v) \leq C'(1, v) = v \quad e \quad 0 = C'(u, 0) \leq C'(u, v) \leq C'(u, 1) = u \quad (1.5)$$

Logo,

$$C'(u, v) \leq \min(u, v)$$

Como C' também é bicrescente, então $V_{C'}(B) \geq 0$. Sendo $B = [u, 1] \times [v, 1]$ temos

$$\begin{aligned} V_{C'}(B) &= C'(1, 1) + C'(u, v) - C'(u, 1) - C'(1, v) \\ &= 1 + C'(u, v) - u - v \geq 0 \end{aligned} \quad (1.6)$$

Portanto pelas equações (1.5) e (1.6), temos $C'(u, v) \geq \max(u + v - 1, 0)$. \square

Conforme Exemplo 1.5, os limites na equação (1.4) são as cópulas M e W . Então para toda cópula C e para todo $(u, v) \in \mathbf{I}^2$,

$$W(u, v) \leq C(u, v) \leq M(u, v) \quad (1.7)$$

A desigualdade (1.7) é denominada *desigualdade dos limites de Frechét*, a cópula M é nomeada por *limite superior de Frechét* e a cópula W por *limite inferior de Frechét*.

A ligação entre funções distribuição multivariada e suas marginais univariadas é feita pelo Teorema de Sklar apresentado a seguir, por meio das cópulas. O nome *cópula* foi escolhido para enfatizar a maneira como a cópula une uma função distribuição conjunta às suas marginais univariadas.

Teorema 1.8. *Seja \mathbb{H} uma função distribuição conjunta com marginais F e G . Então existe uma cópula C tal que para todo $x, y \in \bar{\mathbb{R}}$,*

$$\mathbb{H}(x, y) = C(F(x), G(y)) \quad (1.8)$$

Se F e G são contínuas, então C é única; caso contrário, C é unicamente determinada em $Im(F) \times Im(G)$. Inversamente, se C é uma cópula e F e G são f.d.a., então a função \mathbb{H} definida pela equação (1.8) é uma função distribuição conjunta com marginais F e G .

Demonstração. A prova segue dos lemas 1.9 e 1.10 apresentados a seguir. \square

Lema 1.9. *Seja \mathbb{H} uma função distribuição conjunta com marginais F e G . Então existe uma única subcópula C' tal que*

1. $Dom C' = Im(F) \times Im(G)$;
2. $\forall x, y \in \bar{\mathbb{R}}, \mathbb{H}(x, y) = C'(F(x), G(y))$

Demonstração. Sejam $(x_1, y_1), (x_2, y_2) \in Dom \mathbb{H}$. Pela desigualdade triangular, temos

$$|\mathbb{H}(x_2, y_2) - \mathbb{H}(x_1, y_1)| \leq |\mathbb{H}(x_2, y_2) - \mathbb{H}(x_1, y_2)| + |\mathbb{H}(x_1, y_2) - \mathbb{H}(x_1, y_1)|$$

Considere $x_1 \leq x_2$. Como \mathbb{H} é uma função distribuição conjunta, então

$$0 \leq \mathbb{H}(x_2, y_2) - \mathbb{H}(x_1, y_2) \leq F(x_2) - F(x_1)$$

por \mathbb{H} ser marginalmente decrescente, aplanada, $\mathbb{H}(x, \infty) = F(x)$ e $\mathbb{H}(\infty, y) = G(y)$. Uma desigualdade análoga é válida quando $x_2 \leq x_1$. Portanto segue que para qualquer $x_1, x_2 \in S_1$, $|\mathbb{H}(x_2, y_2) - \mathbb{H}(x_1, y_2)| \leq |F(x_2) - F(x_1)|$. Similarmente para qualquer $y_1, y_2 \in S_2$, $|\mathbb{H}(x_1, y_2) - \mathbb{H}(x_1, y_1)| \leq |G(y_2) - G(y_1)|$, logo

$$|\mathbb{H}(x_2, y_2) - \mathbb{H}(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \quad (1.9)$$

Desta forma, segue que se $F(x_1) = F(x_2)$ e $G(y_1) = G(y_2)$, então $\mathbb{H}(x_1, y_1) = \mathbb{H}(x_2, y_2)$ e conseqüentemente o conjunto de pares $\{(F(x), G(y)), \mathbb{H}(x, y)\}$ permite definir uma função

$$(F(x), G(y)) \xrightarrow{C'} \mathbb{H}(x, y)$$

sendo C' única com domínio $Im(F) \times Im(G)$. A prova que a função C' é uma subcopula segue diretamente das propriedades da distribuição conjunta \mathbb{H} (James (2004) [27]). Para cada $u \in Im(F)$, existe um $x \in \mathbb{R}$ tal que $F(x) = u$, então

$$C'(u, 1) = C'(F(x), G(\infty)) = \mathbb{H}(x, \infty) = F(x) = u$$

$$C'(u, 0) = C'(F(x), G(-\infty)) = \mathbb{H}(x, -\infty) = 0$$

Como \mathbb{H} é bicrescente por definição, logo C' é bicrescente. \square

Lema 1.10. *Seja C' uma subcopula. Então existe uma cópula C tal que $C(u, v) = C'(u, v)$, $\forall (u, v) \in Dom C'$; isto é, qualquer subcopula pode ser estendida a uma cópula. A extensão é geralmente não única.*

Demonstração. Da equação (1.9) e usando o item 2. do Lema 1.9, concluímos que C' é uniformemente contínua em seu domínio $Im(F) \times Im(G)$, ou seja,

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

Denotemos $Im(F)$ por S_1 e $Im(G)$ por S_2 .

Da continuidade de C' podemos estender C' a uma função C'' , ainda subcopula, com domínio em $\bar{S}_1 \times \bar{S}_2$, onde \bar{S}_1 é o fecho de S_1 e \bar{S}_2 é o fecho de S_2 . O próximo passo é estender a subcopula C'' a uma função C com domínio em \mathbf{I}^2 . Para este fim, seja $(a, b) \in \mathbf{I}^2$, e sejam a_1 e a_2 , respectivamente, o menor e o maior elemento de \bar{S}_1 que satisfaça $a_1 \leq a \leq a_2$; e sejam b_1 e b_2 , respectivamente, o menor e o maior elemento de \bar{S}_2 que satisfaça $b_1 \leq b \leq b_2$. Se $a \in \bar{S}_1$, defina $a_1 = a_2 = a$; e se $b \in \bar{S}_2$, defina $b_1 = b_2 = b$. Desta maneira, sejam

$$\lambda_1 = \begin{cases} (a - a_1)/(a_2 - a_1), & \text{se } a_1 < a_2 \\ 1, & \text{se } a_1 = a_2 \end{cases}$$

$$\mu_1 = \begin{cases} (b - b_1)/(b_2 - b_1), & \text{se } b_1 < b_2 \\ 1, & \text{se } b_1 = b_2 \end{cases}$$

e define-se

$$\begin{aligned} C(a, b) &= (1 - \lambda_1)(1 - \mu_1)C''(a_1, b_1) + (1 - \lambda_1)\mu_1C''(a_1, b_2) \\ &+ \lambda_1(1 - \mu_1)C''(a_2, b_1) + \lambda_1\mu_1C''(a_2, b_2). \end{aligned} \quad (1.10)$$

É trivial que $Dom C = \mathbf{I}^2$ e que $C(a, b) = C''(a, b)$, $\forall (a, b) \in Dom C''$. Como λ_1 e μ_1 são lineares em a e b , a forma $C(a, b)$ da equação (1.10) é bilinear em (a, b) propriedade que permite demonstrar que C é aplanada e marginalmente uniforme. Conforme Definição 1.5, para C ser uma cópula falta provar que C é bicscente. Seja (c, d) outro ponto em \mathbf{I}^2 tal que $c \geq a$ e $d \geq b$, e sejam $c_1, d_1, c_2, d_2, \lambda_2, \mu_2$ relacionados a c e d assim como $a_1, b_1, a_2, b_2, \lambda_1, \mu_1$ são relacionados a a e b . Deve-se provar que $V_C(B) \geq 0$ para o retângulo $B = [a, c] \times [b, d]$. Diversos casos são considerados para esta prova. O mais simples dos casos é aquele em que não existe ponto em \bar{S}_1 estritamente entre a e c e não há ponto em \bar{S}_2 estritamente entre b e d , então temos $c_1 = a_1, c_2 = a_2, d_1 = b_1$ e $d_2 = b_2$. Substituindo a equação (1.10) e os termos correspondentes para $C(a, d), C(c, b)$ e $C(c, d)$ na expressão dada pela equação (1.2) para $V_C(B)$, temos

$$V_C(B) = V_C([a, c] \times [b, d]) = (\lambda_2 - \lambda_1)(\mu_2 - \mu_1)V_C([a_1, a_2] \times [b_1, b_2]),$$

de que segue que $V_C(B) \geq 0$ neste caso, pois $c \geq a$ e $d \geq b$ implica $\lambda_2 \geq \lambda_1$ e $\mu_2 \geq \mu_1$.

Outro caso de interesse é quando pelo menos um ponto está em \bar{S}_1 estritamente entre a e c , e pelo menos um ponto está em \bar{S}_2 estritamente entre b e d , então $a < a_2 \leq c_1 < c$ e $b < b_2 \leq d_1 < d$. Para este caso, substituindo a equação (1.10) e os termos correspondentes para $C(a, d), C(c, b)$ e $C(c, d)$ na expressão dada pela equação (1.2) para $V_C(B)$ e reordenando os termos temos

$$\begin{aligned} V_C(B) &= (1 - \lambda_1)\mu_2 V_C([a_1, a_2] \times [d_1, d_2]) + \mu_2 V_C([a_2, c_1] \times [d_1, d_2]) \\ &+ \lambda_2 \mu_2 V_C([c_1, c_2] \times [d_1, d_2]) + (1 - \lambda_1) V_C([a_1, a_2] \times [b_2, d_1]) \\ &+ V_C([a_2, c_1] \times [b_2, d_1]) + \lambda_2 V_C([c_1, c_2] \times [b_2, d_1]) \\ &+ (1 - \lambda_1)(1 - \mu_1) V_C([a_1, a_2] \times [b_1, b_2]) \\ &+ (1 - \mu_1) V_C([a_2, c_1] \times [b_1, b_2]) + \lambda_2(1 - \mu_1) V_C([c_1, c_2] \times [b_1, b_2]). \end{aligned}$$

O lado direito da expressão acima é a soma de combinações de nove quantidades não negativas, representadas por C -volumes, com coeficientes não negativos, logo é não negativa. Os casos remanescentes são similares, que completam a prova. \square

Antes de apresentar o próximo resultado é necessária a definição de *quasi-inversa*.

Definição 1.11. *Seja F uma função distribuição acumulada. Então a quasi-inversa de F é qualquer função $F^{(-1)}$ com domínio em \mathbf{I} tal que*

1. *Se $t \in Im(F)$, então $F^{(-1)}(t) = x$ com $x \in \mathbb{R}$ tal que $F(x) = t$, ou seja, $\forall t \in Im(F)$,*

$$F(F^{(-1)}(t)) = t;$$

2. Se $t \notin \text{Im}(F)$, então

$$F^{(-1)}(t) = \inf\{x|F(x) \geq t\} = \sup\{x|F(x) \leq t\}.$$

Corolário 1.12. *Sejam \mathbb{H}, F, G e C' como no Lema 1.9, e sejam $F^{(-1)}$ e $G^{(-1)}$ as quasi-inversas de F e G , respectivamente. Então para qualquer $(u, v) \in \text{Dom } C'$,*

$$C'(u, v) = \mathbb{H}(F^{(-1)}(u), G^{(-1)}(v)) \quad (1.11)$$

Sejam $x \in \text{Dom}(F)$ e $y \in \text{Dom}(G)$ conforme Teorema 1.8 se $u = F(x)$ e $v = G(y)$, então $x = F^{(-1)}(u)$ e $y = G^{(-1)}(v)$. Quando F e G são contínuas o Corolário 1.12 é válido para cópulas e temos U, V com distribuição $U(0, 1)$ onde $U = F(X)$ e $V = G(Y)$. Deste modo a cópula C é distribuição conjunta do vetor aleatório (U, V) com marginais $U(0, 1)$.

Exemplo 1.6. Um método para construção de cópulas é resultante do último corolário. Seja \mathbb{H} a distribuição da independência

$$\mathbb{H}(x, y) = F(x)G(y) \quad (1.12)$$

com marginais $F(x) = 1 - \exp(-2x)$ e $G(y) = 1 - \exp(-y)$, onde as respectivas quasi-inversas são dada por

$$F^{(-1)}(u) = \begin{cases} -\frac{\log(1-u)}{2}, & \text{se } x \geq 0 \\ 0, & \text{c.c.} \end{cases}$$

$$G^{(-1)}(v) = \begin{cases} -\log(1-v), & \text{se } x \geq 0 \\ 0, & \text{c.c.} \end{cases}$$

Logo, a cópula associada a \mathbb{H} é dada por

$$\begin{aligned} C(u, v) &= \mathbb{H}(F^{(-1)}(u), G^{(-1)}(v)) \\ &= F\{F^{(-1)}(u)\}G\{G^{(-1)}(v)\} \\ &= [1 - e^{-2\log(1-u)/2}][1 - e^{-\log(1-v)}] \\ &= uv \end{aligned}$$

Esta cópula recebe a notação especial Π e é denominada *cópula produto*. Sua estrutura caracteriza a independência entre as v. a. X e Y quando as f.d.a. F e G são contínuas. Note que se a conjunta \mathbb{H} é definida pela equação (1.12), a estrutura de dependência entre as v.a. X e Y independe de suas marginais, pois

$$C(u, v) = \mathbb{H}(F^{(-1)}(u), G^{(-1)}(v)) = F\{F^{(-1)}(u)\}G\{G^{(-1)}(v)\} = uv$$

pela definição de quasi-inversa (Definição 1.11).

Muito da utilidade de cópulas no estudo de estatísticas não paramétricas deriva do fato que sob transformações monótonas estritas de v.a. as cópulas são invariantes, ou são alteradas de modo previsível como visto nos próximos dois resultados.

Teorema 1.13. *Sejam X e Y v.a. contínuas com cópula $C_{X,Y}$. Se α e β são funções estritamente crescentes na $Im(X)$ e $Im(Y)$ respectivamente, então $C_{\alpha(X),\beta(Y)} = C_{X,Y}$.*

Demonstração. Sejam F_1, G_1, F_2 e G_2 as respectivas distribuições de $X, Y, \alpha(X)$ e $\beta(Y)$, respectivamente. Assim, para qualquer $x, y \in \mathbb{R}$,

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(F_2(x), G_2(y)) &= P[\alpha(X) \leq x, \beta(Y) \leq y] = P[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)] \\ &= C_{X,Y}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) = C_{X,Y}(F_2(x), G_2(y)) \end{aligned}$$

Desde que X e Y são contínuas, $Im(F_2) = Im(G_2) = \mathbf{I}$, logo segue que $C_{\alpha(X),\beta(Y)} = C_{X,Y}$ em \mathbf{I}^2 . □

Quando pelo menos uma função, α ou β , é estritamente decrescente, a cópula das v.a. $\alpha(X)$ e $\beta(Y)$ é uma simples transformação de $C_{X,Y}$.

Teorema 1.14. *Sejam X, Y v.a. contínuas com cópula $C_{X,Y}$ e α e β funções estritamente monótonas na $Im(X)$ e $Im(Y)$ respectivamente.*

1. *Se α é estritamente crescente e β é estritamente decrescente, então*

$$C_{\alpha(X),\beta(Y)}(u, v) = u - C_{X,Y}(u, 1 - v), \quad u, v \in \mathbf{I}.$$

2. *Se α é estritamente decrescente e β é estritamente crescente, então*

$$C_{\alpha(X),\beta(Y)}(u, v) = v - C_{X,Y}(1 - u, v), \quad u, v \in \mathbf{I}.$$

3. *Se α e β ambas são estritamente decrescentes, então*

$$C_{\alpha(X),\beta(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v), \quad u, v \in \mathbf{I}.$$

Demonstração. Sejam F_1, G_1, F_2 e G_2 as distribuições de $X, Y, \alpha(X)$ e $\beta(Y)$ respectivamente. Tomemos $u \in \text{Im}(F_2)$ e $v \in \text{Im}(G_2)$ tais que $u = F_2(x)$ para algum $x \in \text{Im}(F_2) = \mathbf{I}$ (devido a v.a. $\alpha(X)$ ser contínua) e $v = G_2(y)$ para algum $y \in \text{Im}(G_2) = \mathbf{I}$ (devido a v.a. $\beta(Y)$ ser contínua).

Demonstração item 1.:

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(u, v) &= C_{\alpha(X),\beta(Y)}(F_2(x), G_2(y)) \stackrel{(1.8)}{=} P(\alpha(X) \leq x, \beta(Y) \leq y) \\ &= P(X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) \end{aligned} \quad (1.13)$$

Como,

$$\begin{aligned} F_2(x) &= P(\alpha(X) \leq x) = P(X \leq \alpha^{-1}(x)) = F_1(\alpha^{-1}(x)) \\ &= P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) + P(X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) \end{aligned} \quad (1.14)$$

e

$$\begin{aligned} G_2(y) &= P(\beta(Y) \leq y) = P(Y \geq \beta^{-1}(y)) = 1 - G_1(\beta^{-1}(y)) \\ &\Rightarrow G_1(\beta^{-1}(y)) = 1 - G_2(y) \end{aligned} \quad (1.15)$$

temos

$$\begin{aligned} P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) &= C_{X,Y}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\ &= C_{X,Y}(F_2(x), 1 - G_2(y)) \\ &= C_{X,Y}(u, 1 - v) \end{aligned} \quad (1.16)$$

onde a segunda igualdade se refere às equações (1.14) e (1.15).

Substituindo a equação (1.16) na equação (1.14)

$$\begin{aligned} F_2(x) &= P(X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) + C_{X,Y}(u, 1 - v) \\ &\Rightarrow P(X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) = u - C_{X,Y}(u, 1 - v) \end{aligned} \quad (1.17)$$

Desta forma, substituindo a equação (1.17) na equação (1.13), se completa a prova.

Demonstração item 2.: análoga a demonstração item 1.

Demonstração item 3.:

$$\begin{aligned}
C_{\alpha(X),\beta(Y)}(u, v) &= C_{\alpha(X),\beta(Y)}(F_2(x), G_2(y)) \stackrel{(1.8)}{=} P(\alpha(X) \leq x, \beta(Y) \leq y) \\
&= P(X \geq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) \\
&= 1 - [P(X \leq \alpha^{-1}(x)) + P(Y \leq \beta^{-1}(y)) - P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y))] \\
&= 1 - F_1(\alpha^{-1}(x)) - G_1(\beta^{-1}(y)) + P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) \quad (1.18)
\end{aligned}$$

Reescrevendo $F_1(\alpha^{-1}(x))$, temos

$$F_1(\alpha^{-1}(x)) = 1 - P(X \geq \alpha^{-1}(x)) = 1 - P(\alpha(X) \leq x) = 1 - F_2(x) \quad (1.19)$$

e, analogamente

$$G_1(\beta^{-1}(y)) = 1 - G_2(y) \quad (1.20)$$

Substituindo as equações (1.19) e (1.20) na equação (1.18) e utilizando o Teorema 1.8,

$$\begin{aligned}
C_{\alpha(X),\beta(Y)}(u, v) &= 1 - [1 - F_2(x)] - [1 - G_2(y)] + C_{X,Y}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\
&= u + v - 1 + C_{X,Y}(1 - u, 1 - v)
\end{aligned}$$

□

1.1.2 Cópulas Arquimedianas

Nesta dissertação a modelagem de dados será trabalhada via cópulas Arquimedianas. O termo *cópula Arquimediana* foi mencionado pela primeira vez na literatura estatística em dois artigos de Genest & Mackay (1986ab) [15] e [16]. Cópulas Arquimedianas também são mencionadas por Schweizer and Sklar (1983) [36], mas sem o nome Arquimediana.

A Classe das cópulas Arquimedianas abrange uma grande variedade de estruturas de dependência, incluindo estruturas próprias de estudos financeiros. Algumas questões em finanças exigem modelos que permitem uma forte dependência entre as perdas extremas (por exemplo, bolsas em colapso) e ganhos extremos. Em particular, as cópulas Arquimedianas podem assumir dependência caudal assimétrica, sendo uma propriedade a favor de sua aplicação à modelagem de dados com estrutura de dependência assimétrica. Em estudos financeiros um aspecto importante a ser analisado é a dependência caudal.

Dependência caudal inferior e superior entre dois mercados financeiros existe quando a probabilidade de valores conjuntos negativos (positivos) em eventos extremos é maior que a que poderia ser prevista a partir das distribuições marginais. Recentes estudos empíricos mostram que períodos de turbulência e calma em finanças são caracterizados por diferentes níveis de dependência caudal, sendo a dependência mais forte sobre a cauda inferior do que na cauda superior.

As cópulas Arquimedianas podem ser construídas facilmente e a forma fechada para sua expressão é simples. Estas facilidades se devem ao fato da representação da cópula Arquimediana permitir reduzir o estudo de cópula multivariada ao estudo de uma função univariada denotada por gerador de uma cópula Arquimediana ϕ . A seguir esta função é apresentada com mais detalhes.

Definição 1.15. *Seja $\phi : \mathbf{I} \rightarrow [0, \infty]$ uma função contínua e estritamente decrescente tal que $\phi(1) = 0$. A pseudo-inversa de ϕ é a função $\phi^{[-1]} : [0, \infty] \rightarrow \mathbf{I}$ dada por*

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

Note que $\phi^{[-1]}$ é contínua e não crescente em $[0, \infty]$, e estritamente decrescente em $[0, \phi(0)]$. Além disso,

$$\phi^{[-1]}(\phi(u)) = u, \quad \forall u \in \mathbf{I} \tag{1.21}$$

e

$$\phi(\phi^{[-1]}(t)) = \begin{cases} t, & \text{se } 0 \leq t \leq \phi(0) \\ \phi(0), & \text{se } \phi(0) \leq t \leq \infty \end{cases} = \min(t, \phi(0))$$

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

Lema 1.16. *Seja ϕ e $\phi^{[-1]}$ como na Definição 1.15. Seja a função $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ dada por*

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{1.22}$$

Então C satisfaz duas das três condições para uma cópula, é aplanada e marginalmente uniforme.

Demonstração.

$$C(u, 0) = \phi^{[-1]}(\phi(u) + \phi(0)) = 0$$

A última igualdade segue da definição de pseudo-inversa, pois $\phi(u) + \phi(0) \geq \phi(0)$ sendo ϕ uma função que assume somente valores positivos.

$$C(u, 1) = \phi^{[-1]}(\phi(u) + \phi(1)) = \phi^{[-1]}(\phi(u)) \stackrel{(1.21)}{=} u \quad (1.23)$$

Analogamente, por simetria, $C(0, v) = 0$ e $C(1, v) = v$. \square

O seguinte lema apresenta uma condição necessária e suficiente para que a função C na equação (1.22) seja bicrescente, e portanto, uma cópula.

Lema 1.17. *Sejam $\phi, \phi^{[-1]}$ e C satisfazendo as hipóteses do Lema 1.16. Então C é bicrescente, se e somente se, sempre que $u_1 \leq u_2$,*

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \quad (1.24)$$

Demonstração. (\Rightarrow)

$$V_C([u_1, u_2] \times [v, 1]) = C(u_1, v) + C(u_2, 1) - C(u_1, 1) - C(u_2, v) \geq 0 \Leftrightarrow u_2 - u_1 \geq C(u_2, v) - C(u_1, v)$$

Então, a equação (1.24) é equivalente à condição $V_C([u_1, u_2] \times [v, 1]) \geq 0$ que é sempre válida quando C é bicrescente.

(\Leftarrow) Considere C satisfazendo a equação (1.24). Sejam $v_1, v_2 \in \mathbf{I}$, $v_1 \leq v_2$, temos $C(0, v_2) = 0 \leq v_1 \leq v_2 = C(1, v_2)$. Como C é contínua, desde que ϕ e $\phi^{[-1]}$ também são, existe $t \in \mathbf{I}$ tal que $C(t, v_2) = v_1$, ou seja, $\phi(v_2) + \phi(t) = \phi(v_1)$. Então

$$\begin{aligned} C(u_2, v_1) - C(u_1, v_1) &= \phi^{[-1]}(\phi(u_2) + \phi(v_1)) - \phi^{[-1]}(\phi(u_1) + \phi(v_1)) \\ &= \phi^{[-1]}(\phi(u_2) + \phi(v_2) + \phi(t)) - \phi^{[-1]}(\phi(u_1) + \phi(v_2) + \phi(t)) \\ &= \phi^{[-1]}(\phi(C(u_2, v_2)) + \phi(t)) - \phi^{[-1]}(\phi(C(u_1, v_2)) + \phi(t)) \\ &= C(C(u_2, v_2), t) - C(C(u_1, v_2), t) \\ &\stackrel{(1.24)}{\leq} C(u_2, v_2) - C(u_1, v_2) \Leftrightarrow V_C([u_1, u_2] \times [v_1, v_2]) \geq 0 \end{aligned}$$

\square

O próximo teorema mostra que a condição da função C ser bicrescente dada pelo Lema 1.17 também encontra-se relacionada com uma propriedade da função ϕ , bem como as outras duas condições para que C seja uma cópula (Lema 1.16).

Deste modo, somente a estrutura da função ϕ pode determinar se a função C da equação (1.22) é ou não uma cópula .

Teorema 1.18. *Sejam ϕ e $\phi^{[-1]}$ como na Definição 1.15. Então a função $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ dada pela equação (1.22) é uma cópula se e somente se ϕ é convexa.*

Demonstração. Como consequência do Lema 1.17 é necessário provar que a equação (1.24) é válida se e somente se ϕ é convexa.

(\Rightarrow) Observe que a equação (1.24) é equivalente a

$$u_1 + \phi^{[-1]}(\phi(u_2) + \phi(v)) \leq u_2 + \phi^{[-1]}(\phi(u_1) + \phi(v))$$

para $u_1 \leq u_2$, se denotarmos $a = \phi(u_1)$, $b = \phi(u_2)$ e $c = \phi(v)$, então a equação (1.24) também é equivalente a

$$\phi^{[-1]}(a) + \phi^{[-1]}(b + c) \leq \phi^{[-1]}(b) + \phi^{[-1]}(a + c) \quad (1.25)$$

onde $a \geq b$, por ϕ ser decrescente, e $c \geq 0$.

Supondo a equação (1.24) válida, ou seja, supondo que $\phi^{[-1]}$ satisfaça a equação (1.25). Sejam $s, t \in [0, \infty]$ tais que $0 \leq s \leq t$. Se definirmos $a = (s + t)/2$, $b = s$ e $c = (t - s)/2$ na equação (1.25), temos

$$\phi^{[-1]} \left(\frac{s + t}{2} \right) \leq \frac{\phi^{[-1]}(s) + \phi^{[-1]}(t)}{2}$$

Logo $\phi^{[-1]}$ é mid-convexa, e desde que $\phi^{[-1]}$ é contínua segue que $\phi^{[-1]}$ é convexa. O fato de $\phi^{[-1]}$ convexa implica na convexidade de ϕ .

(\Leftarrow) Assuma $\phi^{[-1]}$ convexa. Sejam $a, b, c \in \mathbf{I}$ tais que $a \geq b$ e $c \geq 0$; e seja $0 \leq \gamma = (a - b)/(a - b + c) \leq 1$. Deste modo temos $a = (1 - \gamma)b + \gamma(a + c)$ e $b + c = \gamma b + (1 - \gamma)(a + c)$. Logo por definição de função convexa,

$$\phi^{[-1]}(a) \leq (1 - \gamma)\phi^{[-1]}(b) + \gamma\phi^{[-1]}(a + c)$$

e

$$\phi^{[-1]}(b + c) \leq \gamma\phi^{[-1]}(b) + (1 - \gamma)\phi^{[-1]}(a + c)$$

Somando as duas últimas igualdades resulta na equação (1.25), que completa a prova. \square

Cópiulas da forma apresentada na equação (1.22) são denominadas *Cópiulas Arquimedianas*. A função ϕ é denominada *gerador* de uma cópiula Arquimediana. Se $\phi(0) = \infty$, ϕ é denominada *gerador estrito*. Conforme Definição 1.15, $\phi^{[-1]} = \phi^{-1}$ e $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$ é denominada *cópiula Arquimediana estrita*. Se $\phi(0) < \infty$, ϕ é denominada *gerador não estrito*.

Cópiulas Arquimedianas podem ser contruídas usando o Teorema 1.18 - apenas encontrando funções ϕ com propriedades que satisfaçam sua hipótese - e definindo as correspondentes cópiulas via equação (1.22). Em outras palavras, a cópiula Arquimediana C é unicamente determinada pelo gerador ϕ .

Exemplo 1.5.(continuação) Seja $\phi(t) = 1 - t$, $\forall t \in \mathbf{I}$, um gerador não estrito (Figura 1.2). De acordo com a Definição 1.15 $\phi^{[-1]}(t) = \phi^{-1}(t) = 1 - t$, $\forall t \in \mathbf{I}$ e 0 para $t > 1$, ou seja, $\phi^{[-1]}(t) = \max(1 - t, 0)$. Gerando a cópiula C via equação (1.22), temos

$$\phi^{-1}(\phi(u) + \phi(v)) = \max(1 - [(1 - u) + (1 - v)], 0) = \max(u + v - 1, 0) = W(u, v)$$

Sendo então a cópiula limite inferior de Fréchet uma cópiula Arquimediana.

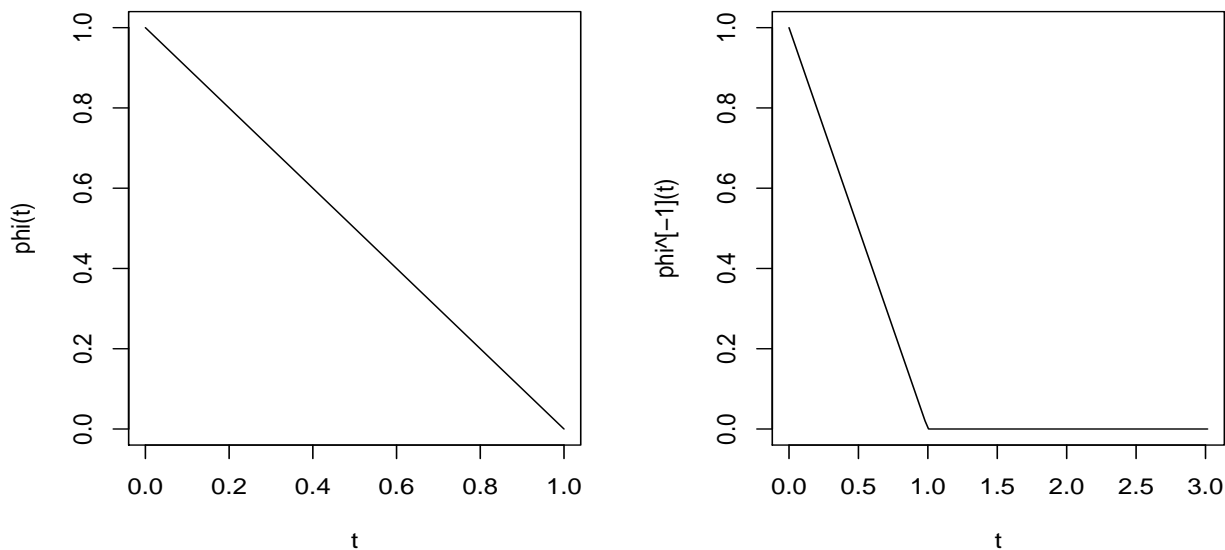


Figura 1.2: Gerador ϕ não estrito e pseudo-inversa $\phi^{[-1]}$ para a cópiula Arquimediana W .

Exemplo 1.6.(continuação) Seja $\phi(t) = -\ln(t)$, $\forall t \in \mathbf{I}$, um gerador estrito (Figura 1.3). Segue da Definição 1.15 que $\phi^{[-1]}(t) = \phi^{-1}(t) = \exp(-t)$, $0 \leq t < \infty$. Pela equação (1.22), temos

$$\phi^{-1}(\phi(u) + \phi(v)) = \exp(-[(-\ln u) + (-\ln v)]) = uv = \Pi(u, v)$$

Deste modo a cópula produto Π é também Arquimediana.

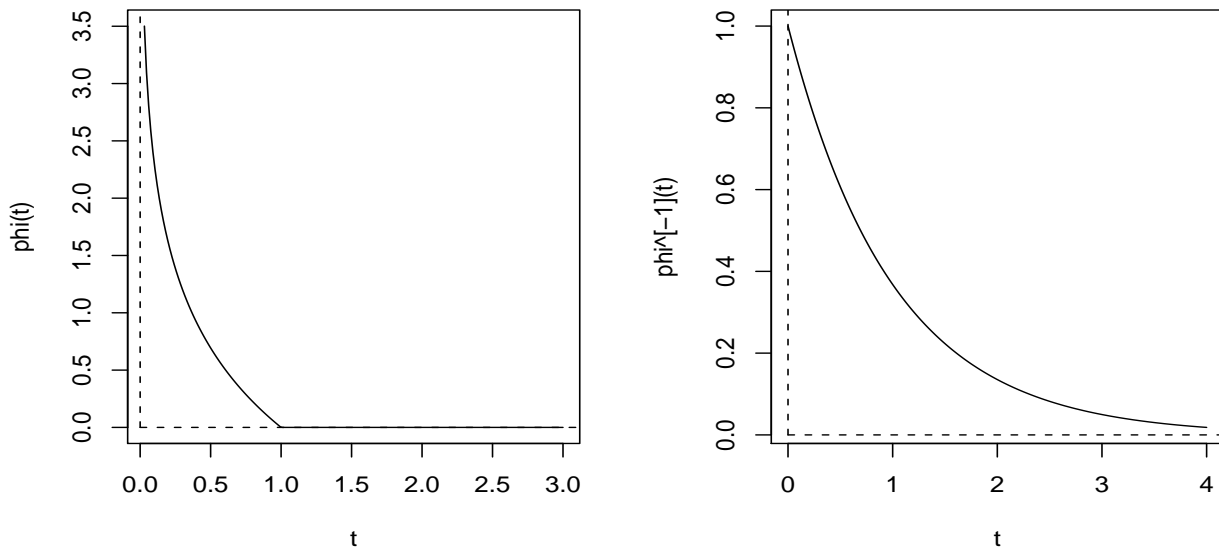


Figura 1.3: Gerador ϕ estrito e pseudo-inversa $\phi^{[-1]}$ para a cópula Arquimediana Π .

Exemplo 1.7. Uma grande variedade de famílias paramétricas de cópulas pertence a classe das cópulas Arquimedianas. Nelsen (1999) [30] apresenta uma lista extensa com as famílias de cópulas Arquimedianas mais comuns, algumas destas famílias que serão trabalhadas nesta dissertação são apresentadas na Tabela 1.1.

Foram escolhidas as famílias de Clayton, Gumbel e Frank pois quando trata-se do estudo de cópulas Arquimedianas em artigos de quaisquer autores estas três cópulas são sempre citadas e exemplificadas. Já a cópula 4.2.12 é citada na dissertação pois é uma exceção entre as famílias Arquimedianas tratando-se que apresenta uma forma fechada (forma analítica) para a inversa de sua função distribuição, fato de importância para o estudo do Kendall Plot e construção das bandas de confiança. Além disso estas quatro

<i>nome</i>	$C_\theta(u, v)$	$\phi_\theta(t)$	$\theta \in$	<i>estrito</i>
Clayton	$\max([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta})$	$\frac{t^{-\theta}-1}{\theta}$	$[-1, \infty] \setminus \{0\}$	$\theta \geq 0^*$
Gumbel	$\exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta})$	$(-\ln t)^\theta$	$[1, \infty)$	<i>sim</i>
Frank	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1} \right)$	$-\ln \frac{e^{-\theta v}-1}{e^{-\theta}-1}$	$(-\infty, \infty) \setminus \{0\}$	<i>sim</i>
4.2.12	$(1 + [(u^{-1} - 1)^\theta + (v^{-1} - 1)^\theta]^{1/\theta})^{-1}$	$(\frac{1}{t} - 1)^\theta$	$[1, \infty)$	<i>sim</i>

Tabela 1.1: Algumas famílias paramétricas de cópulas Arquimedianas com seus geradores e espaços paramétricos (*na cópula de Clayton o gerador é estrito se $\theta \geq 0$, caso contrário é não estrito).

cópulas apresentam diferenças distintas com relação a estrutura de dependência que representam. A família de Gumbel apresenta dependência caudal superior, a família de Clayton apresenta dependência caudal inferior, a família de cópulas 4.2.12 apresenta ambas dependências caudais e a família de Frank não apresenta dependência caudal, é simétrica em relação a diagonal secundária. No Capítulo 5 torna-se a comentar estas características que são essenciais para a análise do comportamento do teste Kendall Plot.

Nelsen (1999) [30] não apresenta nenhum nome especial para a última cópula da Tabela 1.1, logo esta dissertação refere-se a esta cópula conforme notação deste autor.

Nota-se que cópulas já mencionadas na dissertação pertencem a algumas das famílias apresentadas na Tabela 1.1. Por exemplo, a cópula produto Π pertence a família de Gumbel ($\theta = 1$, $\phi_1 = \phi_\Pi$) e a cópula W pertence a família de Clayton ($\theta = -1$, $\phi_{-1} = \phi_W$).

Capítulo 2

Variável BIPIT

Sabe-se que se X é v.a. unidimensional com f.d.a. F contínua, temos $U = F(X)$, a v.a. transformada pela acumulada, com distribuição $U(0, 1)$. Porém, para maiores dimensões isto geralmente não acontece. Suponha (X, Y) com f.d.a. \mathbb{H} , seja $H = \mathbb{H}(X, Y)$, a v.a. transformada pela acumulada conjunta, geralmente a f.d.a. K de H não é $U(0, 1)$.

Exemplo 2.1. Seja $\{(x_1, y_1), \dots, (x_{100}, y_{100})\}$ uma amostra aleatória de uma distribuição \mathbb{H} normal bivariada com média $\mu = (0, 0)$ e matrix de covariância

$$\Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$$

Tomemos a v.a. $H = \mathbb{H}(X, Y)$. O QQplot entre a amostra aleatória da v.a. H calculada por $H_i = \mathbb{H}(x_i, y_i)$ e uma amostra com distribuição $U(0, 1)$ pode-se concluir que H não segue distribuição $U(0, 1)$ (Figura 2.1).

O estudo da v.a. H e de sua f.d.a. K é importante, pois ambas contém informação de dependência sob $\mathbb{H}(X, Y)$, já que dependem apenas da cópula associada a \mathbb{H} pelo Teorema 1.8, e não das marginais F e G ,

$$K(t) = P(\mathbb{H}(X, Y) \leq t) = P(C(F(X), G(Y)) \leq t) = P(C(U, V) \leq t)$$

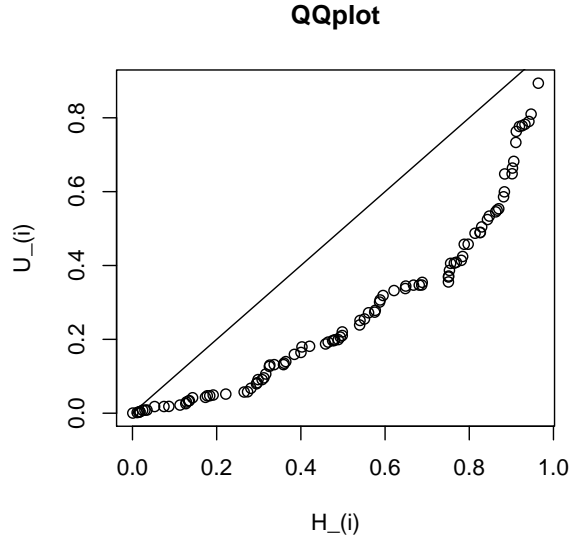


Figura 2.1: QQplot entre a amostra $H_i = \mathbb{H}(x_i, y_i)$ da BIPIT H associada distribuição normal bivariada do Exemplo 2.1 e uma amostra $\{U_1, \dots, U_{100}\}$ com distribuição $U(0, 1)$.

contando que a cópula contém toda informação da estrutura de dependência de (X, Y) . Por sua importância para o estudo de estruturas de dependência é necessário nomear a v.a. H pela seguinte definição.

Definição 2.1. *Sejam X, Y v.a. com f.d.a. F e G respectivamente e (X, Y) vetor aleatório bidimensional com f.d.a. \mathbb{H} . Então $U = F(X)$ ($V = G(Y)$) e $H = \mathbb{H}(X, Y)$ são denominadas respectivamente PIT e BIPIT.*

O Exemplo 2.2 evidencia que a BIPIT independe das marginais, fato que se observa pelo Teorema de Sklar (1.1). Deste modo assumiremos sem perda de generalidade o par aleatório (X, Y) com marginais $U(0, 1)$ ao trabalharmos com conceito de BIPIT.

Exemplo 2.2. *Sejam as BIPIT's H, H_1 associadas aos vetores aleatórios (U, V) e (X_1, Y_1) , respectivamente, com f.d.a. independência \mathbb{H} . Sejam U, V com distribuição marginal $U(0, 1)$ e X_1, Y_1 com marginais F_1 e G_1 exponenciais $\lambda = 2$ e $\lambda = 10$ respectivamente. Conforme QQ-plot da Figura 2.2, H e H_1 são identicamente distribuídas, evidenciando que a estrutura de dependência da BIPIT é invariante sob as marginais.*

Os seguintes teoremas apresentam uma expressão para o cálculo da função distribuição

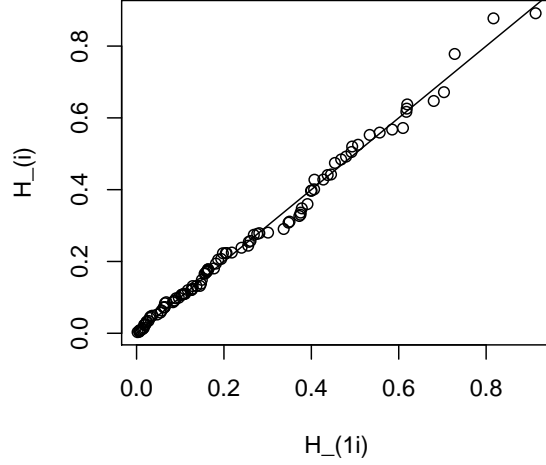


Figura 2.2: QQplot das amostras das BIPITS H e H_1 : $\{H_1, \dots, H_n\}$ e $\{H_{11}, \dots, H_{1n}\}$ respectivamente, $n = 100$, onde $H_i = \mathbb{H}(u_i, v_i) = u_i v_i$ sendo u_i, v_i amostras de $U(0, 1)$ e $H_{1i} = \mathbb{H}(x_{1i}, y_{1i}) = F_1(x_{1i})G_1(y_{1i})$ sendo x_{1i} com distribuição $exp(2)$ e y_{1i} com distribuição $exp(10)$.

K da BIPIT $H = C(X, Y)$.

Teorema 2.2. *Seja (X, Y) um vetor aleatório com f.d.a. C e seja K a f.d.a. da v.a. BIPIT, $H \equiv C(X, Y)$. Então*

$$K(v) = \int_0^1 E(\mathbb{I}\{C(x, Y) \leq v\} | X = x) dx \quad (2.1)$$

Demonstração.

$$\begin{aligned} K(v) &= P(C(X, Y) \leq v) = E[\mathbb{I}\{C(X, Y) \leq v\}] \\ &= E(E[\mathbb{I}\{C(x, Y) \leq v\} | X = x]) \\ &= \int_0^1 E(\mathbb{I}\{C(x, Y) \leq v\} | X = x) dx. \end{aligned} \quad (2.2)$$

□

Teorema 2.3. *Sejam $C_x(y) \equiv C(x, y)$ estritamente crescente e inversível e $y_{x,v} \equiv C_x^{-1}(v)$, $\forall 0 \leq v \leq x$, $0 < x < 1$. Então,*

$$K(v) = v + \int_v^1 C^*(x, y_{x,v}) dx \equiv v - \lambda(v) \quad (2.3)$$

onde, $C^*(x, y) \equiv \frac{\partial C(x, y)}{\partial x}$ e $\lambda(v) \equiv - \int_v^1 C^*(x, y_{x,v}) dx$.

Demonstração. Pela equação (2.1) $K(v)$ pode ser escrito como

$$K(v) = \underbrace{\int_0^v E(\mathbb{I}\{C(x, Y) \leq v\} | X = x) dx}_{(1)} + \underbrace{\int_v^1 E(\mathbb{I}\{C(x, Y) \leq v\} | X = x) dx}_{(2)}.$$

De acordo com o Teorema 1.7 toda cópula C é limitada superiormente pela cópula maximal $M(x, y) = \min(x, y)$. Então, $\forall 0 \leq x \leq v$, $C(x, y) \leq v$, logo $\mathbb{I}\{C(x, Y) \leq v\} = 1$ para a região de integração da integral (1). Desta maneira a integral (1) é igual a v . A integral (2) pode ser reescrita como

$$\int_v^1 P(C(x, Y) \leq v | X = x) dx$$

Conforme hipótese, $C(x, y_{x,v}) = v$ para x fixo, e C é marginalmente não decrescente, logo temos a equivalência $C(x, Y) \leq v \Leftrightarrow Y \leq y_{x,v}$. Então

$$\begin{aligned} \int_v^1 P(C(x, Y) \leq v | X = x) dx &= \int_v^1 P(Y \leq y_{x,v} | X = x) dx \\ &= \int_v^1 \left(\int_0^{y_{x,v}} c(x, y) dy \right) dx \\ &= \int_v^1 C^*(x, y_{x,v}) dx \end{aligned}$$

onde c é a densidade de C , $c(u, v) \equiv \frac{\partial^2 C}{\partial u \partial v}(u, v)$. □

Exemplo 2.3. Cópula Produto: $\Pi(x, y) = xy$; $x, y \in \mathbf{I}$

Sem a aplicação do Teorema 2.3 temos,

$$\begin{aligned} K_{\Pi}(v) &= P(XY \leq v) = \int_0^1 P\left(X \leq \frac{v}{y}\right) dy \\ &= \int_0^v 1 dy + \int_v^1 \frac{v}{y} dy = v - v \ln v \end{aligned}$$

Usando o Teorema 2.3 temos $\Pi^*(x, y) = y$ e $\Pi(x, y_{x,v}) = v$, então $y_{x,v} = \frac{v}{x}$. Logo,

$$\int_v^1 \Pi^*(x, y_{x,v}) dx = \int_v^1 y_{x,v} dx = \int_v^1 \frac{v}{x} dx = -v \ln v$$

Portanto, $K_{\Pi}(v) = v - v \ln v$.

Exemplo 2.4. Cópula Maximal: $M(x, y) = \min(x, y)$; $x, y \in \mathbf{I}$

Observe que $K_M(v)$ deve ser calculada através de uma mistura de cópulas, $C_{\alpha}(x, y) = (1 - \alpha)xy + \alpha \min(x, y)$, $0 \leq \alpha \leq 1$, chamada Família de Fréchet, porque a Cópula M não satisfaz a hipótese do Teorema 2.3 pois $M_x(y)$ não é estritamente crescente. Quando $\alpha \rightarrow 1$, temos $K_{C_{\alpha}}(v) \rightarrow v$ e segue que $K_M(v) = v$.

No Teorema 2.3 $y_{x,v} \equiv C_x^{-1}(v)$, $0 \leq v \leq x$, $0 < x < 1$ fixo, representa a segunda coordenada dos pontos do conjunto de nível v da cópula C dado por

$$\{(x, y_{x,v}) \in \mathbf{I}^2 | C(x, y_{x,v}) = v\}$$

Conforme Nelsen (1999) [30], para uma cópula Arquimediana este conjunto de nível v consiste dos pontos $(x, y_{x,v})$ pertencentes a curva

$$\phi(x) + \phi(y_{x,v}) = \phi(v), \quad (x, y_{x,v}) \in \mathbf{I}^2, \quad v > 0 \quad (2.4)$$

A curva (2.4) pode ser reescrita expressando $y_{x,v}$ em função de x por

$$y_{x,v} = L_v(x) = \phi^{[-1]}(\phi(v) - \phi(x)) = \phi^{-1}(\phi(v) - \phi(x)) \quad (2.5)$$

sendo a última igualdade garantida pela Definição 1.15.

Para $v = 0$, o conjunto $\{(x, y_{x,v}) \in \mathbf{I}^2 | C(x, y_{x,v}) = 0\}$ é chamado *conjunto zero* de C e denotado por $Z(C)$. Para as cópulas Arquimedianas com gerador estrito, por exemplo para família de Frank (Tabela 1.1), $Z(C)$ é composto pelos dois segmentos de reta $\{0\} \times \mathbf{I}$ e $\mathbf{I} \times \{0\}$ de acordo com Definição 1.5. Para as cópulas Arquimedianas com gerador não estrito, por exemplo para família de Clayton $\theta \geq 0$ (Tabela 1.1), $Z(C)$ tem área positiva e é limitado pelos segmentos $\{0\} \times \mathbf{I}$ e $\mathbf{I} \times \{0\}$ e pela curva $\phi(x) + \phi(y_{x,v}) = \phi(0)$, isto é, $y_{x,v} = L_0(x)$ chamada *curva zero* de C .

O Corolário a seguir fornece ferramentas para facilitar o cálculo da f.d.a. K de BIPIT associada a uma classe particular de cópulas, as cópulas Arquimedianas que serão o foco desta dissertação. Em sua demonstração é usado o conceito de curva de nível de cópulas Arquimedianas.

Corolário 2.4. *Seja C uma cópula Arquimediana gerada por $\phi \in \Omega$. A f.d.a. K da v.a. BIPIT, $H \equiv C(X, Y)$, é dada por:*

$$K(v) = v - \frac{\phi(v)}{\phi'(v^+)} \quad (2.6)$$

onde, $\phi'(v^+)$, denota a derivada a direita de ϕ em v e Ω o espaço dos geradores ϕ .

Demonstração. Utilizaremos o Teorema 2.3 para demonstrar que $\lambda(v) = \frac{\phi(v)}{\phi'(v)}$ usando os conceitos de curva de nível e conjunto de nível de uma cópula Arquimediana C .

$$C^*(x, y) \equiv \frac{\partial C(x, y)}{\partial x} = \left(\phi'(x) + \frac{\partial \phi(y)}{\partial x} \right) \phi^{[-1]'}(\phi(x) + \phi(y)) = \phi'(x) \phi^{[-1]'}(\phi(x) + \phi(y))$$

então,

$$C^*(x, y_{x,v}) = \phi'(x) \phi^{[-1]'}(\phi(x) + \phi(y_{x,v})) \stackrel{(2.5)}{=} \phi'(x) \phi^{[-1]'}(\phi(v)) = \frac{\phi'(x)}{\phi'(\phi^{-1}(\phi(v)))} = \frac{\phi'(x)}{\phi'(v)}.$$

para $v > 0$.

Para o caso $v = 0$ temos $(x, y_{x,v})$, os pontos pertencentes ao *conjunto zero* da cópula C , dados não somente pela curva $L_0(x)$ da equação (2.5), mas pertencentes ao conjunto

$$\underbrace{\{(x, y_{x,v}) | x \in \mathbf{I} \text{ e } y_{x,v} = 0\}}_{(1)} \cup \underbrace{\{(x, y_{x,v}) | y \in \mathbf{I} \text{ e } x = 0\}}_{(2)} \cup \underbrace{\{(x, y_{x,v}) | \phi(x) + \phi(y_{x,v}) \geq \phi(0)\}}_{(3)}$$

Se $(x, y_{x,v}) \in (1)$, temos

$$\begin{aligned} C^*(x, y_{x,v}) &= \phi'(x)\phi^{[-1]'}(\phi(x) + \phi(y_{x,v})) = \phi'(x)\phi^{[-1]'}(\phi(x) + \phi(0)) \\ &= \frac{\phi'(x)}{\phi'(\phi^{[-1]}(\phi(x) + \phi(0)))} = \frac{\phi'(x)}{\phi'(0)} \end{aligned}$$

usando a Definição 1.15. Por simetria, para o conjunto (2) segue idêntico resultado.

E se $(x, y_{x,v}) \in (3)$, temos

$$\phi(x) + \phi(y_{x,v}) \geq \phi(x) + \phi(0) - \phi(x) = \phi(0)$$

então,

$$\begin{aligned} C^*(x, y_{x,v}) &= \frac{\phi'(x)\phi^{[-1]'}(\phi(x) + \phi(y_{x,v}))}{\phi'(x)} \\ &= \frac{\phi'(x)}{\phi'(\phi^{[-1]}(\phi(x) + \phi(y_{x,v})))} \\ &= \frac{\phi'(x)}{\phi'(0)} \end{aligned}$$

usando a Definição 1.15.

Deste modo segue que,

$$\lambda(v) = - \int_v^1 \frac{\phi'(x)}{\phi'(v)} dx = \frac{\phi(v)}{\phi'(v)}, \quad v \in \mathbf{I}.$$

<i>cpula</i>	$\phi_\theta(v)$	$K_\theta(v)$
Clayton	$\frac{v^{-\theta}-1}{\theta}$	$v + \frac{v}{\theta}(1 - v^\theta)$
Gumbel	$(-\ln v)^\theta$	$v - \frac{v \ln v}{\theta}$
Frank	$-\ln \frac{e^{-\theta v}-1}{e^{-\theta}-1}$	$v + \frac{\theta e^{-\theta v} \ln \frac{e^{-\theta v}-1}{e^{-\theta}-1}}{e^{-\theta v}-1}$
4.2.12	$(\frac{1}{v} - 1)^\theta$	$v + \frac{v^2(\frac{1}{v}-1)}{\theta}$

Tabela 2.1: Cópulas Arquimedianas da Tabela 1.1 com os respectivos geradores ϕ e funções distribuição K .

Logo, para $v \in \mathbf{I}$ temos $K(v) = v - \lambda(v) = v - \frac{\phi(v)}{\phi'(v)}$ considerando ϕ diferenciável em v . Se ϕ não for diferenciável em v , temos $\phi'(v) \equiv \phi'(v^+)$, pois $v \leq x$ por hipótese do Teorema 2.3.

Concluindo, $K(v) = v - \frac{\phi(v)}{\phi'(v^+)}$ para qualquer gerador ϕ de cópula Arquimediana C . \square

Exemplo 2.5. A Tabela 2.1 apresenta as funções distribuição K , calculadas por meio do Corolário 2.4, das cópulas Arquimedianas utilizadas nesta dissertação.

Como visto no Exemplo 1.7, a cópula de Clayton C_{-1} é a cópula minimal W , então substituindo $\theta = -1$ na função distribuição K de Clayton temos

$$K_W(v) = 1, \forall v \in [0, 1]$$

Do mesmo modo usando $\theta = 1$ na função distribuição K de Gumbel temos

$$K_{\Pi}(v) = v - v \ln v, \forall v \in [0, 1]$$

2.1 Propriedades da função K

Esta seção apresenta propriedades para K relativas a uma ordenação estocástica definida a seguir. Esta ordenação será abordada novamente nos Capítulos 3 e 4 por estar relacionada ao Kendall Plot.

Definição 2.5. A ordem estocástica ordinária entre duas v.a. contínuas X_1 e X_2 com f.d.a. F_1 e F_2 respectivamente é denotada por $X_1 \prec_{st} X_2$ equivalente a $F_1(x) \geq F_2(x), \forall x \in \mathbb{R}$.

Definição 2.6. A ordem estocástica de Kendall entre dois vetores aleatórios contínuos (X_1, Y_1) e (X_2, Y_2) com f.d.a. \mathbb{H}_1 e \mathbb{H}_2 respectivamente e denotada como

$$(X_1, Y_1) \prec_k (X_2, Y_2) \text{ ou } \mathbb{H}_1(X_1, Y_1) \prec_{st} \mathbb{H}_2(X_2, Y_2)$$

representa a ordenação estocástica entre as BIPIT's $H_1 \equiv \mathbb{H}_1(X_1, Y_1)$ e $H_2 = \mathbb{H}_2(X_2, Y_2)$, ou seja, $H_1 \prec_{st} H_2$.

Esta ordem estocástica é chamada *Kendall* devido a estar associada ao coeficiente populacional da medida de associação tau de Kendall estudada no próximo capítulo. Também existe uma associação entre a função K e a medida tau de Kendall, que será apresentada no próximo capítulo, por isso K é chamada frequentemente de *função distribuição de Kendall* (Nelsen (2003) [32]).

O teorema abaixo valida os limites de Frechét também para a ordem estocástica de Kendall.

Teorema 2.7. Sejam K a f.d.a. da v.a. BIPIT $H \equiv C(X, Y)$ e a função $\lambda(v)$, $v \in \mathbf{I}$, definida pelo Teorema 2.3.

$$K_M(v) \leq K(v) \leq K_W(v), \text{ ou seja, } v \leq K(v) \leq 1, v \in \mathbf{I}; \quad (2.7)$$

$$\lambda_W(v) \leq \lambda(v) \leq \lambda_M(v), \text{ ou seja, } v - 1 \leq \lambda(v) \leq 0, v \in \mathbf{I}. \quad (2.8)$$

Demonstração. Equação (2.7): É trivial que $K(v) \leq 1$ com $K(v)$ f.d.a. Sendo $C(x, y)$ não decrescente marginalmente, temos $C^*(x, y)$ positiva, logo $\lambda(v) \equiv - \int_v^1 C^*(x, y_{x,v}) dx \leq 0$, deste modo $K(v) = v - \lambda(v) \geq v$.

Equação (2.8): Provada equação (2.7) temos

$$v \leq K(v) \leq 1 \Leftrightarrow v \leq v - \lambda(v) \leq 1 \Leftrightarrow v - 1 \leq \lambda(v) \leq 0$$

□

Corolário 2.8. Os limites de Frechét são válidos para a ordem estocástica de Kendall

$$W \prec_k C \prec_k M$$

Exemplo 2.6. As Figuras 2.3 e 2.4 ilustram o Teorema 2.7 e o Corolário 2.8 mostrando $K_M \leq K \leq K_W$ e $\lambda_W(v) \leq \lambda(v) \leq \lambda_M(v)$ para as funções K das cópulas de Gumbel e Clayton. Além de também ilustrarem a ordenação estocástica de Kendall para as famílias de Clayton e Gumbel com

$$\theta \leq \theta^* \Rightarrow K_{C_\theta} \geq K_{C_{\theta^*}}, \text{ ou seja } C_\theta \prec_k C_{\theta^*}.$$

Tal ordenação é confirmada através das formas analíticas das funções K apresentadas no Exemplo 2.5 para a família de Clayton e Gumbel.

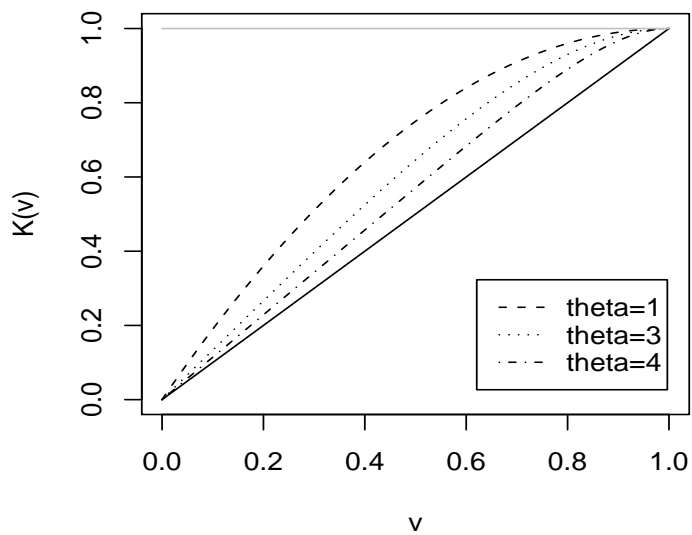
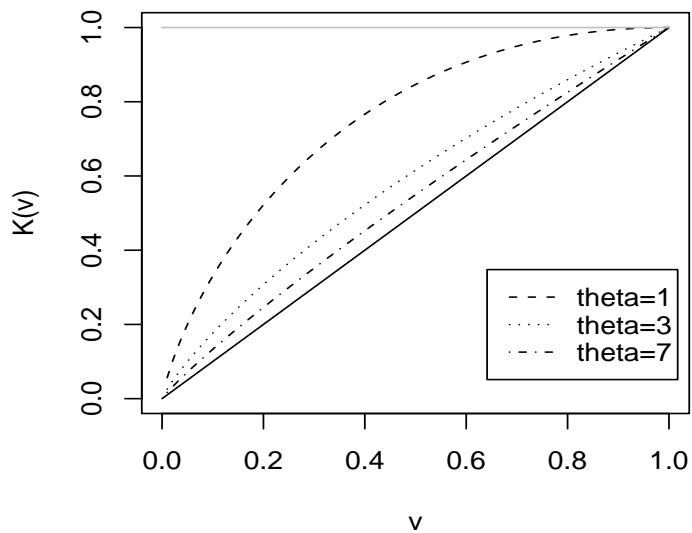


Figura 2.3: Funções distribuição de Kendall das cópulas M (linha cheia preta), W (linha cheia cinza) e família de Gumbel para $\theta = \{1, 3, 7\}$ (primeira figura); família de Clayton para $\theta = \{1, 3, 4\}$ (segunda figura).

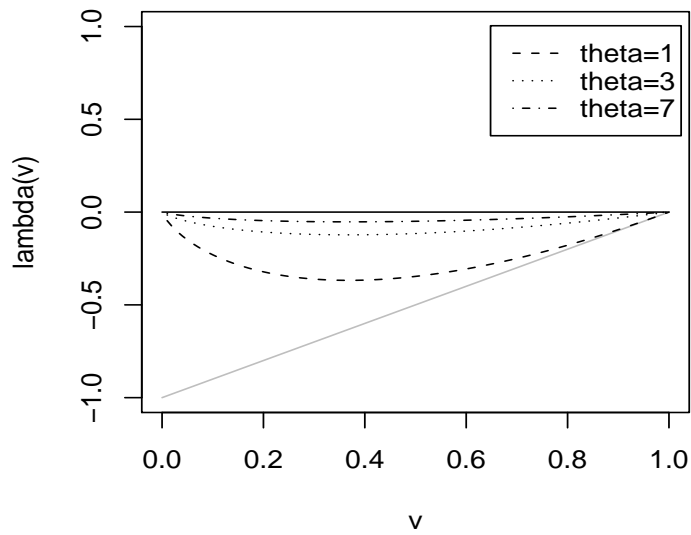
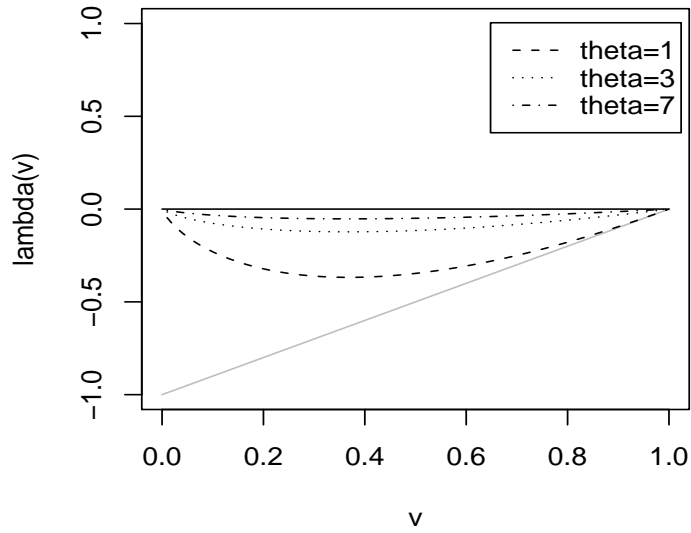


Figura 2.4: Funções $\lambda(v)$, $v \in \mathbf{I}$, das cópulas M (linha cheia preta), W (linha cheia cinza) e família de Gumbel para $\theta = \{1, 3, 7\}$ (primeira figura); família de Clayton para $\theta = \{1, 3, 4\}$.

Capítulo 3

Tau de Kendall

Este capítulo explorará a cópula como objeto de estudo de dependência ou associação entre duas variáveis por meio do coeficiente tau de Kendall. Porém, primeiramente, será apresentado o conceito e algumas propriedades desta medida de associação.

3.1 Concordância

Informalmente, um par de variáveis aleatórias é *concordante* se grandes valores de uma variável estão associados a grandes valores da outra variável ou se pequenos valores de uma variável estão associados a pequenos valores da outra variável. E o par é *discordante* caso contrário. Formalmente, *concordância* é definida por:

Definição 3.1. *Sejam (x, y) e (\tilde{x}, \tilde{y}) duas observações de um vetor (X, Y) de v.a. contínuas. Diz-se que estas observações são*

$$\textit{concordantes} \quad \textit{se} \quad \frac{(y - \tilde{y})}{(x - \tilde{x})} > 0$$

$$\textit{discordantes} \quad \textit{se} \quad \frac{(y - \tilde{y})}{(x - \tilde{x})} < 0$$

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

Definição 3.2. *Sejam (X_1, Y_1) e (X_2, Y_2) vetores aleatórios i.i.d. com a mesma f.d.a. \mathbb{H} . Então o coeficiente populacional Tau de Kendall de um vetor (X, Y) de v.a. contínuas com f.d.a. \mathbb{H} é definido por*

$$\tau \equiv P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (3.1)$$

Teorema 3.3. *Sejam $(X_1, Y_1), (X_2, Y_2)$ vetores independentes de v.a. contínuas com função distribuição conjunta $\mathbb{H}_1, \mathbb{H}_2$ respectivamente, com marginais comuns F (de X_1, X_2) e G (de Y_1, Y_2), C_1, C_2 as respectivas cópulas de (X_1, Y_1) e (X_2, Y_2) e seja*

$$\mathbb{Q} \equiv P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

denominada função de concordância, então

$$\mathbb{Q} \equiv \mathbb{Q}(C_1, C_2) \equiv 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1$$

Demonstração. Como $P[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - P[(X_1 - X_2)(Y_1 - Y_2) > 0]$, a função \mathbb{Q} pode ser reescrita por $\mathbb{Q} = 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1$, onde

$$P[(X_1 - X_2)(Y_1 - Y_2) > 0] = \underbrace{P(X_1 > X_2, Y_1 > Y_2)}_{(1)} + \underbrace{P(X_1 < X_2, Y_1 < Y_2)}_{(2)}$$

$$\begin{aligned} (1) &= P(X_2 < X_1, Y_2 < Y_1) = \iint_{\mathbb{R}^2} P(X_2 < x, Y_2 < y) dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) = \iint_{I^2} C_2(u, v) dC_1(u, v) \end{aligned}$$

$$\begin{aligned} (2) &= P(X_2 > X_1, Y_2 > Y_1) = \iint_{\mathbb{R}^2} P(X_2 > x, Y_2 > y) dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} [1 - F(x) - G(y) + C_2(F(x), G(y))] dC_1(F(x), G(y)) \\ &= \iint_{I^2} [1 - u - v + C_2(u, v)] dC_1(u, v) = \iint_{I^2} C_2(u, v) dC_1(u, v) \end{aligned}$$

A última igualdade segue de U, V serem $U(0, 1)$, logo $E(U) = E(V) = \frac{1}{2}$. Deste modo,

$$P[(X_1 - X_2)(Y_1 - Y_2) > 0] = 2 \iint_{I^2} C_2(u, v) dC_1(u, v)$$

Então segue que

$$\mathbb{Q} = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1$$

□

O próximo resultado estabelece uma ligação entre a medida tau de Kendall entre duas v.a. X e Y e a cópula C associada a estas variáveis por meio da *função de concordância*.

Teorema 3.4. *Sejam X, Y v.a. contínuas com cópula C e seja $H \equiv C(F(X), G(Y))$ a sua v.a. BIPIT com f.d.a. K . O coeficiente populacional tau de Kendall é dado por*

$$\tau \equiv \mathbb{Q}(C, C) \equiv 4 \iint_{I^2} C(u, v) dC(u, v) - 1 = 4E(C(U, V)) - 1 \quad (3.2)$$

e pode ser reescrito por

$$\tau = 3 - 4 \int_0^1 K(t) dt \quad (3.3)$$

Demonstração. A prova da equação (3.2) segue da Definição 3.2 e do Teorema 3.3.

Da equação (3.2) temos

$$\tau = 4E(H) - 1 = 4 \int_0^1 1 - K(t) dt - 1 = 3 - 4 \int_0^1 K(t) dt$$

Como a v.a. BIPIT $H \in [0, 1]$ não assume valores negativos, então $E(H) = \int_0^1 1 - K(t) dt$ (veja Teorema .2 no apêndice 1). □

Assim como a função distribuição K de uma BIPIT associada a uma cópula Arquimediana é expressa em função do gerador ϕ , tau de Kendall também é expresso em função de ϕ para cópulas Arquimedianas.

Corolário 3.5. *Sejam X, Y v.a. contínuas com cópula Arquimediana C gerada por ϕ . O coeficiente populacional τ de Kendall é dado por*

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \quad (3.4)$$

Demonstração. Seja a BIPIT $H \equiv C(F(X), G(Y))$ com f.d.a. K . Pelo Teorema 3.4, equação (3.3)

$$\tau \stackrel{(2.6)}{=} 3 - 4 \int_0^1 t - \frac{\phi(t)}{\phi'(t)} dt = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt$$

□

Exemplo 3.1. Se C_θ é um membro da família de Frank (Tabela 1.1), devido a complexidade de resolução da integral na equação (3.4), o cálculo de τ torna-se mais adequado via integração numérica. Para as demais famílias Arquimedianas citadas nesta dissertação a integral na equação (3.4) é de simples resolução, sendo então possível encontrar uma função de θ que expresse a medida de associação τ de Kendall, tais expressões encontram-se na Tabela 3.1.

<i>família Arquimediana</i>	$\tau = f(\theta)$
<i>Clayton</i>	$\tau = \frac{\theta}{\theta+2}$
<i>Gumbel</i>	$\tau = \frac{\theta-1}{\theta}$
4.2.12	$\tau = 1 - \frac{2}{3\theta}$

Tabela 3.1: Medida de associação τ de Kendall expressa em função do parâmetro θ para algumas famílias de cópulas Arquimedianas.

Exemplo 3.2. Desde que a cópula M possui como suporte a diagonal $v = u$ em I^2 e possui marginais $U(0, 1)$, segue que se g é função integrável cujo domínio é I^2 , então

$$\iint_{I^2} g(u, v) dM(u, v) = \int_0^1 g(u, u) du$$

Similarmente, desde que W possui como suporte a diagonal secundária $v = 1 - u$, temos

$$\iint_{I^2} g(u, v) dW(u, v) = \int_0^1 g(u, 1 - u) du$$

Logo, temos

$$\tau_{\Pi} = \mathbb{Q}(\Pi, \Pi) = 4 \iint_{I^2} uv d\Pi(u, v) - 1 = 4 \int_0^1 uv du dv - 1 = 0$$

$$\tau_M = \mathbb{Q}(M, M) = 4 \iint_{I^2} \min(u, v) dM(u, v) - 1 = 4 \int_0^1 u du - 1 = 1$$

$$\tau_W = \mathbb{Q}(W, W) = 4 \iint_{I^2} \max(u + v - 1, 0) dW(u, v) - 1 = 4 \int_0^1 0 du - 1 = -1$$

A cópula M é chamada *cópula maximal* por estabelecer relação de *dependência perfeitamente positiva* entre as variáveis X, Y , como pode ser evidenciado pelo valor $\tau_M = 1$, ou seja, probabilidade 1 de que X e Y sejam concordantes. O contrário acontece para a cópula W , chamada *cópula minimal* por estabelecer relação de *dependência perfeitamente negativa* entre as variáveis X, Y , $\tau_W = -1$, ou seja, a discordância das variáveis X e Y é certa. O Teorema a seguir mostra a unicidade da relação de dependência perfeita negativa e positiva com respeito às cópulas.

Teorema 3.6. *O coeficiente populacional tau de Kendal associado a cópula C pela equação (3.2) é igual a 1, se e somente se, a cópula C é M . E $\tau = -1$, se e somente se, $C = W$.*

Demonstração. (\Leftarrow) A prova que τ_W e τ_M são iguais a -1 e 1 respectivamente segue do Exemplo 3.2.

(\Rightarrow) Agora considerando X, Y, X', Y' com distribuição $U(0, 1)$ e sejam (X, Y) e (X', Y') *i.i.d.* com f.d.a. cópula C . A seguir a demonstração será dada como consequência de uma sequência de afirmações:

a) $\tau = 1 \Rightarrow C = M$

1. $\tau = 1 \Rightarrow X = Y$ com probabilidade 1;
2. $X = Y \Rightarrow C(x, y) = \min(x, y)$

b) $\tau = -1 \Rightarrow C = W$

1. $\tau = -1 \Rightarrow Y = 1 - X$ com probabilidade 1;
2. $Y = 1 - X \Rightarrow C(x, y) = \max(x + y - 1, 0)$

Demonstração a):

$$\begin{aligned}
\tau = 1 &\Rightarrow P[(X - X')(Y - Y') > 0] = 1 \\
&\Rightarrow \iint_{I^2} [P(X > x, Y > y) + P(X < x, Y < y)] dC(x, y) = 1 \\
&\Rightarrow \iint_{I^2} [1 - F_X(x) - F_Y(y) + 2C(x, y)] dC(x, y) = 1 \\
&\Rightarrow \iint_{I^2} [2C(x, y) + 1 - x - y] dC(x, y) = 1 \\
&\Rightarrow \iint_{I^2} [2C(x, y) + 1 - x - y] dC(x, y) = 1 \Rightarrow 2C(x, y) + 1 - x - y = 1 \\
&\Rightarrow C(x, y) = \frac{x + y}{2}
\end{aligned}$$

Aplicando os limites de Fréchet,

$$C(x, y) = \frac{x + y}{2} \leq \min(x, y) \Leftrightarrow X = Y \text{ com probabilidade } 1$$

logo,

$$C(x, y) = P(X \leq x, Y \leq y, X = Y) = P(X \leq \min(x, y)) = \min(x, y)$$

Demonstração b):

$$\begin{aligned}
\tau = -1 &\Rightarrow P[(X - X')(Y - Y') < 0] = 1 \\
&\Rightarrow \iint_{I^2} [P(X > x, Y < y) + P(X < x, Y > y)] dC(x, y) = 1 \quad (3.5)
\end{aligned}$$

Como,

$$\begin{aligned}
P(X > x) &= P(X > x, Y < y) + P(X > x, Y > y) \\
\Rightarrow P(X > x, Y < y) &= (1 - x) - \bar{C}(x, y) = y - C(x, y)
\end{aligned}$$

e,

$$\begin{aligned}
P(X < x) &= P(X < x, Y < y) + P(X < x, Y > y) \\
&\Rightarrow P(X < x, Y > y) = x - C(x, y)
\end{aligned}$$

Substituindo na equação (3.5), temos

$$x + y - 2C(x, y) = 1 \Rightarrow C(x, y) = \frac{x + y - 1}{2}$$

Por limites de Fréchet,

$$\max(x + y - 1, 0) \leq \frac{x + y - 1}{2} \Leftrightarrow Y = 1 - X \text{ com probabilidade } 1$$

logo,

$$\begin{aligned}
C(x, y) &= P(X < x, Y < y, Y = 1 - X) = P(1 - y \leq X \leq x) \\
&= F(x) - F(1 - y) = \begin{cases} x + y - 1, & \text{se } 0 \leq 1 - y \leq x \leq 1 \\ 0, & \text{se caso contrario.} \end{cases}
\end{aligned}$$

com F denotando a f.d.a. da v.a. X .

Desta forma temos, $C(x, y) = \max(x + y - 1, 0)$. □

Exemplo 3.3. O caso particular $\theta = -1$ da família de Clayton é a cópula minimal, pois para este caso $\tau = -1$ conforme Exemplo 3.1.

No início deste capítulo foi visto que o tau de Kendall é uma medida de associação expressa em termos da concordância, agora veremos que este coeficiente também é uma *medida de concordância* que apresenta as propriedades desta.

Definição 3.7. Uma medida numérica κ de associação entre duas v.a. contínuas X e Y cuja cópula é C é uma medida de concordância se satisfaz as seguintes propriedades

1. κ é definida para todo par X, Y de v.a. contínuas;
2. $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$ e $\kappa_{X,-X} = -1$;

3. $\kappa_{X,Y} = \kappa_{Y,X}$;
4. Se X, Y são independentes, então $\kappa_{X,Y} = 0$;
5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$;
6. Se C_1 e C_2 são cópulas tais que $C_1(u, v) \leq C_2(u, v)$, $\forall (u, v) \in \mathbf{I}^2$, então $\kappa_{C_1} \leq \kappa_{C_2}$;
7. Se $\{(X_n, Y_n)\}$ é uma sequência de v.a. contínuas com cópulas C_n , e se $\{C_n\}$ converge ponto a ponto a C , então $\lim_{n \rightarrow \infty} \kappa_{C_n} = \kappa_C$.

Teorema 3.8. Se X, Y são v.a. contínuas com cópula C , então o coeficiente populacional τ de Kendall satisfaz as propriedades da Definição 3.7, enquadrando-se como medida de concordância.

Demonstração. Cada item da Definição 3.7 é demonstrado separadamente. 2. Como τ é definido na equação (3.1) por uma diferença de probabilidades, então $-1 \leq \tau \leq 1$ e utilizando demonstração do Teorema 3.9 temos $\tau_{X,X} = \tau_M$ e $\tau_{X,-X} = \tau_W$ e por sua vez, pelo Teorema 3.6, temos $\tau_M = 1$ e $\tau_W = -1$.

3. Como $C_{X,Y} = C_{Y,X}$, e por (3.1) temos τ como função da cópula C , então $\tau_{X,Y} = \tau_{Y,X}$.

4. Vide Exemplo 3.2.

5. Seja $W = -X$, pelo Teorema 1.14 temos que $C_{W,Y}$ é dada em função de $C_{X,Y}$ por $C_{W,Y}(u, v) = v - C_{X,Y}(1 - u, v)$, logo

$$\tau_{W,Y} = \mathbb{Q}(C_{W,Y}, C_{W,Y}) = 4 \iint_{\mathbf{I}^2} -v + C_{X,Y}(1 - u, v) dC_{X,Y}(1 - u, v) - 1$$

Aplicando a transformação $z = 1 - u$,

$$4 \int_0^1 \int_0^1 v - C_{X,Y}(z, v) dC_{X,Y}(z, v) - 1 = -4 \iint_{\mathbf{I}^2} C_{X,Y}(z, v) dC_{X,Y}(z, v) + 1 = -\tau_{X,Y}$$

A última igualdade deve-se ao fato de que $Z = 1 - U$ é $U(0, 1)$, pois U é $U(0, 1)$.

Analogamente temos $\tau_{X,-Y} = -\tau_{X,Y}$.

6. Segue da definição de τ dada pelo Teorema 3.4 e pela função concordância \mathbb{Q} (Teorema 3.3) ser não decrescente em cada argumento.

7. Seja $\tau \equiv \mathbb{Q}(C, C)$ conforme Teorema 3.4, como $C_n \rightarrow C$ e C_n é limitada pela cópula Maximal M , usando o Teorema da Convergência Dominada de Lebesgue temos

$$\lim_{n \rightarrow \infty} \tau_n = \lim_{n \rightarrow \infty} 4 \iint_{\mathbf{I}^2} C_n(u, v) dC_n(u, v) - 1 = 4 \iint_{\mathbf{I}^2} C(u, v) dC(u, v) - 1 = \mathbb{Q}(C, C) = \tau$$

□

Pelo Exemplo 3.2 foi visto que $\tau_W = -1$ e $\tau_M = 1$. Mas o resultado abaixo nos permite verificar que se uma v.a. Y é uma função crescente da v.a. X , existindo então uma dependência perfeita positiva entre estas variáveis então caracteriza-se $C_{X,Y} = M$. De maneira análoga, para Y função decrescente de X caracteriza-se $C_{X,Y} = W$. Por meio da medida de concordância τ verifica-se este resultado.

Teorema 3.9. *Seja κ medida de concordância para v.a. contínuas X e Y .*

1. *Se Y é quase certamente uma função crescente de X , então $\kappa_{X,Y} = \kappa_M = 1$;*
2. *Se Y é quase certamente uma função decrescente de X , então $\kappa_{X,Y} = \kappa_W = -1$;*

Demonstração. Primeiro, por intermédio de τ , como um particular exemplo de medida de concordância, se provam que se Y é quase certamente uma função crescente de X sendo (X, Y) com cópula C , então $C = M$ e que se Y é quase certamente uma função decrescente de X , então $C = W$.

Sejam (X, Y) e (X_1, Y_1) duas realizações i.i.d. com f.d.a. \mathbb{H} , f não decrescente, $X = f(Y)$ e $X_1 = f(Y_1)$ com probabilidade 1.

Como $P[(X - X_1)(Y - Y_1) \geq 0] + P[(X - X_1)(Y - Y_1) < 0] = 1$, pela equação (3.1) temos $\tau = 2P[(X - X_1)(Y - Y_1) \geq 0] - 1$, onde

$$\begin{aligned} P[(X - X_1)(Y - Y_1) \geq 0] &= P[(X - X_1)(Y - Y_1) \geq 0, Y = f(X), Y_1 = f(X_1)] \\ &= P[(X - X_1)(f(X) - f(X_1)) \geq 0] = 1 \end{aligned}$$

A última igualdade deve-se ao fato de f não decrescente.

Logo, $\tau = 2 \times 1 - 1 = 1$ e pelo Teorema 3.6 temos $C = M$. Considerando f não crescente, do mesmo modo, obtém-se $\tau = -1$, e conseqüentemente, $C = W$.

Pelo item 6. da Definição 3.7 e pelos limites de Fréchet, $W \leq C \leq M$, temos $\kappa_W \leq \kappa \leq \kappa_M$ e como pelo item 2., $-1 \leq \kappa \leq 1$, então $\kappa_M = 1$ e $\kappa_W = -1$.

□

A medida de concordância tau de Kendall nomeia o método gráfico de ajuste de cópulas apresentado no próximo capítulo, chamado Kendall Plot. Embora o Kendall Plot seja detalhado no próximo capítulo, há a necessidade de uma breve introdução de seu conceito neste capítulo para a justificativa de seu nome. O Kendall Plot pode ser interpretado como um teste gráfico de cópulas baseado em um QQplot entre duas v.a. BIPIT's $H \equiv C(U, V)$ e $H_0 \equiv C_0(U, V)$ com cópulas C desconhecida e C_0 conhecida. Se o QQplot diagnosticar H e H_0 identicamente distribuídas, então a BIPIT H está associada a cópula C_0 . Caso contrário, o Kendall Plot revela uma informação a respeito entre a relação dos coeficientes populacionais τ de Kendall associados às cópulas C e C_0 . Esta relação entre τ de Kendall e Kendall Plot se deve ao fato da associação entre τ e K por τ ser uma função de K (Teorema 3.4). Então pela equação (3.3) temos,

$$K(w) \geq K_0(w), \forall 0 \leq w \leq 1 \Rightarrow \tau \leq \tau_0 \quad (3.6)$$

ou então, conforme Definição 2.6 temos,

$$(X, Y) \prec_k (X_0, Y_0) \Rightarrow \tau \leq \tau_0 \quad (3.7)$$

Assim justificando o nome da ordem \prec_k como *ordem estocástica de Kendall*.

No contexto do QQplot as v.a. BIPIT's H e H_0 são vistas como quantis de suas respectivas distribuições K e K_0 , ou seja, os n pontos (H_i, H_{0i}) do QQplot são dados por $H_i = K^{-1}(p_i)$ e $H_{0i} = K_0^{-1}(p_i)$, $\forall 0 = p_1 < \dots < p_n = 1$. Como K e K_0 são f.d.a., logo a implicação (3.6) é equivalente a

$$H_i \leq H_{0i}, \forall 1 \leq i \leq n \Rightarrow \tau \leq \tau_0 \quad (3.8)$$

Exemplo 3.4. A recíproca da equação (3.8) nem sempre é válida. Seja a cópula C dada por $C(u, v) = \min(C_M(u, v), 1/4 + C_W(u, v))$ com $K_C(t) = \max(t, (3/4)[t + 3/4])$ (Nelsen (2003) [32]), temos $\Pi(u, v) \leq C(u, v)$, $\forall (u, v) \in \mathbf{I}^2$, deste modo pela propriedade 6. da Definição 3.7, temos $\tau_\Pi \leq \tau_C$, porém $K_C(1/e) = 3/4 > 2/e = K_\Pi(1/e)$, ou seja, não é verdade que $\Pi \prec_k C$.

Exemplo 3.5. Conforme Exemplo 2.6 a família de Clayton segue a ordem estocástica de Kendall, ou seja, $\theta \leq \theta^* \Rightarrow C_\theta \prec_k C_{\theta^*}$, então pela implicação (3.7) temos $\tau_\theta \leq \tau_{\theta^*}$. A Figura 3.1 mostra o QQplot de (H_0, H_1) e (H_0, H_2) com H_0 v.a. BIPIT associada a

cópula Π , H_1 e H_2 as v.a. BIPIT associadas a cópula de Clayton com $\theta_1 = -0.5$ e $\theta_2 = 2$ respectivamente. Conforme ordenação estocástica de Kendall para a família de Clayton, observa-se que os gráficos referentes a $\theta_1 = -0.5$ e $\theta_2 = 2$ encontram-se respectivamente abaixo e acima da diagonal principal, ou seja, $H_{(1i)} \leq H_{0i} \leq H_{(2i)}$, caracterizando $\tau_{\theta_1} < \tau_0 < \tau_{\theta_2}$. De maneira geral, sendo H v.a. BIPIT de uma determinada cópula C , os QQplot's (H_0, H) situados acima da diagonal principal revelam uma estrutura de dependência positiva para H e situados abaixo revelam uma estrutura de dependência negativa, já que $\tau_0 = 0$ pois é associado a cópula Π (Definição 3.7 item 4.).

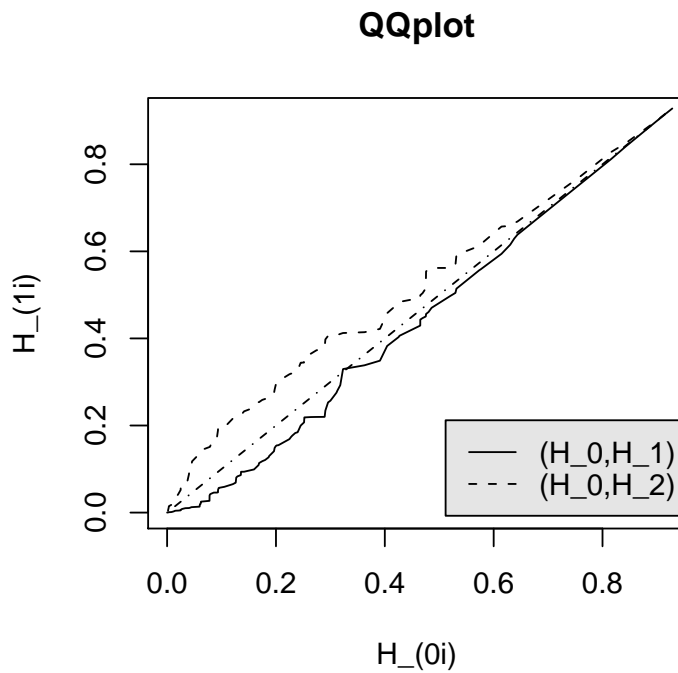


Figura 3.1: QQplots referentes Exemplo 3.5: $H_{0i} = \Pi(u_i, v_i)$, $H_{1i} = C_{-0.5}(u_i, v_i)$, $H_{2i} = C_2(u_i, v_i)$ com C_θ cópula de Clayton com parâmetro θ (A linha com ponto e tracejado é a diagonal principal do gráfico).

Capítulo 4

Kendall Plot

Em meio a um recente estudo de ferramentas para o ajuste de cópulas ainda há a necessidade de um mecanismo simples e eficiente para a modelagem de dependência. A motivação para a criação deste mecanismo partiu do mais simples modo para o ajuste de distribuições: o gráfico QQplot. Porém em cópulas trabalha-se num contexto multivariado diferentemente do QQplot. Este problema é resolvido usando a v.a. BIPIT que traduz um problema multivariado ao contexto univariado preservando as características da estrutura de dependência dos dados. Neste contexto Genest & Frave (2007) utilizam o gráfico QQplot usando a v.a. BIPIT para a modelagem de cópulas e denominam esta ferramenta gráfica por Kendall Plot. Assim como o QQplot padrão compara os quantis amostrais aos quantis teóricos da normal padrão, o gráfico proposto compara os quantis amostrais aos quantis teóricos sob hipótese nula de uma específica função distribuição K de uma v.a. BIPIT H associada a uma cópula C . Quando os quantis teóricos sob hipótese nula referem-se a cópula da independência Π (Exemplo 1.6), o gráfico proposto mostra se há evidência de dependência entre as variáveis. Esta foi a primeira proposta do Kendall Plot sugerida por Genest e Boies(2003).

Este capítulo também apresenta a relação existente entre a medida de concordância tau de Kendall e o gráfico Kendall Plot, sendo por essa razão o gráfico denominado desta maneira.

4.1 Construção do Kendall Plot

O Kendall Plot é uma adaptação do gráfico de normalidade. Sua construção é similar ao QQplot, porém usando o conceito de BIPIT.

4.1.1 QQplot

Uma maneira gráfica de verificar se uma amostra aleatória univariada X_1, \dots, X_n é Gaussiana é comparar os quantis amostrais com os quantis teóricos de uma normal padrão. O QQplot é o gráfico dos pares $(Z_{[np_i]:n}, X_{([np_i])})$, onde $[np_i]$ denota o menor inteiro maior ou igual a np_i , a sequência $\{X_{([np_i])}\}$, $0 \leq p_1 < \dots < p_n \leq 1$ corresponde às estatísticas de ordem da amostra e $Z_{[np]:n} \equiv E(Z_{([np_i])})$ onde $\{Z_1, \dots, Z_n\}$ é uma amostra aleatória com distribuição normal padrão.

Sendo F_n a f.d.a. empírica da v.a. X dada por

$$F_n(X_i) = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i\} \quad (4.1)$$

observa-se que a estatística de ordem $X_{(i)} \equiv X_{([np_i])}$ é definida como o p_i -quantil amostral de uma amostra de tamanho n . Mais especificamente, de acordo com a equação (4.1), para todo $1 \leq i \leq n$ e denotando $i \equiv [np_i]$, temos

$$F_n(X_{(i)}) = \frac{i-1}{n-1} \equiv p_i \quad (4.2)$$

com $0 = p_1 < \dots < p_n = 1$. E de acordo com Sen & Singer (1993) [37], sendo F a f.d.a da v.a. X , temos

$$X_{([np])} \rightarrow F^{-1}(p), \quad \forall p \in [0, 1]$$

em probabilidade e quase certamente quando $n \rightarrow \infty$ (para demonstração veja apêndice 2). Ou seja, é garantida a convergência do quantil amostral ao quantil teórico da dis-

tribuição. Deste modo a convergência de $Z_{\lceil np \rceil:n} \equiv E(Z_{(\lceil np \rceil)})$ ao quantil teórico da normal padrão é garantida pelo Teorema da Convergência Dominada (James (2004) [27])

$$\begin{aligned} \lim_{n \rightarrow \infty} E(Z_{(\lceil np \rceil)}) &= E\left(\lim_{n \rightarrow \infty} Z_{(\lceil np \rceil)}\right) \\ &= E[\phi^{-1}(p)] = \phi^{-1}(p) \end{aligned}$$

pois $\forall n$, $|Z_{(\lceil np \rceil)}| \leq Y$, com v.a. Y seguindo distribuição $Exp(\lambda)$, $\lambda \cong 0$.

Logo, para n suficientemente grande se os pontos $(Z_{(\lceil np_i \rceil)}, X_{(\lceil np_i \rceil)})$, $0 = p_1 < \dots < p_n = 1$ concentram-se sob a diagonal principal, então pode-se concluir que a f.d.a. da v.a. X é a f.d.a. da normal padrão.

O QQplot é utilizado não somente para verificar se uma amostra aleatória univariada $\{X_1, \dots, X_n\}$ com distribuição desconhecida F segue distribuição normal, mas também para testar se esta amostra segue qualquer outra distribuição F_0 , isto é,

$$\mathbf{H}_0 : F(x) = F_0(x), \forall x \in Dom F$$

$$\mathbf{H}_1 : F(x) \neq F_0(x), \text{ para pelo menos um valor de } x \in Dom F$$

O procedimento para este teste também é baseado em uma comparação quantílica, assim como para o teste com a distribuição normal padrão sob hipótese nula. Considere p_i , $0 = p_1 < \dots < p_n = 1$. Para cada p_i há dois quantis a considerar: $E(Y_{(\lceil np_i \rceil)})$ sob hipótese de que Y segue distribuição F_0 , o p_i -quantil teórico sob hipótese nula, e conforme equação (4.2) $F_n^{-1}(p_i) = X_{(i)} \equiv X_{\lceil np_i \rceil}$, o p_i -quantil amostral. Logo, se para n suficientemente grande o gráfico dos pares $(E(Y_{(\lceil np_i \rceil)}), F_n^{-1}(p_i))$, concentra-se sob a diagonal principal, então a v.a. X segue distribuição F_0 .

4.1.2 Kendall Plot

Seja uma amostra aleatória bivariada $(X_1, Y_1), \dots, (X_n, Y_n)$. É necessário transformar esta amostra bivariada em uma amostra univariada que contenha as mesmas informações de dependência da amostra original para que se possa trabalhar com o conceito do QQplot

com a finalidade de ajustar estruturas de dependência, ou seja, cópulas. A melhor maneira de transformar dados bivariados em univariados preservando a estrutura de dependência é utilizar a BIPIT desses dados. Como desconhece-se a distribuição dos dados, se trabalha com a BIPIT empírica \widehat{H} definida pela distribuição empírica \mathbb{H}_n dada por

$$\widehat{H}_i \equiv \mathbb{H}_n(X_i, Y_i) = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\} \quad (4.3)$$

Desta maneira a hipótese a ser testada pelo Kendall Plot é dada por

$$\begin{aligned} \mathbf{H}'_0 & : K(H) = K_0(H), \quad \forall H \in \mathbf{I} \\ \mathbf{H}'_1 & : K(H) \neq K_0(H), \quad \text{para pelo menos um valor de } H \in \mathbf{I} \end{aligned} \quad (4.4)$$

onde K é a f.d.a. desconhecida da BIPIT H e K_0 a f.d.a. a ser testada.

A hipótese (4.4) funciona como hipótese auxiliar, pois o objetivo principal é testar a estrutura de dependência dos dados, logo

$$\mathbf{H}_0 : C = C_0 \quad \text{vs.} \quad \mathbf{H}_1 : C \neq C_0 \quad (4.5)$$

onde a BIPIT H está associada a cópula desconhecida C com distribuição desconhecida K e C_0 com distribuição K_0 é a cópula a ser testada. Sob a hipótese nula a BIPIT H está associada a cópula C_0 e segue distribuição K_0 , portanto a estrutura de dependência dos dados bivariados associados a H é modelada pela cópula C_0 .

Deve-se tomar cuidado ao usar o Kendall Plot, pois as hipóteses (4.4) e (4.5) não são equivalentes, a não rejeição de \mathbf{H}'_0 não implica na não rejeição de \mathbf{H}_0 . Temos $\mathbf{H}_0 \subset \mathbf{H}'_0$, pois a implicação $C \neq C^* \Rightarrow K \neq K^*$ é falsa e para sua verificação veja exemplo a seguir.

Exemplo 4.1. Considere uma distribuição pertencente a classe de distribuições do valor-extremo, cujas cópulas (veja Capérea, Fougères e Genest (1997,2000) [3] e [4]) são da forma

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

para alguma função convexa $A : [0, 1] \rightarrow [1/2, 1]$, tal que $A(0) = A(1) = 0$ e $A(w) \geq \max(w, 1 - w)$, $\forall 0 \leq w \leq 1$ e sendo a função A definida como *gerador* da cópula do valor-extremo C . De acordo com Ghoudi, Khoudraji e Rivest (1998) [24] a v.a. BIPIT $H = C_A(U, V)$ é distribuída por

$$K_A(w) = w - (1 - \tau_A)w \log(w), \quad \forall 0 \leq w \leq 1$$

onde

$$\tau_A = \int_0^1 \frac{w(1-w)}{A(w)} dA'(w)$$

é o valor populacional de tau de Kendall, desde que A é convexa. Então, se duas distribuições do valor-extremo com geradores $A \neq A^*$ verificam $\tau_A = \tau_{A^*}$, logo $K_A = K_{A^*}$, ou seja, a cópula do valor-extremo C_A não é unicamente determinada por sua f.da. K_A . Mas esta questão não será preocupante nesta dissertação, já que trabalhamos com a classe de cópulas Arquimedianas bivariadas para a qual as hipóteses (4.4) e (4.5) são equivalentes, já que para esta família a cópula C é unicamente determinada pela função K . Este fato se deve a função distribuição K de uma cópula Arquimediana C ser definida em função do gerador ϕ desta cópula (Corolário 2.4). E pela construção de uma cópula Arquimediana (equação 1.22) observa-se que o gerador ϕ a determina unicamente.

Assim como o QQplot, o Kendall Plot realiza uma comparação quantílica. Compara-se os quantis amostrais de K aos quantis teóricos de K_0 sob hipótese nula. Os p_i -quantis amostrais são representados por $\widehat{H}_{(\lceil np_i \rceil)} \equiv \widehat{H}_{(i)}$ e os p_i -quantis teóricos denotados por $W_{\lceil np_i \rceil : n, K_0} = E(H_{(\lceil np_i \rceil)}^0) \equiv W_{i:n, K_0}$ sendo p_i definido como na equação (4.2). Os resultados da próxima seção permitem verificar que $\widehat{H}_{(\lceil np_i \rceil)}$ e $E(H_{(\lceil np_i \rceil)}^0)$ realmente representam o p_i -quantil amostral e teórico da distribuição sob hipótese nula K_0 .

Sejam $H_{(1)}^0 \leq \dots \leq H_{(n)}^0$ estatísticas de ordem de uma amostra aleatória H_1^0, \dots, H_n^0 da BIPIT H^0 com distribuição K_0 , pela definição de densidade de uma estatística de ordem em Casella & Berger (2002) [5] (para demonstração veja apêndice 3)

$$W_{i:n, K_0} \equiv E(H_{(i)}^0) = n \binom{n-1}{i-1} \int_0^1 w \{K_0(w)\}^{i-1} \{1 - K_0(w)\}^{n-i} dK_0(w) \quad (4.6)$$

A integral em questão não possui primitiva, portanto foi calculada por integração numérica através do comando *integrate* do software estatístico *R*. Este comando realiza integração numérica para variável unidimensional através de um método de quadratura adaptativa baseado nas rotinas *dqags* e *dqagi* em Piessens & Doncker-Kapenga (1983) [33]. Este algoritmo não funciona bem para integrandos que assumem valores constantes (em particular, próximos de zero) em todo seu domínio. O integrando em questão é composto por um produto de potências (que dependem de n) de valores entre 0 e 1, apresentando então valores muito próximos de zero em todo seu domínio, e sendo decrescente em n . Portanto para grandes valores de n , por exemplo $n > 200$, para algumas funções K_0 , por exemplo K_π , a integração usando *integrate* não corresponde ao valor esperado. Como o Kendall Plot, assim como o QQplot, é uma ferramenta assintótica foi necessário recorrer a outro método de integração adaptativa para o cálculo de $W_{i:n,K_0}$ para um grande tamanho amostral. Veja apêndice 4 para maiores detalhes do método de integração numérica.

Os passos para construção do Kendall Plot, listados a seguir, são semelhantes ao do QQplot.

1. Para cada $1 \leq i \leq n$, calcula-se a BIPIT empírica \hat{H}_i como na equação (4.3);
2. Ordena-se \hat{H}_i : $\hat{H}_{(1)} \leq \dots \leq \hat{H}_{(n)}$, igualdades são possíveis;
3. Para cada $1 \leq i \leq n$, calcula-se $W_{i:n,K_0}$ pela equação (4.6);
4. Plotar os pares $(W_{i:n,K_0}, \hat{H}_{(i)})$, $1 \leq i \leq n$.

4.2 Resultados e Fundamentos

Esta seção apresenta os resultados teóricos necessários para a validade do Kendall Plot como método assintótico para comparação de quantis, bem como exemplos decorrentes da teoria para auxiliar na interpretação do gráfico Kendall Plot. Um dos resultados evidencia a relação existente entre Kendall Plot e τ de Kendall.

Definição 4.1. *Sejam $\hat{H}_1, \dots, \hat{H}_n$ calculados pela equação (4.3) realizações da BIPIT $H = \mathbb{H}(X, Y)$ com função distribuição de Kendall K . A função distribuição empírica K_n dos H_i 's é dada por*

$$K_n(v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{\hat{H}_i \leq v\} \quad (4.7)$$

Teorema 4.2. *A função distribuição empírica de Kendall K_n converge em probabilidade à função distribuição de Kendall K sendo a convergência válida também para as inversas, ou seja,*

$$K_n(v) \xrightarrow{P} K(v) \quad e \quad K_n^{-1}(p) \xrightarrow{P} K^{-1}(p), \quad \forall v, p \in \mathbf{I}$$

quando $n \rightarrow \infty$.

Demonstração. Seja $(X_1, Y_1), \dots, (X_n, Y_n)$ uma amostra aleatória com f.d.a. cópula C e $\{\widehat{H}_1, \dots, \widehat{H}_n\}$, calculados pela equação (4.3), realizações da BIPIT $H = C(X, Y)$. Provenmos que a distribuição de \widehat{H}_i , $P(\widehat{H}_i \leq v)$, converge a $K(v) = P\{C(X, Y) \leq v\}$ e que a função distribuição empírica K_n dos \widehat{H}_i 's, dada pela equação (4.7), é um estimador \sqrt{n} -consistente de $K(v)$, ou seja, no limite a esperança do estimador converge a esperança de $K(v)$ e no limite a variância do estimador é nula. Consequentemente seguem as convergências em probabilidade.

$$E\{K_n(v)\} = E[\mathbb{I}\{\widehat{H}_1 \leq v\}] = P(\widehat{H}_1 \leq v) \quad (4.8)$$

Dados X_1 e Y_1 com distribuição $U(0, 1)$ sendo (X_1, Y_1) com f.d.a. cópula C . Seja (x_1, y_1) uma realização de (X_1, Y_1) , pela equação (4.3) temos que a quantidade $(n-1)\widehat{H}_1$ é distribuída de acordo com uma binomial com parâmetros $(n-1)$ e $C(x_1, y_1)$. Desta maneira a função geradora de momentos condicional a (x_1, y_1) de \widehat{H}_1 é dada por $\{1 - C(x_1, y_1) + C(x_1, y_1)e^{t/(n-1)}\}^{(n-1)}$. Como $\widehat{H}_1 = C_n(X_1, Y_1) \rightarrow C(X_1, Y_1)$ quando n tende ao infinito, onde C_n denota a cópula empírica, segue do teorema da convergência dominada de Lebesgue que $E(e^{t\widehat{H}_1})$ converge a $E\{e^{tC(X_1, Y_1)}\}$ quando n tende ao infinito. Portanto, assintoticamente, \widehat{H}_1 tem a mesma distribuição que $C(X_1, Y_1)$, ou seja,

$$\lim_{n \rightarrow \infty} P(\widehat{H}_1 \leq v) = K(v) \quad (4.9)$$

Logo, pela equação (4.8) temos

$$E[K_n(v)] \rightarrow K(v)$$

quando $n \rightarrow \infty$.

$$\begin{aligned} Var[K_n(v)] &= Var \left[\frac{1}{n} \sum_{i=1}^n \mathbb{I}\{\widehat{H}_i \leq v\} \right] \\ &= Var[\mathbb{I}\{\widehat{H}_1 \leq v\}]/n + (n-1)Cov(\mathbb{I}\{\widehat{H}_1 \leq v\}, \mathbb{I}\{\widehat{H}_2 \leq v\})/n \end{aligned}$$

sendo $\mathbb{I}\{\widehat{H}_1 \leq v\}$ e $\mathbb{I}\{\widehat{H}_2 \leq v\}$ v.a. de Bernoulli identicamente distribuídas com parâmetro $p(v) = P(\widehat{H}_1 \leq v)$, logo a variância de $K_n(v)$ pode ser escrita como

$$p(v)\{1 - p(v)\}/n + (n-1)[P(\widehat{H}_1 \leq v, \widehat{H}_2 \leq v) - p(v)^2]/n \quad (4.10)$$

Usando o termo de ordem $1/n$ da transformada de Laplace-Stieljes bivariada (Abramowitz & Stegun (1972) [1]) de

$$P(\widehat{H}_1 \leq v_1, \widehat{H}_2 \leq v_2) - P(\widehat{H}_1 \leq v_1)P(\widehat{H}_2 \leq v_2) \quad (4.11)$$

mostra-se que

$$P(\widehat{H}_1 \leq v, \widehat{H}_2 \leq v) - P(\widehat{H}_1 \leq v)^2 = k(v)\{k(v)R(v) - 2v\overline{K}(v)\}/n + o(1/n) \quad (4.12)$$

onde $\overline{K}(v) = 1 - K(v)$, $k(v) = K'(v)$ é a densidade de $H = C(X, Y)$ e

$$R(v) = E[C\{\min(X_1, X_2), \min(Y_1, Y_2)\} - v^2 | C(X_1, Y_1) = C(X_2, Y_2) = v]$$

Para detalhes veja Genest & Rivest (1993) [17].

Como a distribuição de \widehat{H}_1 no limite é dada por $K(v)$ (equação 4.9) e substituindo a equação (4.12) na equação (4.10), então uma aproximação de ordem $o(1/n)$ para a variância de $K_n(v)$ é dada por

$$[K(v)\overline{K}(v) + k(v)\{k(v)R(v) - 2v\overline{K}(v)\}]/n$$

E deste modo

$$\lim_{n \rightarrow \infty} \text{Var}[K_n(v)] = 0$$

e completa-se a prova. \square

O corolário a seguir é de extrema importância para a característica assintótica do Kendall Plot mostrando que $\widehat{H}_{(\lceil np \rceil)}$ e $W_{\lceil np \rceil:n,K}$ convergem ao p -quantil teórico da verdadeira distribuição K e ao p -quantil teórico sob hipótese nula de distribuição K_0 respectivamente.

Corolário 4.3. *Sejam $\{\widehat{H}_1, \dots, \widehat{H}_n\}$ realizações da BIPIT H , $\{H_1, \dots, H_n\}$ amostra aleatória de H com f.d.a. K , K_n como na Definição 4.1 e $0 \leq p \leq 1$, então*

$$\widehat{H}_{(\lceil np \rceil)} = K_n^{-1}(p) \xrightarrow{\mathbf{P}} K^{-1}(p) \quad (4.13)$$

e

$$W_{\lceil np \rceil:n,K} \equiv E(H_{(\lceil np \rceil)}) \rightarrow K^{-1}(p) \quad (4.14)$$

quando $n \rightarrow \infty$.

Demonstração. Demonstração da equação (4.13):

De acordo com a Definição 4.1 pode-se considerar $\widehat{H}_{(\lceil np \rceil)} = K_n^{-1}(p)$, logo a convergência segue do Teorema 4.2.

Demonstração da equação (4.14):

$$\begin{aligned} \lim_{n \rightarrow \infty} W_{\lceil np \rceil:n,K} &\equiv \lim_{n \rightarrow \infty} E(H_{(\lceil np \rceil)}) = E\left(\lim_{n \rightarrow \infty} H_{(\lceil np \rceil)}\right) \\ &= E[K^{-1}(p)] = K^{-1}(p) \end{aligned}$$

Por $|H_{(\lceil np \rceil)}| \leq 1$, $\forall n$ e a convergência quase certa,

$$H_{(\lceil np \rceil)} \xrightarrow{n \rightarrow \infty} K^{-1}(p)$$

demonstrada por Sen & Singer (1993) [37] (veja Apêndice .2), a igualdade entre limite da esperança e esperança do limite é garantida pelo Teorema da Convergência Dominada (James (2004) [27]). \square

Exemplo 4.2. A Figura 4.1 ilustra a convergência da BIPIT empírica $\widehat{H}_{(\lceil np \rceil)} \equiv K_n^{-1}(p)$ e $W_{\lceil np \rceil:n,K} \equiv E(H_{(\lceil np \rceil)})$ a $K^{-1}(p)$, inversa da distribuição acumulada da BIPIT H . Usa-se no exemplo a BIPIT H associada a cópula 4.2.12 (Tabela 1.1) que possibilita o cálculo direto da inversa $K^{-1}(p)$. Esta cópula Arquimediana torna-se uma exceção frente que a distribuição das demais cópulas Arquimedianas não apresentam uma forma fechada para o cálculo da inversa da distribuição.

O Corolário 4.3 confirma o mencionado na seção 4.1.2, ou seja, que os elementos do Kendall Plot, $\widehat{H}_{(i)}$ e $W_{i:n,K_0} \equiv E(H_{(i)}^0)$, representam respectivamente o p_i -quantil da distribuição K (desconhecida) e o p_i -quantil da distribuição sob a hipótese nula K_0 para n suficientemente grande. Logo, se os pontos $(W_{i:n,K_0}, \widehat{H}_{(i)})$ concentram-se sob a diagonal principal evidencia-se que a BIPIT H segue distribuição K_0 , ou seja, a não rejeição da hipótese nula do Kendall Plot. Este fato é verificado formalmente através do corolário abaixo.

Corolário 4.4. *Os pares $(W_{i:n,K_0}, \widehat{H}_{(i)})$ do gráfico Kendall Plot concentram-se ao longo da curva*

$$p \mapsto K^{-1}\{K_0(p)\}, \quad \forall p \in \mathbf{I}. \quad (4.15)$$

Demonstração. Reparametrizando $p = K_0(w)$ temos $w = K_0^{-1}(p)$, $w \in \mathbf{I}$, então

$$(K_0^{-1}(p), K^{-1}(p)) = (w, K^{-1}\{K_0(w)\})$$

logo os pontos do Kendall Plot se comportam como os pontos do gráfico $w \mapsto K^{-1}\{K_0(w)\}$. \square

Exemplo 1.1 (continuação)

O gráfico dos pontos $(W_{i:n,K_0}, \widehat{H}_{(i)})$, $1 \leq i \leq n$, de acordo com a equação (4.15) concentra-se ao longo da diagonal principal sob a hipótese nula, ou seja, quando $K = K_0$. A Figura 4.2 mostra o Kendall Plot dos dados simulados no Exemplo 1.1 sob a hipótese nula de independência, ou seja, $K_0(v) = K_{\Pi}(v)$ (veja Exemplo 2.3), evidenciando independência nos dados.

Exemplo 4.3. O gráfico dos pontos $(W_{i:n,K_0}, \widehat{H}_{(i)})$, $1 \leq i \leq n$, de acordo com a equação (4.15) concentra-se sob o eixo horizontal - $\widehat{H}_{(i)} = 0$, $1 \leq i \leq n$ - quando o vetor aleatório

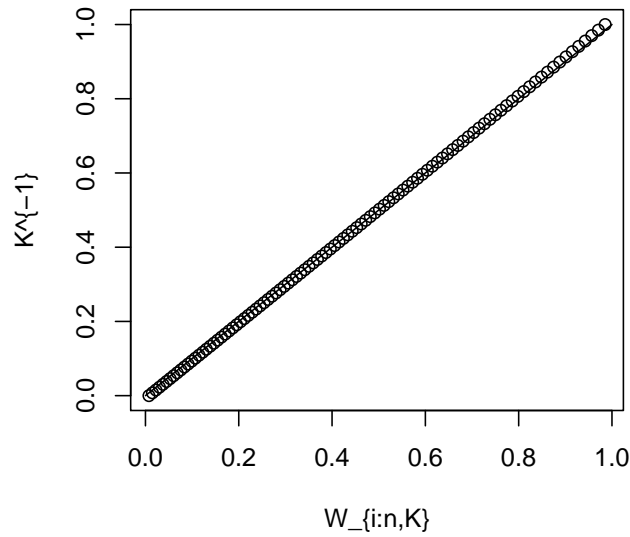
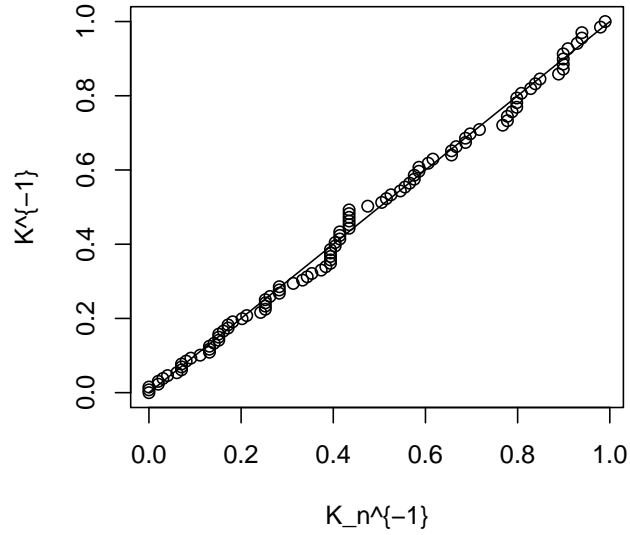


Figura 4.1: Gráficos da inversa da função distribuição $K^{-1}(p)$ vs. $\hat{H}_{([np])} \equiv K_n^{-1}(p)$ proveniente de $n = 100$ pseudo-observações da cópula 4.2.12 (primeira figura); e $W_{[np]:n,K} = E(H_{([np])})$, $n = 100$, sob hipótese da mesma cópula (segunda figura).

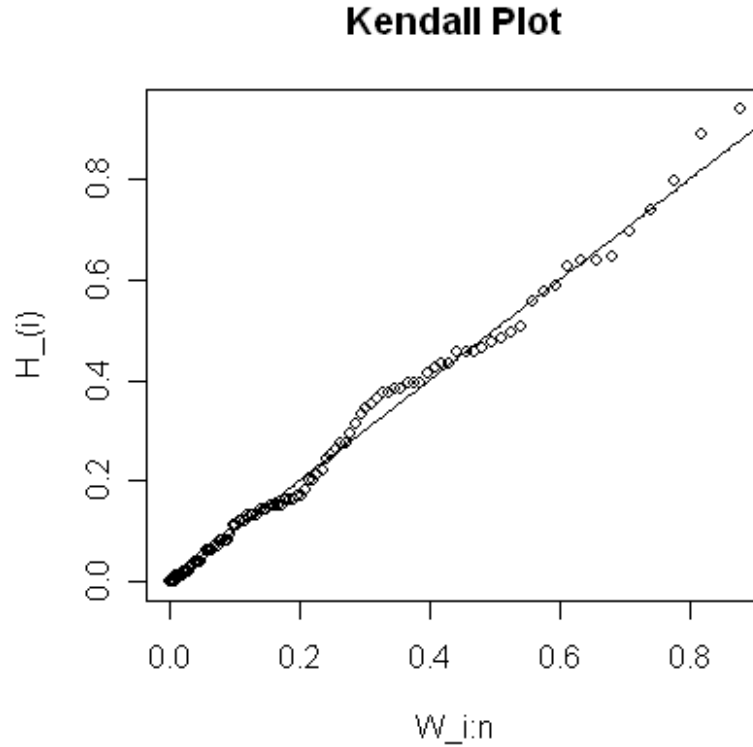


Figura 4.2: Kendall Plot sob hipótese nula de independência para a amostra $\{(x_1, y_1), \dots, (x_{100}, y_{100})\}$ sendo x_i com distribuição $exp(2)$ e y_i com distribuição $exp(10)$ com $\{x_i\}$ gerados de forma independente de $\{y_i\}$.

(X, Y) testado está associado a cópula minimal W (veja Exemplo 1.5). Como $K_W(v) = 1, \forall v \in \mathbf{I}$ (Exemplo 2.5), de acordo com a definição de quasi-inversa (Definição 1.11) temos $K_W^{-1}(p) = 0, \forall p \in \mathbf{I}$. A Figura 4.3 apresenta o Kendall Plot para as observações do vetor aleatório (X, Y) sendo $Y = 1 - X$ que de acordo com demonstração do Teorema 3.6, $C_{X,Y} = W$.

Exemplo 4.4. O gráfico dos pontos $(W_{i:n, K_0}, \hat{H}_{(i)})$, $1 \leq i \leq n$, de acordo com a equação (4.15) concentra-se sob a curva $K_0(p)$ quando o vetor aleatório (X, Y) sendo testado está associado a cópula maximal M (veja Exemplo 1.5), pois $K_M^{-1}(p) = p, \forall p \in \mathbf{I}$ (Exemplo 2.4). A Figura 4.4 apresenta o Kendall Plot para as observações do vetor aleatório (X, Y) sendo $Y = X$ que de acordo com demonstração do Teorema 3.6, $C_{X,Y} = M$.

Teorema 4.5. *Seja a BIPIT H associada a cópula C desconhecida. Se o gráfico do*

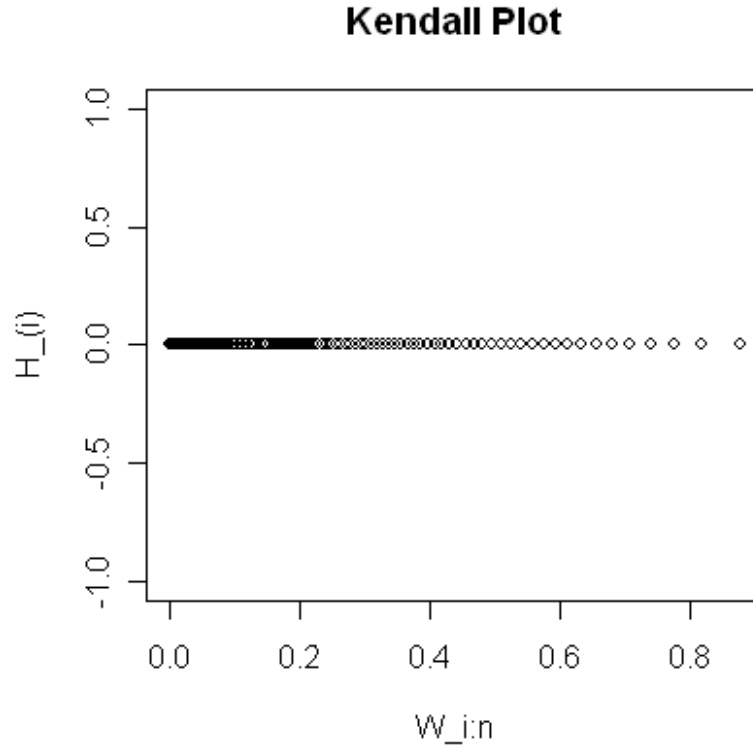


Figura 4.3: Kendall Plot sob hipótese nula de independência de $n = 100$ pares de observações do vetor aleatório (X, Y) com $Y = 1 - X$.

Kendall Plot, sob hipótese nula $C = C_0$, apresenta $\hat{H}_{(i)} \leq W_{i:n, K_0}$, $\forall 1 \leq i \leq n$ para n suficientemente grande, então $\tau_C \leq \tau_{C_0}$.

Demonstração. De acordo com as convergências de $\hat{H}_{(i)}$ e $W_{i:n, K}$ evidenciadas pelo Corolário 4.3 e pela relação entre v.a. BIPIT e tau de Kendall estabelecida na equação (3.8), para n suficientemente grande temos

$$\hat{H}_{(i)} \leq W_{i:n, K_0}, \quad \forall 1 \leq i \leq n \Rightarrow \tau_C \leq \tau_{C_0} \quad (4.16)$$

□

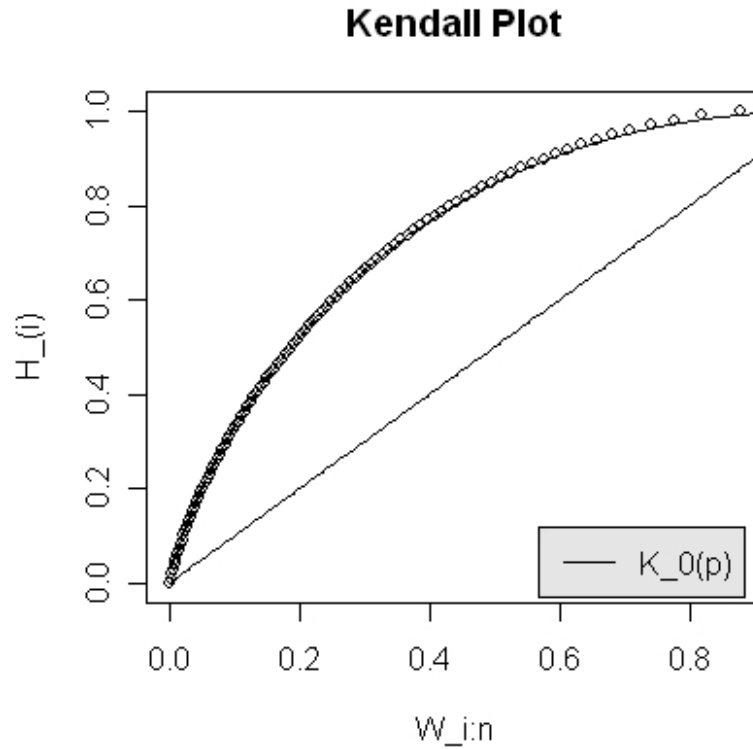


Figura 4.4: Kendall Plot sob hipótese nula de independência, de $n = 100$ pares de observações do vetor aleatório (X, Y) com $Y = X$.

A recíproca da equação (4.16) nem sempre é válida, veja Exemplo 3.4.

Exemplo 4.5. Simulamos três amostras aleatórias de tamanho $n = 150$ da cópula de Clayton com os respectivos parâmetros $\theta = \{0.5, 2, 5\}$ e construímos os Kendall Plot's sob hipótese nula de Clayton $\theta = 2$ (Figura 4.5). Conforme esperado o plot dos dados com $\theta = 2$ encontra-se sob a diagonal principal, e observa-se que os dados com $\theta = 0.5$ e $\theta = 5$ encontram-se respectivamente abaixo e acima do plot para $\theta = 2$, conforme esperado de acordo com a ordenação estocástica de Kendall para a família de Clayton (veja Exemplo 3.5).

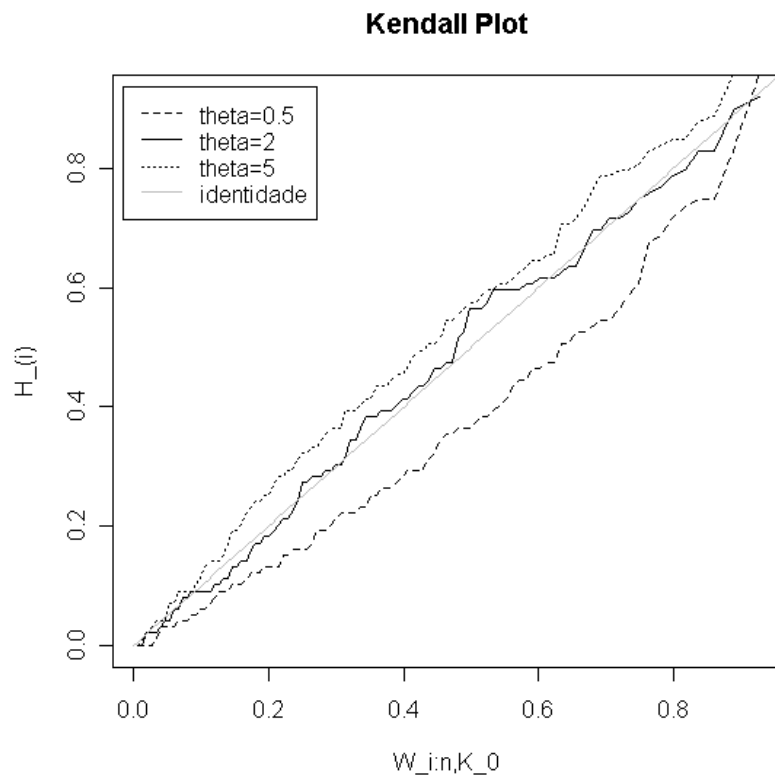


Figura 4.5: Kendall Plot sob hipótese nula de cópula de Clayton $\theta = 2$ ($W_{i:n,K_0}$ referente a Clayton com $\theta = 2$) de amostras aleatórias de tamanho $n = 100$ associadas a cópula de Clayton com os respectivos parâmetros $\theta = \{0.5, 2, 5\}$.

Capítulo 5

Testes de Aderência para Cópulas

Enquanto testes para distribuições univariadas estão bem documentados, o estudo de metodologias formais para o ajuste de cópulas surgiu apenas recentemente como um problema inferencial desafiador.

As diferentes características de cada família de cópulas \mathbb{C} possibilitam que se modele diferentes contextos em cada família. Por exemplo, as cópulas Arquimedianas podem assumir dependência caudal assimétrica, sendo uma propriedade a favor de sua aplicação à modelagem de dados com esta estrutura de dependência. Recentes estudos empíricos mostram que períodos de turbulência e calma em finanças são caracterizados por diferentes níveis de dependência caudal, sendo a dependência mais forte sobre na cauda inferior do que na cauda superior. Devido a importância de se detectar em qual família os dados se adequam, ultimamente diversas contribuições tem sido feitas para se testar as hipóteses

$$\mathbf{H}_0 : C \in \mathbb{C} = \{C_\theta | \theta \in \Theta\} \quad vs. \quad \mathbf{H}_1 : C \notin \mathbb{C} = \{C_\theta | \theta \in \Theta\},$$

onde Θ é o espaço paramétrico da família de cópulas \mathbb{C} . Autores como Genest, Quessy & Rémillard (2006) [20], Genest & Rémillard (2008) [22], Dobrić & Schmid (2007) [7] e Genest, Rémillard & Beaudoin (2008) [23] apresentam estatísticas utilizadas em um procedimento bootstrap paramétrico para testar estas hipóteses. No apêndice 5 são apresentadas duas destas estatísticas e o procedimento bootstrap para a realização do teste.

Com a construção de uma banda de confiança, permitida pela estrutura da v.a. BIPIT,

o Kendall Plot deixa de ser apenas um meio auxiliar gráfico para o ajuste de cópulas e passa a ser um teste gráfico completo. A proposta deste capítulo é postular este teste gráfico para se testar a hipótese

$$\mathbf{H}_0 : C = C_{\theta_0} \quad vs. \quad \mathbf{H}_1 : C \neq C_{\theta_0}$$

sendo $\theta_0 \in \Theta$ fixo e, em particular, a família paramétrica sob \mathbf{H}_0 uma família Arquimediana. Há dois elementos sendo testados pelo Kendall Plot: a família paramétrica e o respectivo parâmetro θ . Em uma família Arquimediana uniparamétrica o parâmetro θ é determinado completamente pela medida de associação τ de Kendall, sendo assim por meio desta medida calcula-se um valor θ_0 (usando equação (3.4) conforme Exemplo 3.1), referente ao nível de associação τ dos dados, para compor a hipótese nula. O Kendall Plot então compara a estrutura global de dependência dos dados (com nível de associação τ) a estrutura de dependência da família Arquimediana sob hipótese nula (com o mesmo nível de associação τ expresso por θ_0).

Embora o objetivo principal deste capítulo seja o estudo do Kendall Plot munido de bandas de confiança para testar estruturas de dependência por meio de cópulas, o Kendall Plot também pode ser utilizado para se testar apenas o parâmetro θ quando conhecida a família Arquimediana adequada aos dados. Como em uma família Arquimediana o parâmetro θ é determinado por τ de Kendall é razoável pensar que este teste pode se comportar como um teste para a medida de associação τ de Kendall. Para informações sobre o comportamento do poder deste teste veja o apêndice 7.

A próxima seção apresenta o resultado de Walker(1968) [39] que permite a construção das bandas de confiança para o Kendall Plot. Desta maneira o teste gráfico Kendall Plot torna-se uma ferramenta formal, ao contrário do QQplot que funciona como uma ferramenta gráfica auxiliar em um teste de ajuste de distribuição.

5.1 Teste Gráfico - Intervalos de Confiança para Quantis

Walker (1968) [39], enuncia que para fazer inferência sobre um ou mais quantis de função distribuição F com densidade f , sendo o p -quantil teórico ξ_p definido unicamente por $F(\xi_p) = p$ para $0 < p < 1$, é necessário considerar apenas o conjunto das estatísticas de

ordem $X_{(1)} \leq \dots \leq X_{(n)}$, obtido a partir de uma amostra aleatória $\{X_1, \dots, X_n\}$, onde o p -quantil amostral deve ser definido por $Y_{n,p} = X_{[np]}$, com $[np]$ denotando a parte inteira de np . Deste modo, sob certas condições, o p -quantil amostral $Y_{n,p}$ é aproximadamente normal com média ξ_p e variância $p(1-p)/\{nf^2(\xi_p)\}$ quando n é suficientemente grande.

Segue o resultado formal de Walker (1968) [39] para a distribuição assintótica de quantis amostrais.

Teorema 5.1. *Sejam $X_{(1)}, \dots, X_{(n)}$ estatísticas de ordem de uma amostra aleatória de f.d.a. F absolutamente contínua e estritamente crescente para $0 < F(x) < 1$. Seja $\{r(n)\}$ uma sequência de inteiros tais que $n^{-\frac{1}{2}}(r(n) - np) \rightarrow 0$ quando $n \rightarrow \infty$, onde $0 < p < 1$ e seja $\xi_p = F^{-1}(p)$. Então se $F'(\xi_p) = f(\xi_p)$ existe e é positivo, $n^{\frac{1}{2}}(X_{(r(n))} - \xi_p)$ converge para*

$$N\left(0, \frac{p(1-p)}{f^2(\xi_p)}\right) \quad (5.1)$$

quando $n \rightarrow \infty$.

Demonstração. Seja $Z = \sqrt{nf^2(\xi_p)}(X_{(r(n))} - \xi_p)$, logo

$$P(Z \leq z) = P\left(X_{(r(n))} \leq \xi_p + z/(n^{\frac{1}{2}}f(\xi_p))\right)$$

Devido a F ser estritamente crescente, consideremos $X_{(r)} = F^{-1}(U_{(r)})$, onde $U_{(r)}$ é a r -ésima estatística de ordem de uma amostra aleatória de $U(0, 1)$. Então se $U_{(r(n))} = F(X_{(r(n))})$

$$P(Z \leq z) = P\left(U_{(r(n))} \leq F\left[\xi_p + z/(n^{\frac{1}{2}}f(\xi_p))\right]\right) \quad (5.2)$$

que desde a definição de derivada

$$\lim_{n \rightarrow \infty} \frac{F(\xi_p + z/(n^{\frac{1}{2}}f(\xi_p))) - p}{z/(n^{\frac{1}{2}}f(\xi_p))} = f(\xi_p)$$

temos

$$F\left(\xi_p + \frac{z}{n^{\frac{1}{2}}f(\xi_p)}\right) \cong (f(\xi_p) + \epsilon_n)\frac{z}{n^{\frac{1}{2}}f(\xi_p)} + p \quad (5.3)$$

e definindo

$$\varsigma_n \equiv \frac{z\epsilon_n}{f(\xi_p)}$$

onde

$$\lim_{n \rightarrow \infty} \varsigma_n = 0$$

temos a equação (5.2) equivalente a

$$P\left(U_{(r(n))} \leq p + n^{-\frac{1}{2}}(z + \varsigma_n)\right) \quad (5.4)$$

Segue da equação (5.4) que se N_n denota o número de observações que não excedem $p + n^{-\frac{1}{2}}(z + \varsigma_n)$ em uma amostra aleatória $U(0,1)$ de tamanho n , então

$$P(Z \leq z) = P(N_n \geq r(n)) \quad (5.5)$$

Desde que N_n segue distribuição $Bin(n, p_n)$ com $p_n = n^{-\frac{1}{2}}(z + \varsigma_n) + p$.

Seja $W_n = n^{-\frac{1}{2}}(N_n - np_n)$, a função característica de W_n é dada por

$$\begin{aligned}
\varphi_{W_n}(t) &= E(e^{itW_n}) = E(e^{it[n^{-\frac{1}{2}}(N_n - np_n)]) \\
&= E(e^{itn^{-\frac{1}{2}}N_n} e^{-itn^{\frac{1}{2}}p_n}) \\
&= e^{-itn^{\frac{1}{2}}p_n} \varphi_{N_n}(n^{-\frac{1}{2}}t) \\
&= e^{-n^{\frac{1}{2}}p_n it} [p_n(e^{itn^{-\frac{1}{2}}} - 1) + 1]^n
\end{aligned}$$

E o log da função característica de W_n é dado por

$$\begin{aligned}
\psi_{W_n}(t) &= -n^{\frac{1}{2}}p_n it + n \log[1 + p_n(\exp(n^{-\frac{1}{2}}it) - 1)] \\
&= np_n[\exp(-n^{-\frac{1}{2}}it) - 1 - n^{-\frac{1}{2}}it] - \frac{1}{2}np_n^2[\exp(-n^{-\frac{1}{2}}it) - 1]^2 + nO(n^{-\frac{3}{2}}|t|^3) \\
&= -\frac{1}{2}p_n(1 - p_n)t^2 + O(n^{-\frac{1}{2}}|t|^3)
\end{aligned}$$

Logo

$$\lim_{n \rightarrow \infty} \psi_{W_n}(t) = -\frac{1}{2}p(1 - p)t^2$$

Como o log da função característica de W_n converge para o log da função característica da distribuição $N(0, p(1 - p))$, segue que W_n converge para uma variável $N(0, p(1 - p))$.

Pela equação (5.5) e substituindo p_n segue

$$\begin{aligned}
P(Z \leq z) &= P[W_n \geq n^{-\frac{1}{2}}(r(n) - np_n)] \\
&= P[W_n \geq n^{-\frac{1}{2}}(r(n) - np) - (z + \varsigma_n)] \tag{5.6}
\end{aligned}$$

E para n suficientemente grande $P(Z \leq z)$ na equação (5.6) está entre $P(W_n \geq -z + \varsigma_n)$ e $P(W_n \geq -z - \varsigma_n)$ sendo ς_n muito pequeno. Pela normalidade e simetria de W_n , temos

$$P(W_n \geq -z + \varsigma_n) = \Phi\left(\frac{z - \varsigma_n}{\sqrt{p(1 - p)}}\right) \quad e \quad P(W_n \geq -z - \varsigma_n) = \Phi\left(\frac{z + \varsigma_n}{\sqrt{p(1 - p)}}\right)$$

Logo,

$$\lim_{n \rightarrow \infty} P(Z \leq z) = \Phi \left(\frac{z}{\sqrt{p(1-p)}} \right)$$

Portanto a distribuição assintótica de Z é $N(0, p(1-p))$ e conseqüentemente $n^{\frac{1}{2}}(X_{(r(n))} - \xi_p)$ segue distribuição assintótica $N(0, p(1-p)/f^2(\xi_p))$. \square

5.1.1 Construção de Intervalos de Confiança para Quantis

Sejam $\{H_1, \dots, H_n\}$ uma amostra aleatória da BIPIT $H = C(U, V)$ com f.d.a. desconhecida K , $K'(x) = k(x)$, a seqüência $\{r(n)\}$ dada por $r(n) = n^{-1/2}(\lceil np \rceil - np) \rightarrow 0$ quando $n \rightarrow \infty$ onde $0 < p < 1$ e $\xi_p \equiv K^{-1}(p)$. Aplicando o Teorema 5.1 temos

$$\sqrt{n}(H_{(\lceil np \rceil)} - \xi_p) \xrightarrow{D} N(0, p(1-p)/k^2(\xi_p)), \quad \forall p \in (0, 1) \quad (5.7)$$

quando $n \rightarrow \infty$.

Sejam K_0 uma f.d.a. absolutamente contínua e estritamente crescente (a f.d.a. K de uma cópula Arquimediana satisfaz estas condições: um dos motivos de sua utilização nesta dissertação), $\xi_p^0 \equiv K_0^{-1}(p)$ e $s_{p, K_0}^2 \equiv p(1-p)/k_0^2(\xi_p^0)$ sendo k_0 a densidade de K_0 . Definamos a estatística

$$Q(K_0, K) \equiv \frac{\sqrt{n}(H_{(\lceil np \rceil)} - \xi_p)}{s_{p, K_0}} \quad (5.8)$$

De acordo com Teorema 5.1, se K_0 é a verdadeira distribuição de $\{H_1, \dots, H_n\}$, então

$$Q(K_0, K) \xrightarrow{D} N(0, 1), \quad \forall p \in (0, 1)$$

quando $n \rightarrow \infty$, ou seja, $Q(K_0, K)$ constitui uma quantidade pivotal para o parâmetro ξ_p . Deste modo, se $\alpha \in (0, 1)$ temos

$$\lim_{n \rightarrow \infty} P(-z_\alpha < Q(K_0, K) < z_\alpha) = 1 - \alpha$$

equivalente a

$$P\left(H_{(\lceil np \rceil)} - \frac{z_\alpha s_{p, K_0}}{\sqrt{n}} \leq \xi_p \leq H_{(\lceil np \rceil)} + \frac{z_\alpha s_{p, K_0}}{\sqrt{n}}\right) = 1 - \alpha$$

Ou seja, sob $\mathbf{H}_0 : K^{-1}(p) = K_0^{-1}(p)$ a estatística $Q(K_0, K)$ comporta-se de acordo com a distribuição $N(0, 1)$ e o intervalo

$$IC(\xi_p, \alpha) \equiv \left[H_{(\lceil np \rceil)} \pm \frac{z_\alpha s_{p, K_0}}{\sqrt{n}} \right] \quad (5.9)$$

é um intervalo assintótico para $\xi_p = K^{-1}(p)$ de confiança $1 - \alpha$. Desta maneira, rejeitamos $\mathbf{H}_0 : K^{-1}(p) = K_0^{-1}(p)$ ao nível de confiança α , se $\xi_p^0 \notin IC(\xi_p, \alpha)$.

5.1.2 Teste gráfico para o Kendall Plot: uma aplicação para intervalos de confiança de quantis

A hipótese nula do Kendall Plot $\mathbf{H}'_0 : K = K_0$ pode ser reescrita como uma interseção de hipóteses nulas. Formalmente

$$\mathbf{H}'_0 : \bigcap_{j \in \mathcal{I}} \{K^{-1}(p_j) = K_0^{-1}(p_j)\}$$

sendo $p_j \in (0, 1)$ e \mathcal{I} um conjunto de índices. Então, a região de rejeição para o teste é

$$\bigcup_{j \in \mathcal{I}} \{\xi_{p_j}^0 : \xi_{p_j}^0 \notin IC(\xi_{p_j}, \alpha^*)\}$$

Ou seja, se pelo menos uma das hipóteses nulas $\mathbf{H}_{0_{p_j}} : K^{-1}(p_j) = K_0^{-1}(p_j)$ for rejeitada, isto é, pelo menos um $\xi_{p_j}^0 \notin IC(\xi_{p_j}, \alpha^*)$, então \mathbf{H}_0 é rejeitado. Apenas, se cada hipótese nula $\mathbf{H}_{0_{p_j}} : K^{-1}(p_j) = K_0^{-1}(p_j), \forall p_j \in (0, 1)$, não for rejeitada, então não haverá evidências para se rejeitar \mathbf{H}_0 .

Ao se utilizar uma amostra de tamanho n o Kendall Plot permite realizar no máximo $n - 2$ testes conjuntos, $\mathbf{H}_{0_{p_j}} : K^{-1}(p_j) = K_0^{-1}(p_j)$ com $0 < p_2 < \dots < p_{n-1} < 1$ usando os intervalos de confiança $IC(\xi_{p_j}, \alpha^*)$. Não são realizados testes para $\mathbf{H}_{0_{p_1=0}}$ e $\mathbf{H}_{0_{p_n=1}}$, pois os intervalos de confiança $IC_p(\xi_p, \alpha^*)$ são válidos para $0 < p < 1$ de acordo com Teorema 5.1. Cada teste $\mathbf{H}_{0_{p_j}}$ é realizado com nível de significância $\alpha^* = \alpha/|\mathcal{I}|$, para que de acordo com a correção de Bonferroni (veja apêndice 6) o teste \mathbf{H}'_0 tenha nível de significância α .

A utilização do método de Bonferroni para se controlar o nível de confiança α do teste implica em quanto maior o número de testes individuais envolvidos mais conservador será o teste para se rejeitar a hipótese nula quando ela é verdadeira. Logo, sendo o Kendall Plot um teste assintótico, para grandes tamanhos amostrais o nível de significância do teste tende a zero ao utilizar o número máximo de testes individuais. Para contornar este problema, uma solução a ser considerada é estabelecer um número fixo de testes. Ou seja, estipular quais e quantos p_j -quantis são comparados. Neste capítulo os Kendall Plot's construídos como exemplo são resultantes da intersecção de $n - 2$ testes, ou seja, são comparados os p_j -quantis dados por $p_j = (j - 1)/(n - 1)$ (equação (4.2)), $j = 2, \dots, n - 1$. Na seção 5.2 analisa-se o poder do teste Kendall Plot utilizando todos os $n - 2$ p_j -quantis e utilizando no máximo 20 p_j -quantis igualmente espaçados.

De acordo com o Corolário 4.4, quando $K = K_0$ a curva $p_j \mapsto K^{-1}\{K_0(p_j)\}$, $p_j \in [0, 1]$ do Kendall Plot concentra-se sob a diagonal principal. Os intervalos $IC(\xi_{p_j}, \alpha^*)$ (equação (5.9)) permitem decidir quando os pontos do Kendall Plot $(\xi_{p_j}^0, H_{(\lceil np_j \rceil)})$ encontram-se próximos a diagonal principal. Se todos os pontos do Kendall Plot pertencerem a banda de confiança construída com os intervalos, ou seja, se

$$\xi_{p_j}^0 \in IC(\xi_{p_j}, \alpha^*), \forall p_j \in (0, 1) \text{ e } j \in \mathcal{I}$$

pode-se concluir a proximidade da curva a diagonal principal, e por consequência, não há evidências para a rejeição da hipótese nula do Kendall Plot $\mathbf{H}_0 : K = K_0$.

Abaixo segue o algoritmo para a construção do Kendall Plot com a banda de confiança construída a partir dos intervalos de confiança $IC_p(\xi_{p_j}, \alpha^*)$, $0 < p_j < 1$, constituindo então

um teste de nível α para o teste da hipótese nula $\mathbf{H}'_0 : K = K_0$.

Se n for suficientemente grande $\{\widehat{H}_i\}_{i=1}^n$ como definido na equação (4.3) permite a construção de uma amostra de H visto que $\mathbb{H}_n \rightarrow \mathbb{H}$ quase certamente quando $n \rightarrow \infty$. Desta maneira o Teorema de Walker 5.1 torna-se válido para as BIPIT's empíricas \widehat{H}_i e permite a construção de *IC* assintóticos utilizando BIPIT's empíricas.

1. Para cada $1 \leq i \leq n$, calcula-se a BIPIT empírica \widehat{H}_i pela equação (4.3);
2. Ordena-se \widehat{H}_i : $\widehat{H}_{(1)} \leq \dots \leq \widehat{H}_{(n)}$, igualdades são possíveis;
3. Para cada $j \in \mathcal{I}$, calcula-se $\xi_{p_j}^0 \equiv K_0^{-1}(p_j)$.
4. Para cada $j \in \mathcal{I}$ calcula-se $s_{p_j, K_0}^2 \equiv p_j(1 - p_j)/k_0^2(\xi_{p_j}^0)$, onde k_0 é a densidade de K_0 .
5. Calcula-se os extremos inferiores l_{p_j} e superiores u_{p_j} dos IC de confiança $IC_{p_j}(\xi_{p_j}, \alpha^*)$, $\alpha^* = \alpha/|\mathcal{I}|$ por

$$l_{p_j} = \widehat{H}_{(\lceil np_j \rceil)} - \frac{z_{\alpha^*} s_{p_j, K_0}}{\sqrt{n}}$$

$$u_{p_j} = \widehat{H}_{(\lceil np_j \rceil)} + \frac{z_{\alpha^*} s_{p_j, K_0}}{\sqrt{n}}$$

onde z_{α^*} é determinado de modo que

$$P(Z \leq z_{\alpha^*}) = 1 - \frac{\alpha^*}{2}$$

sendo Z v.a. com distribuição $N(0, 1)$.

6. Plota-se os pontos $(\xi_{p_j}^0, \widehat{H}_{(\lceil np_j \rceil)})$ que compõem a curva do Kendall Plot
7. Plota-se os pontos $(u_{p_j}, \widehat{H}_{(\lceil np_j \rceil)})$ que compõem o extremo superior e os pontos $(l_{p_j}, \widehat{H}_{(\lceil np_j \rceil)})$, $j \in |\mathcal{I}|$, que compõem o extremo inferior da banda de confiança.
8. Se pelo menos um ponto $(\xi_{p_j}^0, \widehat{H}_{(\lceil np_j \rceil)})$ do Kendall Plot não pertencer a banda de confiança, caracteriza-se a rejeição da hipótese nula.

No trabalho com cópulas Arquimedianas, devido a complexidade das funções de Kendall K , fica impossibilitado o cálculo do verdadeiro valor de ξ_p^0 , por isso trabalha-se com sua estimativa calculada integração numérica $W_{\lceil np \rceil; n, K_0} = E(H_{(\lceil np \rceil)}^0)$. A menos quando

trabalha-se com a cópula 4.2.12 (Tabela 1.1) que apresenta inversa usual para a f.d.a. K (Tabela 2.1). Contudo, a convergência de $Q(K_0, K)$ a distribuição $N(0, 1)$ se preserva ao substituir ξ_p^0 por $\hat{\xi}_p^0 = E(H_{(\lceil np \rceil)}^0)$. Formalmente, temos

$$E(H_{(\lceil np \rceil)}^0) \rightarrow \xi_p^0$$

quando $n \rightarrow \infty$. Logo pelo Teorema de Slutsky (James (2004) [27]), temos

$$\frac{\sqrt{n}(\hat{H}_{(\lceil np \rceil)} - \xi_p)k_0\{E(H_{(\lceil np \rceil)}^0)\}}{\sqrt{p(1-p)}} \xrightarrow{D} \frac{\sqrt{n}(\hat{H}_{(\lceil np \rceil)} - \xi_p)k_0(\xi_p^0)}{\sqrt{p(1-p)}} = Q(K_0, K)$$

quando $n \rightarrow \infty$.

Exemplo 5.1. Como ilustração de que a substituição de $\xi_p^0 \equiv K_0^{-1}(p)$ pela estimativa $\hat{\xi}_p^0 = E(H_{(\lceil np \rceil)}^0)$ não influi no comportamento das bandas de confiança do Kendall Plot a Figura 5.1 mostra que os Kendall Plots são praticamente idênticos para o caso da cópula 4.2.12. Além de permitir verificar que sob a hipótese \mathbf{H}_0 verdadeira os pontos do Kendall Plot encontram-se no interior da banda de confiança. A Figura 5.2 apresenta um Kendall Plot quando \mathbf{H}_0 é falsa, verificando que alguns de seus pontos encontram-se fora da banda de confiança conforme esperado caracterizando a rejeição da hipótese nula.

Exemplo 5.2. Esboça-se dois Kendall plots ao nível de significância $\alpha = 0.05$ sob hipótese nula de cópula de Clayton $\theta = 20$, equivalente a $\tau \approx 0.9$ conforme Exemplo 3.1, para amostras de tamanho $n = 100$ e $n = 500$ com cópula de Clayton $\theta = 6$, equivalente a $\tau = 0.75$ conforme Exemplo 3.1. Observando a Figura 5.3 vê-se que para o tamanho amostral $n = 100$ o Kendall Plot não apresenta evidências para rejeição da hipótese nula, embora os gráficos das amostras $n = 100$ com cópulas de Clayton $\theta = 20$ e $\theta = 6$ manifestem claramente que estas estruturas de dependência são diferentes (Figura 5.4). Ao contrário, para $n = 500$ o Kendall Plot apresenta evidências para rejeição da hipótese nula. Essa é uma indicativa de que o poder do teste de fato aumenta com o tamanho amostral.

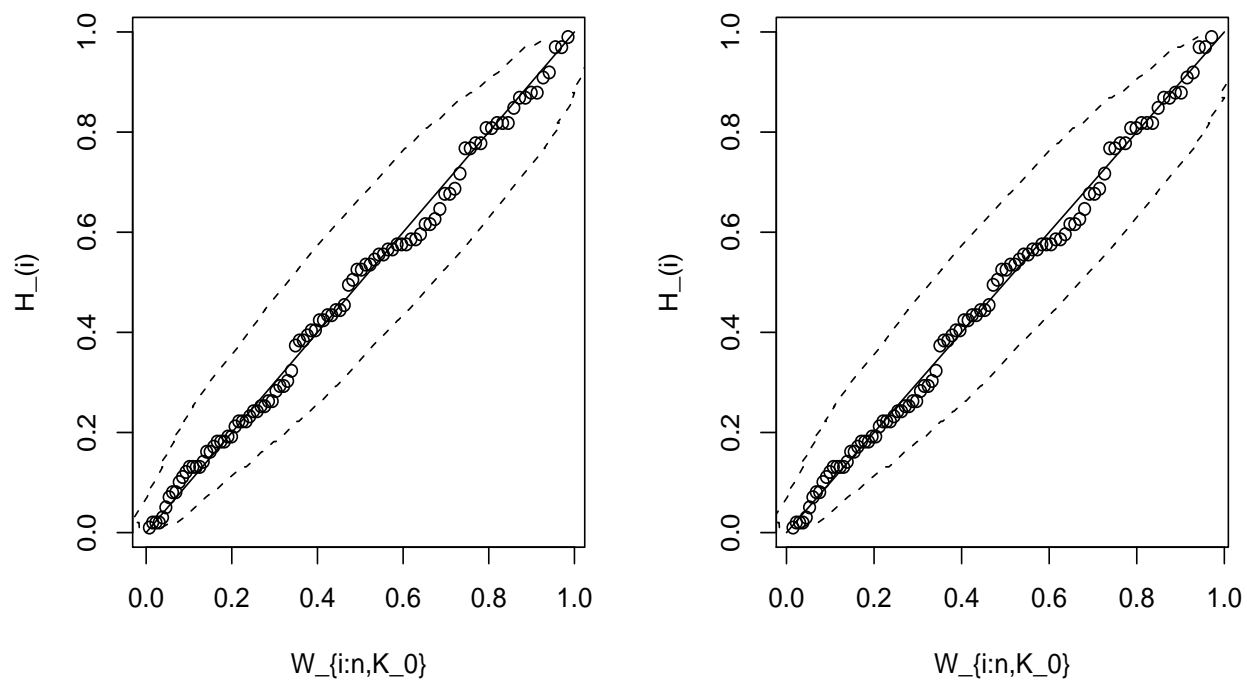


Figura 5.1: Kendall Plots sob hipótese nula de cópula 4.2.12 com $\theta = 3$, nível de significância $\alpha = 0.05$, de amostra aleatória $n = 100$ associada a mesma cópula $\theta = 3$. Primeira figura: $(\xi_p^0, \widehat{H}_{(\lceil np \rceil)})$. Segunda figura: $(\widehat{\xi}_p^0, \widehat{H}_{(\lceil np \rceil)})$.

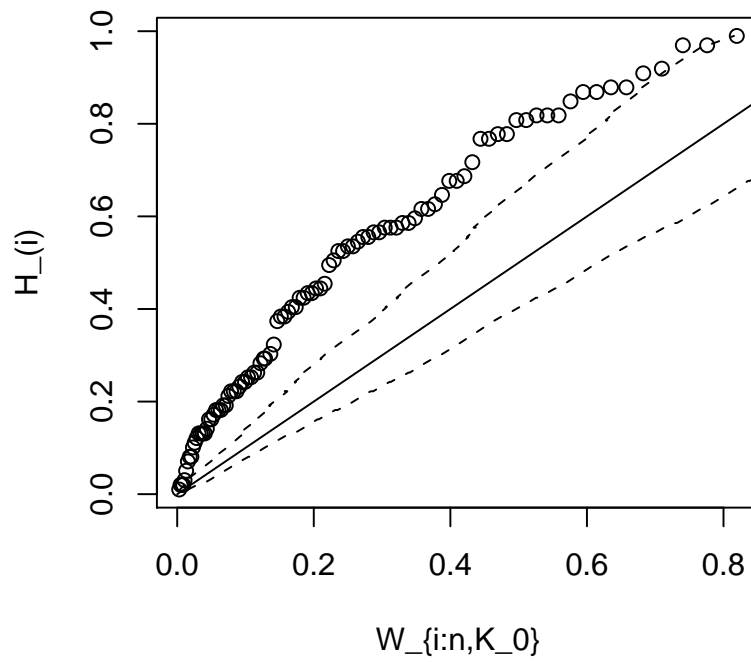


Figura 5.2: Kendall Plot $(\hat{\xi}_p^0, \hat{H}_{(\lceil np \rceil)})$ sob hipótese nula de cópula de independência, nível de significância $\alpha = 0.05$, de amostra aleatória $n = 100$ associada a cópula 4.2.12 com $\theta = 3$.

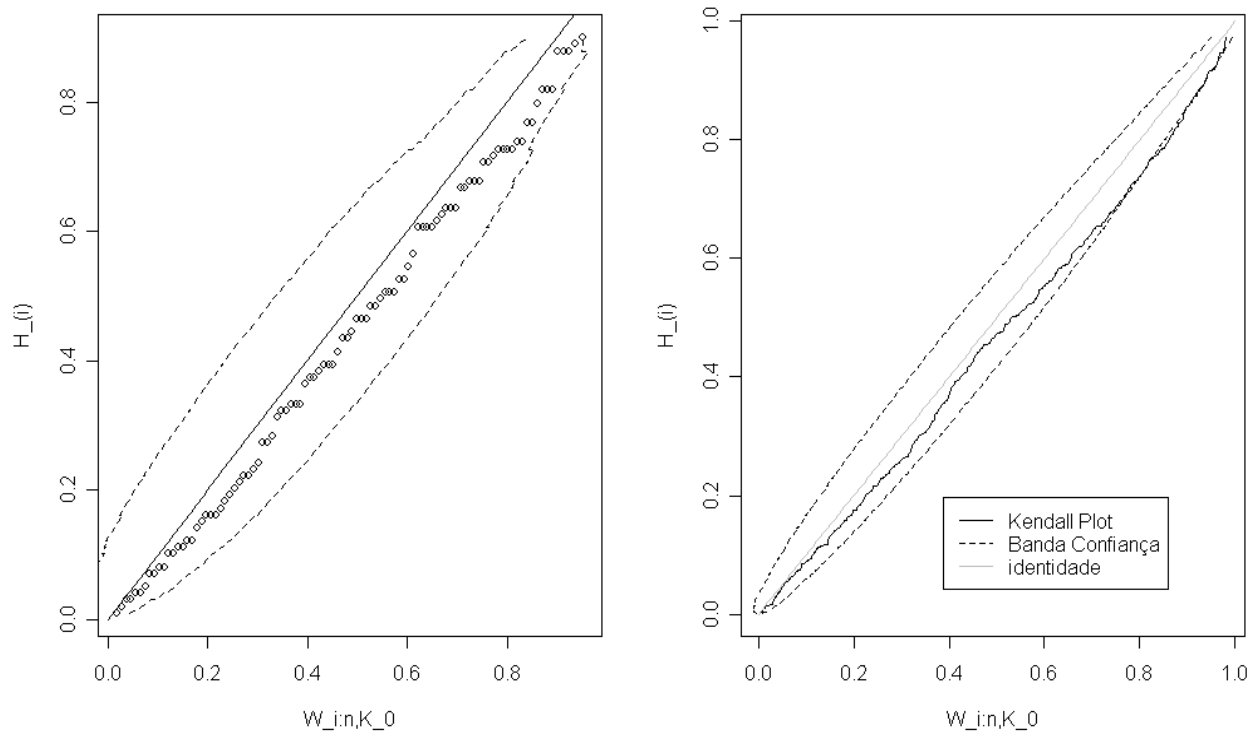


Figura 5.3: Kendall Plot $(\hat{\xi}_p^0, \hat{H}_{(\lceil np \rceil)})$ sob hipótese nula de cópula de Clayton $\theta = 20$, nível de significância $\alpha = 0.05$, de amostras aleatórias $n = 100$ (primeira figura) e $n = 500$ (segunda figura) associadas a cópula de Clayton $\theta = 6$.

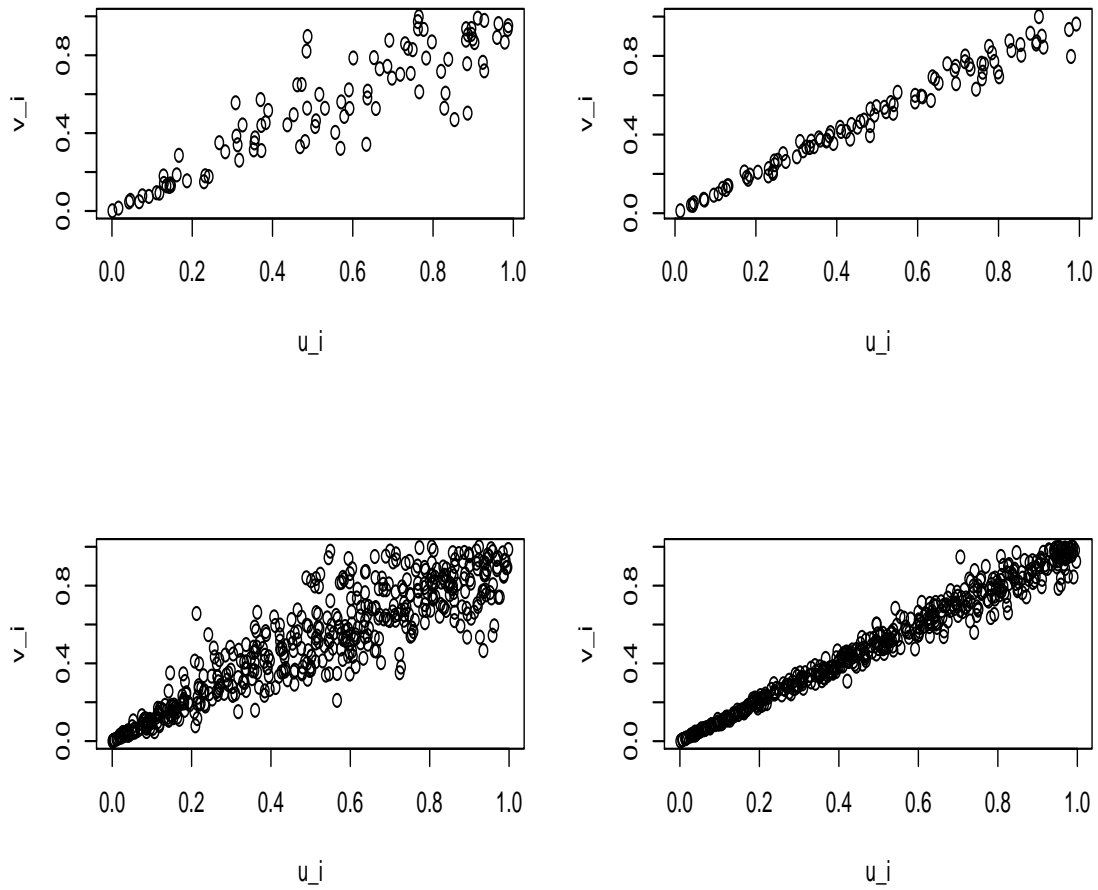


Figura 5.4: Gráficos de amostras tamanho $n = 100$ com cópula de Clayton com $\theta = 6$ e $\theta = 20$ (gráficos acima da esquerda a direita). Gráficos de amostras tamanho $n = 500$ com cópula de Clayton com $\theta = 6$ e $\theta = 20$ (gráficos abaixo da esquerda a direita).

5.2 Poder do Teste

Com o objetivo de analisar a função poder do teste se consideram dados simulados associados às copulas de Clayton, Gumbel, Frank e copula 4.2.12 com os níveis de associação τ de Kendall $\tau \approx 0.3$ e $\tau \approx 0.7$ sob hipótese nula de copula de Clayton e Gumbel com idênticos níveis de associação τ ao nível de significância $\alpha = 0.05$. Considera-se os tamanhos amostras $n = 150$ e $n = 500$. Veja nas Tabelas 5.2, 5.3, 5.4 e 5.5 o esquema de simulação, bem como a porcentagem de rejeição de cada um dos testes analisados, considerando o teste Kendall Plot resultante da intersecção de $n - 2$ testes com os p_j -quantis dados por $p_j = (j - 1)/(n - 1)$, $j = 2, \dots, n - 1$. As Tabelas consideram o teste Kendall Plot resultante da intersecção de 20 testes com os p_j -quantis, $j = 1, \dots, 20$, dados por $p_j = 0.05j - 0.025$.

Toda a análise do poder do teste é feita em função da medida τ de Kendall associada às famílias de copulas Arquimedianas, porém é necessário para a construção do Kendall Plot determinar o valor do parâmetro θ_0 sob hipótese nula. O parâmetro θ_0 foi calculado usando o Exemplo 3.1 que expressa τ como uma função de θ , logo a inversa desta função é um modo de calcular θ em função de τ .

A função poder do teste é dada pela probabilidade de rejeitar a hipótese nula $\mathbf{H}_0 : C_\theta = C_{\theta_0}$ dado que a hipótese nula é verdadeira. Quando a verdadeira copula C_θ é a copula sob hipótese nula o poder do teste deve ser próximo ao nível de significância do teste. Por se desconhecer a distribuição de $Q(K_0, K)$ (equação 5.8) sob H_0 quando a verdadeira distribuição K é diferente da distribuição K_0 da copula sob hipótese nula é utilizada a probabilidade de rejeição empírica como estimativa da função poder do teste gráfico. Cada um dos testes é realizado 5.000 vezes com base em 5.000 amostras seguindo copula determinada pela respectiva hipótese alternativa. A probabilidade de rejeição empírica de cada teste é dada pelo número de vezes que \mathbf{H}_0 foi rejeitada dividido pelo número total de testes realizados que é igual a 5.000.

5.2.1 Análise dos resultados

Espera-se que o teste gráfico Kendall Plot consiga detectar as diferenças entre as estruturas de dependência analisadas. A estrutura de dependência de uma copula C é analisada pela forma como os pontos $\{u_i, v_i\}$, seguindo copula C , concentram-se em torno da diagonal principal (dependência positiva) ou diagonal secundária (dependência negativa). Difer-

entes estruturas de dependência são caracterizadas por padrões de concentração em torno da diagonal: afunilamento na cauda superior e/ou inferior e simetria na concentração em torno da diagonal secundária. A Tabela 5.1 apresenta os padrões de dependência apresentados para cada copula Arquimediana analisada.

<i>copula</i>	<i>simetria</i>	<i>afunilamento na cauda</i>
<i>Clayton</i>	não	inferior
<i>Frank</i>	sim	não
4.2.12	não	superior e inferior*
<i>Gumbel</i>	não	superior

Tabela 5.1: Algumas características das estruturas de dependência de algumas copulas Arquimedianas (*a manifestação do afunilamento na cauda superior depende do nível de associação τ , veja Figura 5.5)

A estrutura de dependência da copula de Clayton é caracterizada por apresentar dependência caudal inferior, ou seja, é mais provável que ocorram eventos extremos na cauda inferior da distribuição que na cauda superior. O contrário acontece para a copula de Gumbel que apresenta dependência caudal superior. A copula de Frank não apresenta dependência caudal, é simétrica em relação a diagonal secundária. A copula 4.2.12, em geral, é caracterizada pela presença de ambas dependências do tipo caudal superior e caudal inferior, observando que a manifestação de dependência caudal superior depende do nível de associação τ : quanto maior τ mais forte a dependência caudal superior (e inferior). Veja as Figuras 5.6 e 5.7 para ilustração, apresentando as quatro copulas citadas com mesmo nível de associação $\tau \approx 0.7$ porém é clara a diferença entre suas estruturas de dependência. Sendo então que as copulas utilizadas para análise do teste apresentam estruturas de dependência distintas entre si. Logo são esperadas altas porcentagens de rejeição pelo Kendall Plot para os diferentes testes em questão.

5.2.2 Intersecção de $n - 2$ testes

As porcentagens de rejeição dos testes referentes às hipóteses nulas de copula de Gumbel encontram-se nas Tabelas 5.2 e 5.3 e as porcentagens de rejeição referentes às hipóteses nulas de Clayton encontram-se nas Tabelas 5.4 e 5.5.

De maneira geral a estimativa para o nível de significância do teste é inferior ao nível nominal $\alpha = 0.05$. Este fato se deve ao nível do teste gráfico ser determinado pelo método

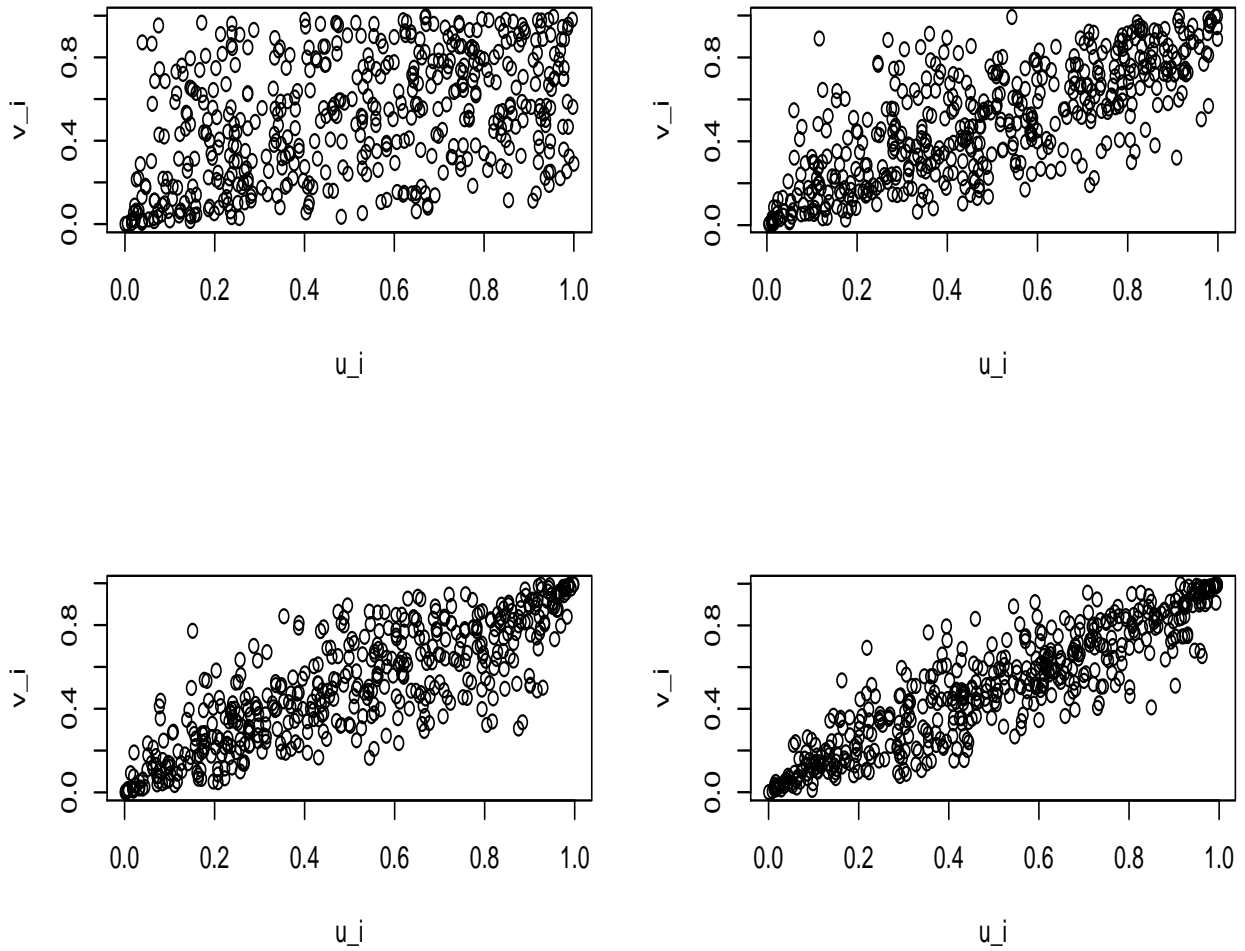


Figura 5.5: Amostras $n = 500$ da copula 4.2.12 para $\theta = \{1, 1.5, 2, 2.5\}$ associado respectivamente a $\tau \approx \{0.3, 0.5, 0.6, 0.7\}$ (da esquerda para direita).

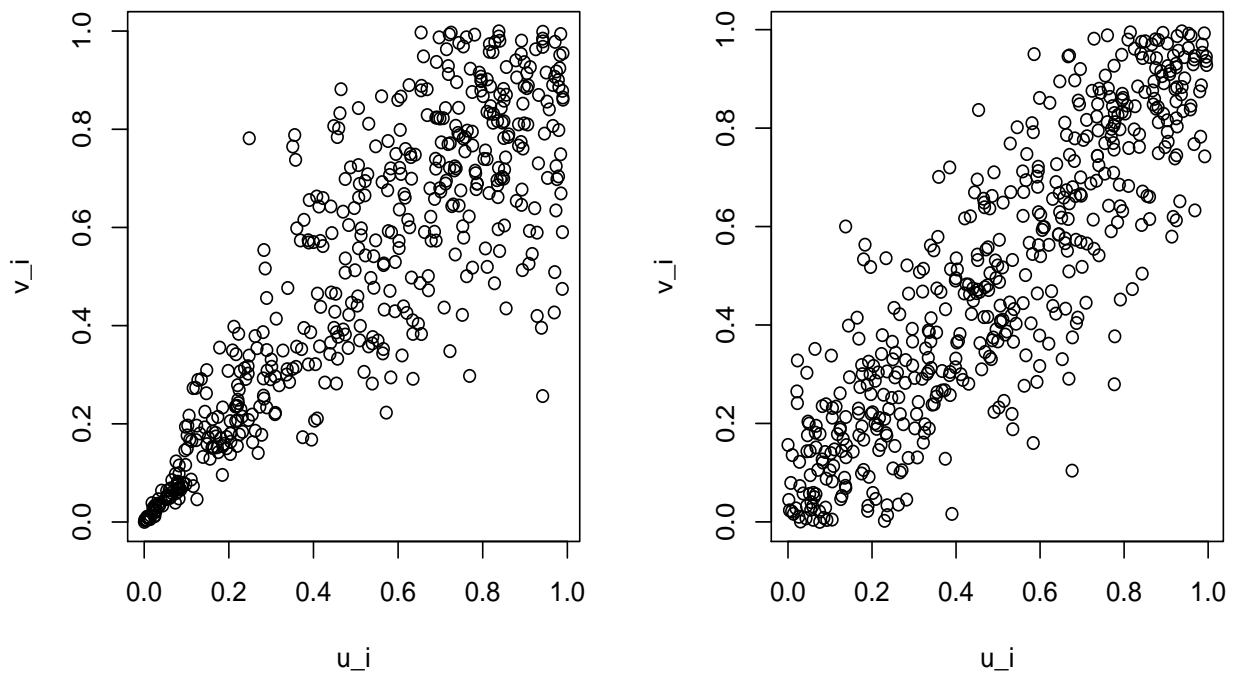


Figura 5.6: Amostras $n = 500$ das copulas de Clayton e Frank com nível de associação $\tau \approx 0.7$ (da esquerda para direita).

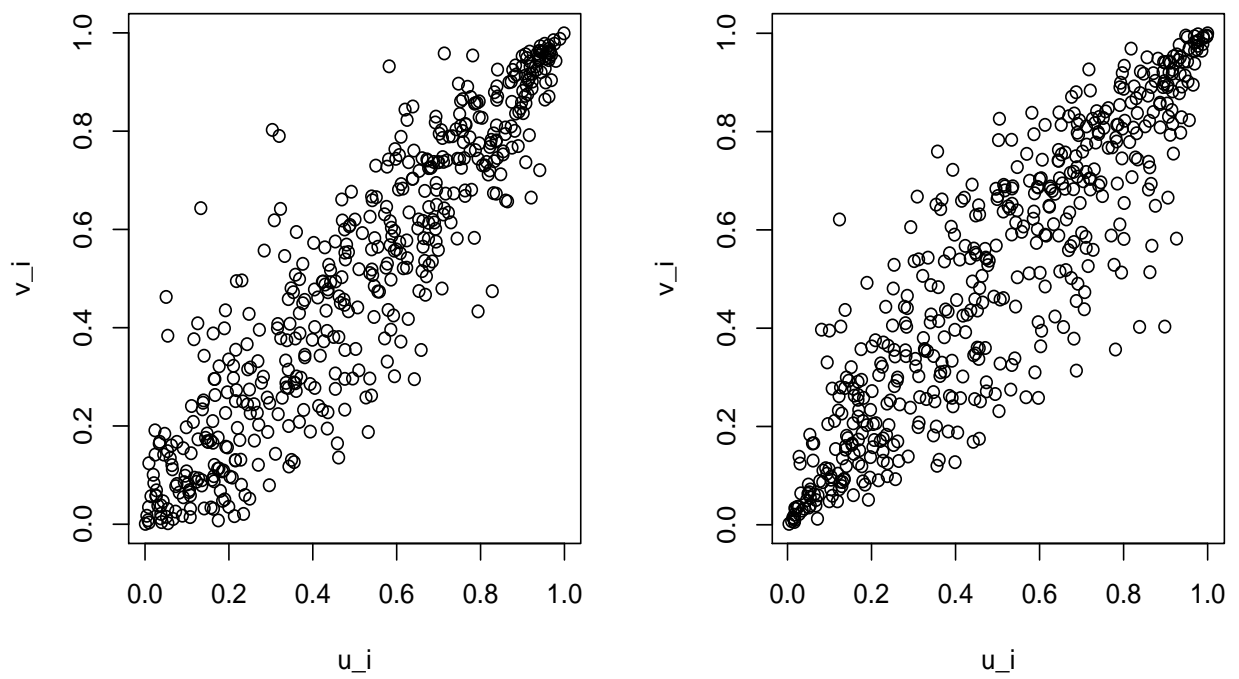


Figura 5.7: Amostras $n = 500$ das copulas de Gumbel e 4.2.12 com nível de associação $\tau \approx 0.7$ (da esquerda para direita).

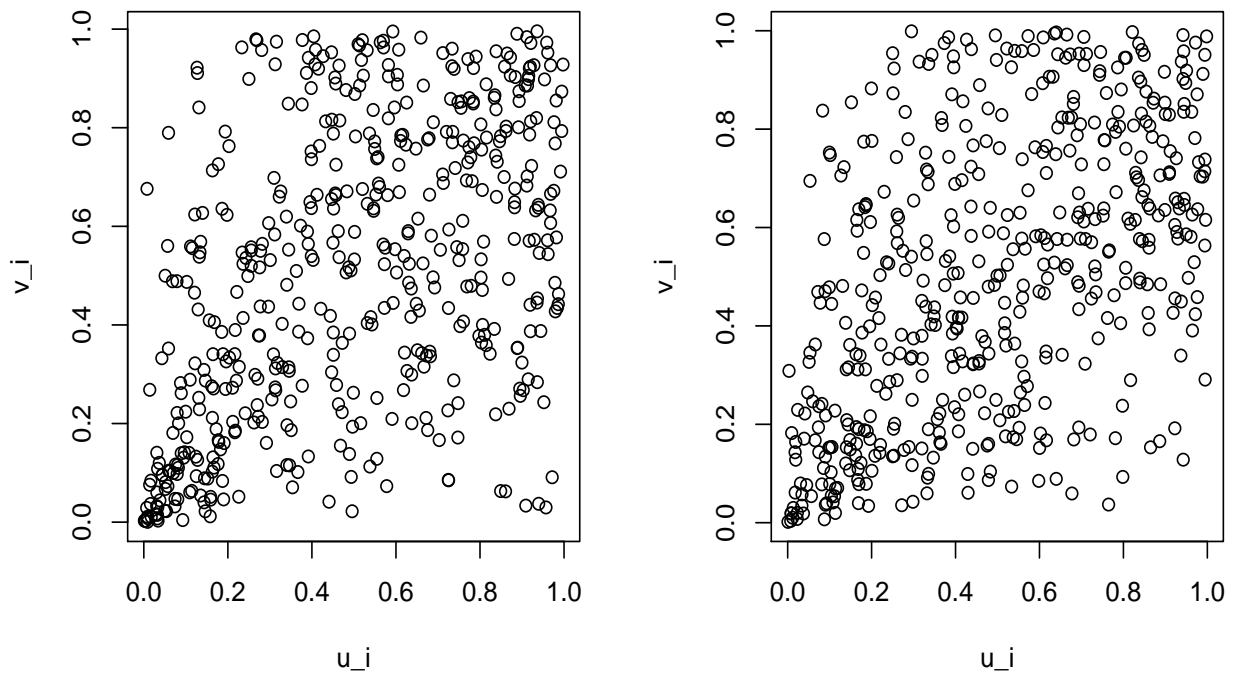


Figura 5.8: Amostras $n = 500$ das copulas de Clayton e 4.2.12 com nível de associação $\tau \approx 0.3$ (da esquerda para direita).

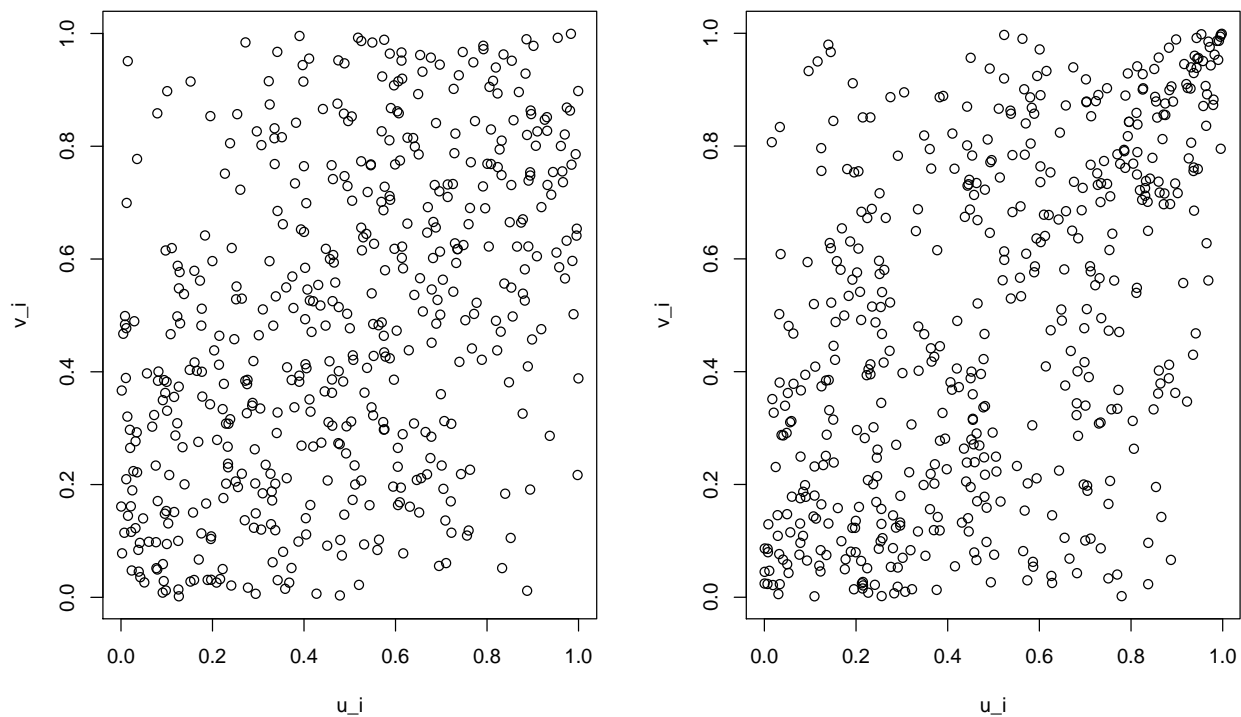


Figura 5.9: Amostras $n = 500$ das copulas de Frank e Gumbel com nível de associação $\tau \approx 0.3$ (da esquerda para direita).

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Clayton</i>	77.5	99.9
<i>Frank</i>	31.3	91.4
4.2.12	77.3	99.9
<i>Gumbel</i>	1.1	0.4

Tabela 5.2: Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Clayton</i>	81.7	100
<i>Frank</i>	38.7	97.5
4.2.12	1.6	1.8
<i>Gumbel</i>	0.02	0

Tabela 5.3: Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$.

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Gumbel</i>	31.3	98.6
<i>Frank</i>	3	68.1
4.2.12	0.3	0.4
<i>Clayton</i>	0.32	0.26

Tabela 5.4: Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.

de correção de Bonferroni (apêndice 6) que determina que a probabilidade de rejeitar \mathbf{H}_0 quando \mathbf{H}_0 é verdadeira é menor ou igual ao nível α especificado para o teste.

Em geral, para ambas hipóteses nulas de Clayton e Gumbel o poder do teste apresenta altas porcentagens de rejeição para as diferentes estruturas de dependência consideradas, bem como o poder do teste aumenta com o tamanho amostral. Embora o poder do teste

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Gumbel</i>	41.8	100
<i>Frank</i>	0.9	55.4
4.2.12	4.5	96
<i>Clayton</i>	0.7	0.29

Tabela 5.5: Intersecção de $n - 2$ testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$.

aumente com o tamanho amostral o nível de significância do teste diminui. O teste gráfico é composto por uma intersecção de $n - 2$ testes baseados em intervalos de confiança. Cada um destes testes é construído com nível de significância $\alpha^* = \alpha / (n - 2)$ (de acordo com o método de correção de Bonferroni), logo quanto maior o tamanho amostral menor será a probabilidade de rejeitar a hipótese nula dado que a hipótese nula é verdadeira para cada um destes testes. Sendo o teste gráfico composto pela intersecção dos $n - 2$ testes, o nível de significância global α também decrescerá com n .

Sob hipótese nula de copula de Clayton observa-se grandes diferenças entre as porcentagens de rejeição para os tamanhos amostrais $n = 150$ e $n = 500$. Este fato indica que o tamanho amostral $n = 150$ é considerado pequeno para a validade do Kendall Plot por este ser um método assintótico. Para este tamanho amostral melhores resultados são obtidos ao considerar como hipótese nula a copula de Gumbel.

Ao comparar a copula de Frank à copula de Clayton ou Gumbel (para ambos níveis de associação analisados) a diferença entre estas estruturas de dependência está localizada em uma das caudas. Entre Gumbel e Frank a diferença encontra-se na cauda superior e entre Clayton e Frank a diferença encontra-se na cauda inferior. Porém observa-se que as porcentagens de rejeição ao comparar Gumbel e Frank (91.4% e 97.5% para $n = 500$) são muito maiores que as porcentagens de rejeição ao comparar Clayton e Frank (68.1% e 55.4% para $n = 500$). Este fato está relacionado a questão de se verificar qual tamanho amostral é suficiente para a eficiência do Kendall Plot. Além de se observar que sob hipótese nula de Clayton o tamanho amostral $n = 150$ é insuficiente, se pode observar que para hipótese nula de Clayton é exigido um maior tamanho amostral que para hipótese nula de Gumbel. Por exemplo, ao analisar o teste para copula de Frank sob hipótese nula de Clayton (ambas copulas com nível de associação $\tau \approx 0.3$) utilizando o tamanho amostral $n = 1000$ se obtém mais de 90% de rejeição. Deste modo, se confirma que baixa porcentagem de rejeição apresentada para distinção entre padrões de dependência

na região da cauda inferior está relacionada ao tamanho amostral insuficiente, logo não constituindo uma deficiência do Kendall Plot em distinguir este tipo de diferença entre estruturas de dependência.

Em dois dos testes realizados o Kendall Plot não consegue identificar diferença entre as estruturas de dependência: ao testar copula 4.2.12 com $\tau \approx 0.7$ sob hipótese nula de Gumbel e copula 4.2.12 com $\tau \approx 0.3$ sob hipótese nula de Clayton. A estrutura de dependência da copula 4.2.12 ao nível de associação $\tau \approx 0.3$ não manifesta claramente a presença de dependência caudal superior, portanto sua estrutura de dependência assemelha-se a estrutura de Clayton (veja Figura 5.8). A semelhança entre ambos padrões de dependência justifica a baixa porcentagem de rejeição apresentada pelo Kendall Plot. Situação semelhante acontece com o teste entre copula de Gumbel e copula 4.2.12 ao nível de associação $\tau \approx 0.7$, apresentando porcentagem de rejeição de 1.8% ($n = 500$). Observando a Figura 5.7 é evidente a característica de presença de dependência caudal superior e inferior em 4.2.12 e dependência caudal superior em Gumbel. Porém observa-se que embora Gumbel não apresente dependência caudal inferior, os pontos localizados nesta região não encontram-se dispersos em relação a diagonal principal assemelhando-se ao padrão de dependência sob a cauda inferior da copula 4.2.12. Sendo assim, a única diferença entre ambas estruturas de dependência em questão, região da cauda inferior, não encontra-se bem definida.

A falta de sensibilidade do Kendall Plot para distinguir estruturas de dependência semelhantes também é evidenciada ao analisar o comportamento deste teste como um teste do parâmetro θ de uma família Arquimediana (veja apêndice 7). Observa-se que embora exista significativa diferença entre θ e θ_0 não significa a existência de grandes diferenças entre os referentes níveis de associação τ e τ_0 . Como o valor de τ quantifica a intensidade da dependência na estrutura de dependência de uma família Arquimediana, valores próximos de τ e τ_0 podem não produzir mudança significativa nesta estrutura. Este fato justifica as baixas porcentagens de rejeição apresentadas pelo Kendall Plot para estes casos. Por exemplo, considerando a estrutura de dependência da copula de Clayton, ao testar $\theta = 8$ ($\tau \approx 0.8$) sob hipótese nula de $\theta_0 = 4.66$ ($\tau \approx 0.7$) a porcentagem de rejeição deste teste é de apenas 2.2% (Tabela 13 do apêndice 7). Pela Figura 5.11 nota-se que o padrão de dependência é o mesmo tanto para θ como para θ_0 . A mesma situação se repete considerando a estrutura de dependência da copula de Gumbel. Ao testar $\theta = 6.66$ ($\tau \approx 0.85$) sob hipótese nula de $\theta_0 = 3.5$ ($\tau \approx 0.7$). A porcentagem de rejeição para este teste é de 0.58% (Tabela 11 do apêndice 7) justificada pela indistinção entre os padrões de dependência associados a θ e θ_0 (Figura 5.10).

Observa-se que as maiores porcentagens de rejeição do Kendall Plot são referentes aos

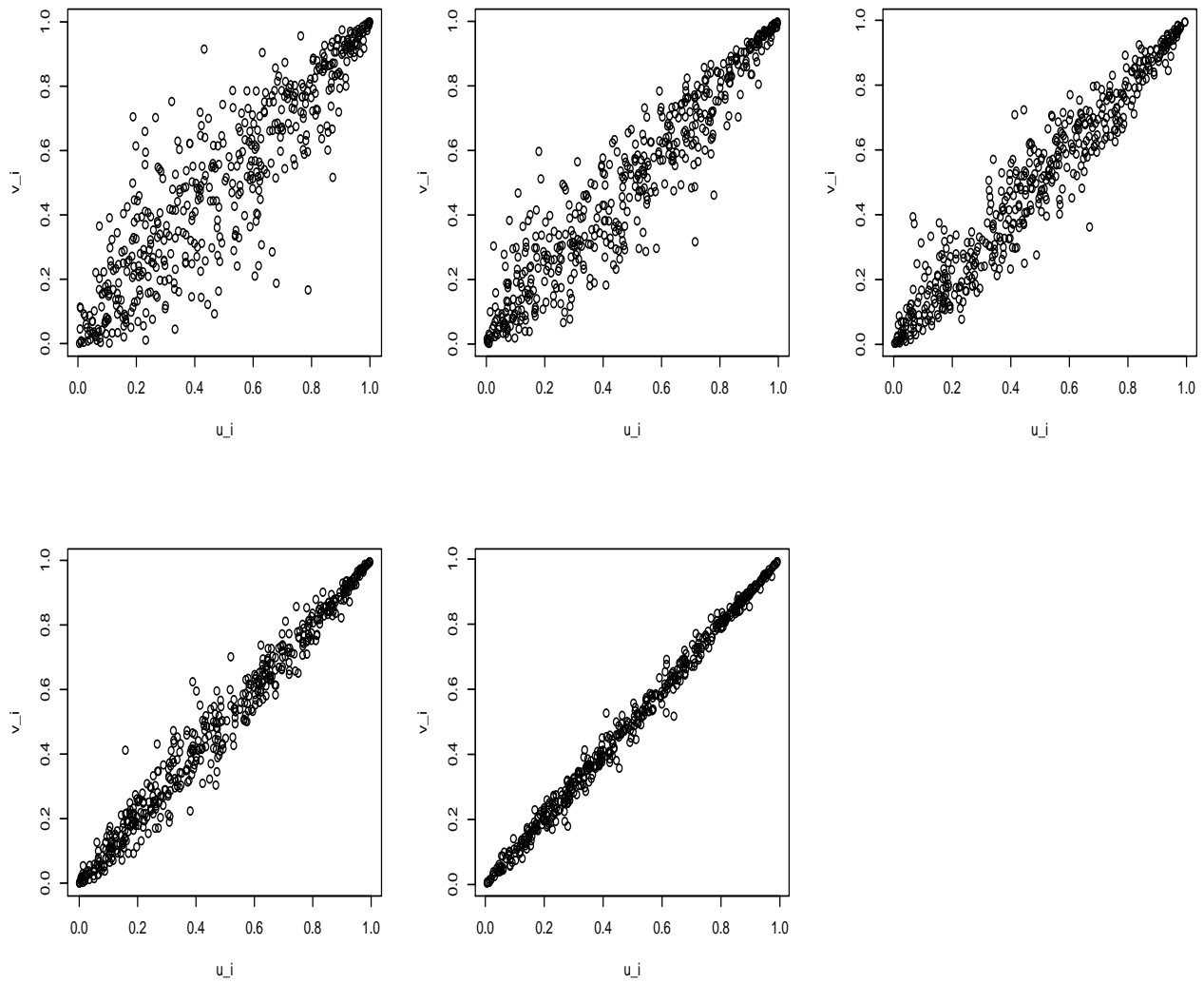


Figura 5.10: Amostras $n = 500$ de copula de Gumbel $\theta = \{3.5, 5, 6.66, 10, 20\}$ (de cima para baixo da esquerda para direita).

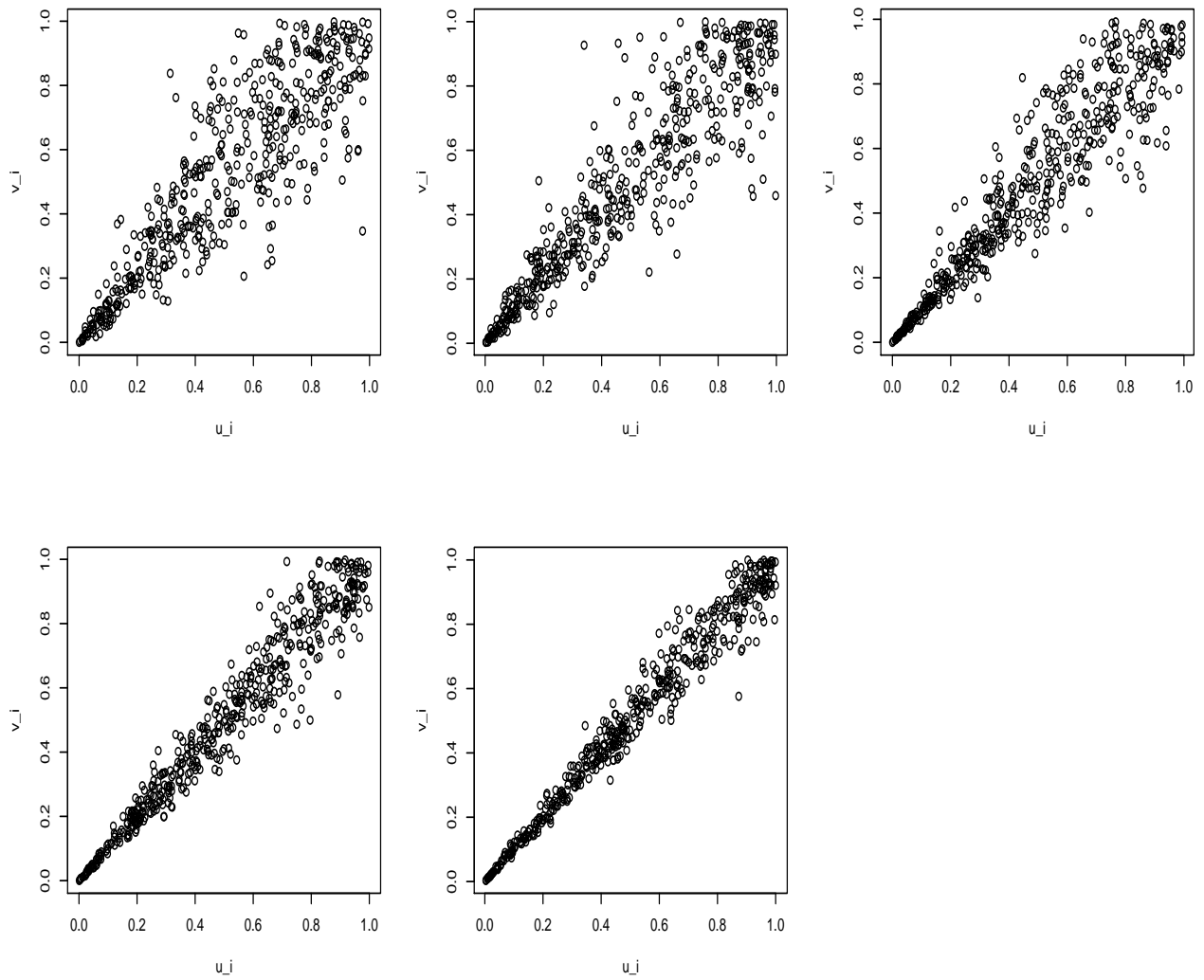


Figura 5.11: Amostras $n = 500$ de copula de Clayton $\theta = \{4.66, 6, 8, 12, 18\}$ (de cima para baixo da esquerda para direita).

testes entre as copulas de Gumbel e Clayton, sendo que a diferença entre estas estruturas de dependência, caracterizada pela clara distinção entre os padrões de dependência sob ambas regiões caudais, se encontra ao longo de toda a diagonal principal.

5.2.3 Intersecção de 20 testes

As porcentagens de rejeição dos testes referentes às hipóteses nulas de copula de Gumbel encontram-se nas Tabelas 5.6 e 5.7 e as porcentagens de rejeição referentes às hipóteses nulas de Clayton encontram-se nas Tabelas 5.8 e 5.9.

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Clayton</i>	61.1	100
<i>Frank</i>	9.4	63.5
4.2.12	61.6	100
<i>Gumbel</i>	0.22	0.22

Tabela 5.6: Intersecção de 20 testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Clayton</i>	59.6	100
<i>Frank</i>	12.8	85.9
4.2.12	0.34	4.56
<i>Gumbel</i>	0	0

Tabela 5.7: Intersecção de 20 testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Gumbel com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$.

Ao realizar o teste Kendall Plot com uma menor quantidade de p_j -quantis, quantidade esta independente do tamanho da amostra considerada, se tem por objetivo viabilizar o uso do Kendall Plot para grandes amostras de maneira que o nível de significância nominal α^* de cada teste individual não seja próximo de zero. Como segundo objetivo, observar se o aumento do nível de significância nominal α^* de cada teste individual acarreta uma

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Gumbel</i>	44.7	99.8
<i>Frank</i>	14.7	95.9
4.2.12	0.2	0.4
<i>Clayton</i>	0.24	0.26

Tabela 5.8: Intersecção de 20 testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.3$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.3$.

<i>copula verdadeira</i>	$n = 150$	$n = 500$
<i>Gumbel</i>	55.2	100
<i>Frank</i>	3.4	94.3
4.2.12	7.5	98.7
<i>Clayton</i>	0.06	0.08

Tabela 5.9: Intersecção de 20 testes: porcentagem de rejeição sob \mathbf{H}_0 copula de Clayton com $\tau \approx 0.7$ de testes para 5.000 amostras de diferentes copulas Arquimedianas de tamanho $n = 150$ e $n = 500$ com $\tau \approx 0.7$.

aproximação do nível de significância empírico do teste ao nível de significância nominal α .

Os resultados são semelhantes ao utilizar $n - 2$ testes. Altas porcentagens de rejeição são exibidas para um suficiente tamanho amostral, com exceção dos casos em que as estruturas de dependência dos dados e sob hipótese nula são semelhantes como discutido na seção anterior. Há evidências de que a redução no número de testes conjuntos aumente o poder do teste, por exemplo, ao considerar a cópula de Clayton sob hipótese nula, porém não há significativas alterações no nível de significância empírico. A hipótese de que o tamanho amostral $n = 500$ seja insuficiente para analisar o teste é descartada ao analisar o nível de significância empírico para amostras de tamanho $n = 1000$: para a cópula de Clayton com $\tau \approx 0.3$, observou-se $\hat{\alpha} = 0.56\%$ e para cópula de Gumbel com $\tau \approx 0.7$ o nível de significância empírico manteve-se inalterado, $\hat{\alpha} = 0\%$. Estes resultados evidenciam que o Kendall Plot pode ser considerado um teste conservador para se rejeitar uma hipótese nula verdadeira.

5.2.4 Conclusão

Pode-se inferir que o Kendall Plot é um eficiente teste para acusar diferenças entre padrões de dependência, em especial sob as caudas da distribuição. Por ser um teste assintótico sua eficiência também está relacionada ao tamanho amostral. Observou-se que para um tamanho amostral ser considerado suficiente depende de qual hipótese nula é considerada. Logo é sugerida a utilização de maior tamanho amostral possível no Kendall Plot. Ao utilizar grande tamanho amostral sugere-se a realização de comparação de uma quantidade fixa de p_j -quantis, no máximo 20, para que não se anule o nível de significância nominal de cada teste individual de quantis. Por se tratar de um teste caudal, o Kendall Plot pode ser usado, por exemplo, para dados financeiros que são caracterizados por esta estrutura de dependência.

Apêndices

.1 Resultado sobre Esperança de uma v.a. aleatória

Teorema .2. (James (2004) [27]) *Seja F a f.d.a da v.a. X que assume somente valores não negativos, então*

$$E(X) = \int_0^{\infty} [1 - F(x)]dx$$

Demonstração. Para a prova usa-se integração por partes, com a diferencial $d(xF(x)) = x dF(x) + F(x)dx$:

$$\forall b > 0, \int_0^b x dF(x) = bF(b) - \int_0^b F(x)dx = \int_0^b [F(b) - F(x)]dx$$

Como $F(b) \leq 1$ e $1 - F(x) \geq 0$, temos

$$\int_0^b x dF(x) = \int_0^b [F(b) - F(x)]dx \leq \int_0^b [1 - F(x)]dx, \quad \forall b > 0,$$

de modo que

$$\int_0^{\infty} x dF(x) = \lim_{b \rightarrow \infty} \int_0^b x dF(x) \leq \int_0^{\infty} [1 - F(x)]dx$$

Por outro lado, seja $\lambda > 0$. Se $b > \lambda$, então

$$\begin{aligned}
\int_0^b [F(b) - F(x)]dx &\geq \int_0^\lambda [F(b) - F(x)]dx \\
&= \int_0^\lambda [F(b) - 1]dx + \int_0^\lambda [1 - F(x)]dx \\
&= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx,
\end{aligned}$$

e portanto,

$$\begin{aligned}
\int_0^\infty x dF(x) &= \lim_{b \rightarrow \infty} \int_0^b x dF(x) = \lim_{b \rightarrow \infty} [F(b) - F(x)]dx \\
&\geq \int_0^\lambda [1 - F(x)]dx + \lim_{b \rightarrow \infty} \lambda[F(b) - 1] = \int_0^\lambda [1 - F(x)]dx.
\end{aligned}$$

Como a equação acima vale para todo $\lambda > 0$, temos

$$\int_0^\infty x dF(x) \geq \lim_{\lambda \rightarrow \infty} \int_0^\lambda [1 - F(x)]dx = \int_0^\infty [1 - F(x)]dx.$$

□

.2 Convergência do quantil amostral ao quantil teórico de uma distribuição

Teorema .3. (Sen (1993) [37]). *Seja $\{X_1, \dots, X_n\}$ uma amostra aleatória de uma v.a. com distribuição F , sendo o p -ésimo quantil ξ_p , $0 < p < 1$ unicamente definido por $\xi_p = F^{-1}(p)$. Definido o p -ésimo quantil amostral como $X_{(q_p)}$ com*

$$q_p = [np] + 1 \quad \text{ou} \quad q_p = [(n + 1)p]$$

temos

$$X_{(q_p)} \rightarrow \xi_p$$

em probabilidade e quase certamente quando $n \rightarrow \infty$.

Demonstração. Para todo $\epsilon > 0$, seja $p_\epsilon = F(\xi_p + \epsilon)$ sendo então $p_\epsilon > p$. Observe que

$$\begin{aligned} P(X_{(q_p)} \leq \xi_p + \epsilon) &= P(q_p \text{ ou mais entre } (X_1, \dots, X_n) \leq \xi_p + \epsilon) \\ &= P\left(\sum_{i=1}^n \mathbb{I}\{X_i \leq \xi_p + \epsilon\} \geq q_p\right) \\ &= P\left(\frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq \xi_p + \epsilon\} - p_\epsilon \geq \frac{q_p}{n} - p_\epsilon\right) \\ &\equiv P(U_n \geq a_n) \end{aligned}$$

onde $U_n \equiv \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq \xi_p + \epsilon\} - p_\epsilon \right\} \xrightarrow{p} 0$ quando $n \rightarrow \infty$ pela Lei Fraca de Kintchin (James (2004) [27]), pois a média amostral das indicadores converge para

$$E(\mathbb{I}\{X_i \leq \xi_p + \epsilon\}) = P(X_i \leq \xi_p + \epsilon) = F(\xi_p + \epsilon)$$

Sendo $a_n \equiv q_p/n - p_\epsilon$, temos $a_n \rightarrow a = p - p_\epsilon < 0$, logo $\forall \epsilon' > 0 \exists n_0(\epsilon')$ tal que quando $n \rightarrow \infty$, $|a_n - a| < \epsilon'$ e segue que

$$\begin{aligned} P(U_n \geq a_n) &= P(U_n - a \geq a_n - a) \geq P(U_n - a \geq -\epsilon) \\ &= P(U_n \geq a - \epsilon') \geq P(\underbrace{a - \epsilon'}_{<0} \leq U_n \leq \underbrace{-a + \epsilon'}_{>0}) \xrightarrow{n \rightarrow \infty} 1 \end{aligned}$$

Fazendo $p_\epsilon^* = F(\xi_p - \epsilon)$ temos $p_\epsilon^* < p$ e procedendo analogamente mostra-se que

$$P(X_{(q_p)} \leq \xi_p - \epsilon) = P(-X_{(q_p)} + \xi_p > \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

portanto

$$P(|X_{(q_p)} - \xi_p| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

provando que $X_{(q_p)} \rightarrow \xi_p$ em probabilidade.

Para provar a convergência quase certa considere

$$A_n = \{X_{(q_p)} \geq \xi_p + \epsilon\} = \left\{ \sum_{i=1}^n \mathbb{I}\{X_i \leq \xi_p + \epsilon\} \leq q_p \right\} = \{nF_n(\xi_p + \epsilon) \leq q_p\}$$

e

$$B_n = \{X_{(q_p)} \leq \xi_p - \epsilon\} = \left\{ \sum_{i=1}^n \mathbb{I}\{X_i \leq \xi_p - \epsilon\} \geq q_p \right\} = \{nF_n(\xi_p - \epsilon) \geq q_p\}$$

Note que $nF_n(\xi_p \pm \epsilon)$ segue distribuição $Bin(n, F(\xi_p \pm \epsilon))$. Usando a desigualdade de Bernstein para eventos binominais (Sen & Singer (1993) [37]) temos

$$P(A_n) \leq \{\rho[\epsilon, F(\xi_p)]\}^n$$

$$P(B_n) \leq \{\rho^*[\epsilon, F(\xi_p)]\}^n$$

onde $0 \leq \{\rho[\epsilon, F_n(\xi_p)]\}, \{\rho^*[\epsilon, F(\xi_p)]\} \leq 1$. Logo temos $\sum P(A_n) < \infty$ e $\sum P(B_n) < \infty$, então pelo teorema de Borel Cantelli (James (2004) [27]) os eventos A_n e B_n ocorrem infinitas vezes com probabilidade zero. Desta maneira o evento

$$\{|X_{(q_p)} - \xi_p| > \epsilon\} = A_n \cup B_n$$

ocorre infinitas vezes com probabilidade zero, ou seja, $X_{(q_p)} \rightarrow \xi_p$ quase certamente. \square

.3 Expressão para $W_{i:n, K_0} \equiv E(H_{(i)}^0)$

De acordo com Casella & Berger (2002) [5], sejam $H_{(i)}^0, \dots, H_{(n)}^0$ estatísticas de ordem de uma amostra H_i^0, \dots, H_n^0 de uma BIPIT H com distribuição K e densidade k . Então a densidade da estatística de ordem $H_{(i)}^0$ é dada por

$$f_{H_{(i)}^0}(v) = \frac{n!}{(i-1)!(n-i)!} k(v)[K(v)]^{i-1}[1-K(v)]^{n-i}$$

Desta forma temos

$$\begin{aligned}
E(H_{(i)}^0) &= \int_0^1 w f_{H_{(i)}^0}(w) dw \\
&= n \binom{n-1}{i-1} \int_0^1 w \{K_0(w)\}^{i-1} \{1 - K_0(w)\}^{n-i} dK_0(w)
\end{aligned}$$

.4 Integração Numérica

A idéia básica da integração numérica reside na aproximação da função integrando $f(x)$ por um polinômio. A escolha desse polinômio e dos pontos que serão usados na sua determinação vão definir os diversos métodos de integração. Uma vez estabelecida a aproximação polinomial conveniente, o problema seria resolvido pela integração trivial de polinômios.

As fórmulas de integração são somatórios cujas parcelas são valores da função $f(x)$ calculados em pontos e multiplicados por pesos convenientemente escolhidos. Assim, as fórmulas de integração numérica são do tipo:

$$\int_a^b f(x) dx \cong \sum_{j=0}^n w_j f(x_j)$$

onde $a \leq x_0 < x_1 < \dots < x_n \leq b$ são chamados *pontos de integração* e w_j são os *pesos* da fórmula de integração. Para cada fórmula de integração, são estabelecidos os pesos e os pontos de integração correspondentes ao critério de aproximação adotado. A análise do comportamento da função integrando é de extrema importância para determinar quais os pontos de integração a serem usados e seus respectivos pesos na escolha da fórmula de integração adequada.

Em particular, o integrando para o cálculo de

$$W_{i;n,K_0} \equiv E(H_{(i)}^0) = n \binom{n-1}{i-1} \int_0^1 w \{K_0(w)\}^{i-1} \{1 - K_0(w)\}^{n-i} dK_0(w)$$

é expresso por produto de potências (que dependem de n) de valores entre 0 e 1 ($K_0(w)$), logo assumindo valores não negativos muito próximos de zero em todo seu domínio para qualquer $i \in \{1, 2, \dots, n\}$ e sendo decrescente em n .

Por exemplo, para a cópula da independência Π e $n = 100$ o integrando assume valores desde ordem 10^{-27} até 10^{-100} , e para $n = 400$ chega a atingir a ordem 10^{-200} .

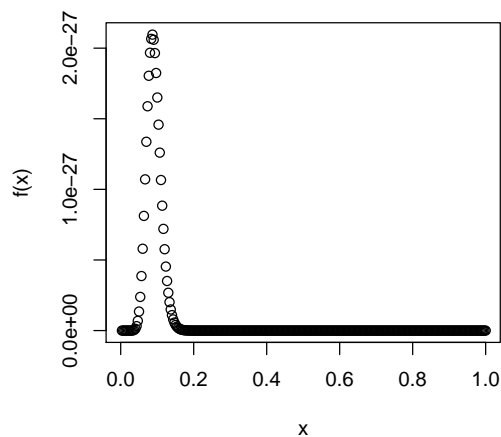


Figura 12: Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 30$.

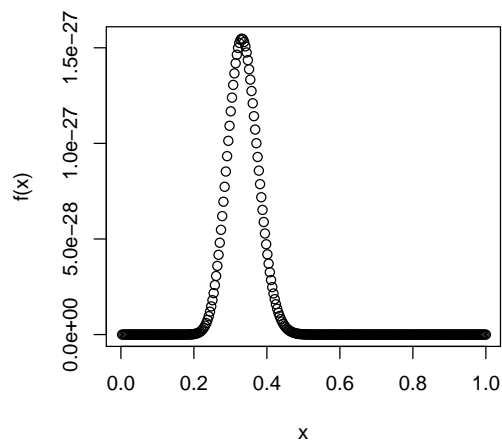


Figura 13: Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 70$.

Como observado pelos gráficos da função integrando (Figuras 12, 13, 14 e 15), dentre os valores assumidos, destaca-se uma parcela de valores com modulo relativamente maior que os demais, que chamaremos de a porção onde o integrando é denso, ou seja, $D = \{w \in [0, 1] | f(w) > f(w^*)\}$, onde $w^* \in [0, 1]$ são os pontos onde o integrando f assume valores muito pequenos. Podemos concluir também que o parâmetro i de $W_{i:n}$ funciona como um parâmetro de locação para o conjunto D .

Quando considerada uma família paramétrica de cópulas, por exemplo cópulas Arquimedianas, a função distribuição de Kendall K é função de um parâmetro $\theta \in \Theta$, logo o comportamento também é influenciado pelo parâmetro θ além dos parâmetros i e n .

Observadas estas peculiaridades do comportamento da função integrando para estimar com eficiência a integral em questão é necessário um método de integração que dentre os pontos de integração escolhidos estejam os pontos pertencentes ao conjunto D para qualquer $i \in \{1, 2, \dots, n\}$.

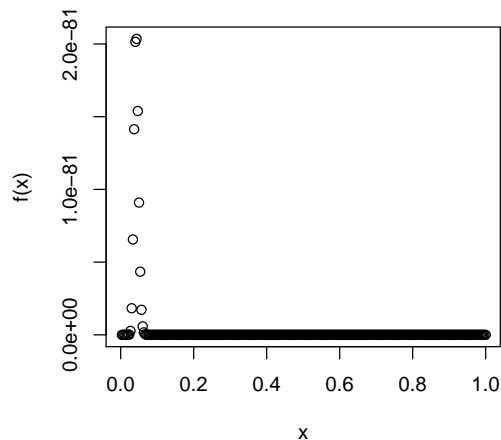


Figura 14: Integrando $f(x)$ para cópula da independência com $n = 400$ e $i = 70$.

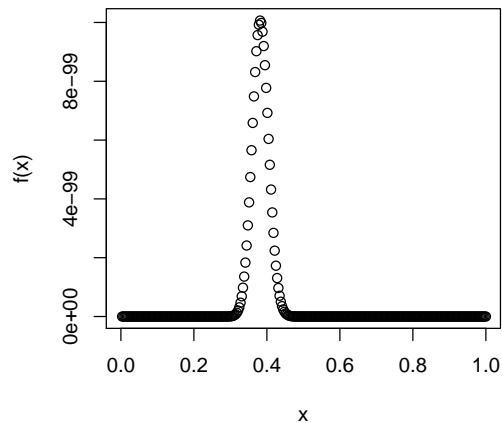


Figura 15: Integrando $f(x)$ para cópula da independência com $n = 100$ e $i = 300$.

A solução usada foi a utilização de *fórmulas de Newton-Cotes* (1711). Nestas fórmulas a idéia de polinômio que aproxime $f(x)$ razoavelmente é que este polinômio interpole $f(x)$ em pontos de $[a, b]$ igualmente espaçados. Desta maneira dividindo o intervalo de integração em um grande número m de subintervalos de comprimento $h = (b - a)/m$ percorre-se todo o domínio de integração sempre sendo escolhidos pontos de integração pertencentes ao conjunto D , não importando sua localização.

O integrando em questão não é definido em pelo menos um dos limites de integração, pois algumas funções de Kendall K de cópulas Arquimedianas não são definidas no ponto 0 ou 1. Para integrandos deste tipo, existem as *fórmulas abertas de Newton-Cotes*. Para o cálculo da integral foi então utilizada a fórmula aberta do trapézio, pertencente a família das fórmulas de Newton-Cotes, dada por

$$\int_a^b f(x)dx = \left(\sum_{j=1}^m f(x_j) \right) (b - a)/m + O\left(\frac{(b - a)^3}{m^2}\right)$$

onde $x_j = a + (j - 1/2)(b - a)/m$. Para esta fórmula usa-se como polinômio interpolador do integrando o polinômio de Lagrange de ordem 1 em torno do ponto $p = (a + b)/2$.

Para maiores detalhes sobre a fórmula do trapézio veja Gray (2001) [26].

Esta fórmula aproxima a integral pela área de retângulos. Para $m = 1$, a integral é aproximada pela área do retângulo de base $[a, b]$ e altura $f((a + b)/2)$. Para $m = 2$, a integral é estimada pela soma da área de dois retângulos: de base $[a, (a + b)/2]$ e altura $f(x_1)$ e base $[(a + b)/2, b]$ e altura $f(x_2)$, onde x_1 e x_2 são os respectivos pontos médios dos segmentos que representam as bases dos retângulos. Generalizando para $m = k$ a integral é estimada pela soma da área de k retângulos de base $[x_{j-1}^*, x_j^*]$, onde $x_j^* = a + j(b - a)/k$, $j = 1, 2, \dots, k$ e respectivas alturas calculadas pelo integrando no ponto médio da base.

Na prática esta regra é frequentemente aplicada calculando a aproximação para integral para uma sequência crescente de valores de m , deste modo transformando-a em um método iterativo. Desde de que a fórmula usa pontos de integração igualmente espaçados, se para iteração k definirmos $m \equiv 3^{(k-1)}$, então triplica-se o número de pontos de integração a cada iteração, porém sendo reutilizados os cálculos do passo anterior. A sequência é usualmente terminada quando a mudança de uma iteração para a próxima é considerada suficientemente pequena.

Algoritmo: Quadratura adaptativa usando formula do trapézio aberta para $\int_a^b f(x)dx$

1. Para cada m_k com $m_k = 3^{(k-1)}$, iterações $k = 1, 2, \dots$, calcula-se

$$T_{m_k} = \sum_{j=1}^{m_k} f(x_j)(b - a)/m_k$$

onde $x_j = a + (j - 1/2)(b - a)/m_k$.

2. Calcula-se T_{m_k} , $k = 1, 2, \dots$ até que a precisão desejada seja atingida, ou seja

$$|T_{m_k} - T_{m_{k-1}}| < \epsilon |T_{m_{k-1}}|$$

onde ϵ representa a precisão desejada. Logo, a integral é dada por T_{m_k} .

.5 Teste bootstrap paramétrico

Dada uma amostra aleatória $\{(X_1, Y_1, \dots, (X_n, Y_n))\}$ com f.d.a. desconhecida \mathbb{H} e

marginais F e G sabe-se que pelo Teorema de Sklar 1.8 os vetores

$$\mathbf{U}_1 = (F(X_1), G(Y_1)), \dots, \mathbf{U}_n = (F(X_n), G(Y_n))$$

são distribuídos de acordo com cópula C associada a \mathbb{H} . Como na prática F e G são desconhecidas, utiliza-se as estimativas para os vetores $\mathbf{U}_1, \dots, \mathbf{U}_n$ dadas por

$$\hat{\mathbf{U}}_1 = (F_n(X_1), G_n(Y_1)), \dots, \hat{\mathbf{U}}_n = (F_n(X_n), G_n(Y_n))$$

para se testar a hipótese

$$\mathbf{H}_0 : C \in \mathbb{C} = \{C_\theta | \theta \in \Theta\} \quad vs. \quad \mathbf{H}_1 : C \notin \mathbb{C} = \{C_\theta | \theta \in \Theta\}$$

onde Θ é o espaço paramétrico.

Foram propostas diversas estatísticas para o teste, como por exemplo, \hat{S}_n baseada na cópula empírica proposta por Genest et. al. (2008) [23] e $\hat{S}_n^{(K)}$ baseada na função de Kendall proposta por Genest et. al. (2006) [20], dadas por

$$\begin{aligned} \hat{S}_n &\equiv n \int_{\mathbf{I}^2} \{C_n(\mathbf{u}) - C_{\hat{\theta}}(\mathbf{u})\}^2 dC_n(\mathbf{u}) \\ &= \sum_{j=1}^n \{C_n(\hat{\mathbf{U}}_j) - C_{\hat{\theta}}(\hat{\mathbf{U}}_j)\}^2 \end{aligned}$$

onde

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(\hat{\mathbf{U}}_i \leq \mathbf{u})$$

e

$$\begin{aligned}\widehat{S}_n^{(K)} &\equiv n \int_{\mathbf{I}^2} \{K_n(v) - K_{\widehat{\theta}}(v)\}^2 dK_{\widehat{\theta}}(v) \\ &= \sum_{j=1}^n \left\{ K_n\left(\frac{j}{n+1}\right) - K_{\widehat{\theta}}\left(\frac{j}{n+1}\right) \right\}^2\end{aligned}$$

onde

$$K_n(v) = \frac{1}{n+1} \sum_{j=1}^n \mathbb{I}\{C_n(\widehat{U}_j) \leq v\}$$

A idéia para a construção de ambas estatísticas parte do caso univariado ao se comparar função distribuição empírica a função distribuição sob hipótese nula. A validade do procedimento bootstrap para a realização do teste está ligada a existência de distribuição limite para as estatísticas. Genest & Rémillard (2008) [22] estabelecem a validade para o procedimento bootstrap para as duas estatísticas apresentadas por meio da análise do comportamento dos processos empíricos

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\widehat{\theta}}) \quad e \quad \mathbb{K}_n = \sqrt{n}(K_n - K_{\widehat{\theta}})$$

através da teoria de funcionais contínuos.

Algoritmo para procedimento bootstrap paramétrico

1. Calcular as pseudo-observações $\{\widehat{\mathbf{u}}_1, \dots, \widehat{\mathbf{u}}_n\}$ - as marginais empíricas da amostra aleatória $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ do vetor aleatório bivariado \mathbf{X} ;
2. Estimar o parâmetro θ com um estimador consistente $\widehat{\theta} = \widehat{\nu}(\widehat{\mathbf{u}}_1, \dots, \widehat{\mathbf{u}}_n)$;
3. Calcule a estatística estimada \widehat{S} ;
4. Para algum inteiro M , repita os seguintes passos para todo $m \in \{1, \dots, M\}$:
 - (a) Gere uma amostra aleatória $\{\mathbf{x}_{1,m}^*, \dots, \mathbf{x}_{n,m}^*\}$ sob hipótese nula $C_{\widehat{\theta}}$ e calcule as pseudo-observações associadas $\{\widehat{\mathbf{u}}_{1,m}^*, \dots, \widehat{\mathbf{u}}_{n,m}^*\}$;

- (b) Estime o parâmetro θ^* com um estimador consistente $\theta_m^* = \hat{\nu}(\hat{\mathbf{u}}_{1,m}^*, \dots, \hat{\mathbf{u}}_{n,m}^*)$;
- (c) Calcule a estatística estimada \hat{S}^* ;
- (d) Se $\hat{S}_{(1)}^* \leq, \dots, \leq \hat{S}_{(M)}^*$ denotam os valores ordenados das estatísticas calculadas no passo 4.(c), uma estimativa do valor crítico do teste no nível α para a estatística \hat{S} é dada por

$$\hat{S}_{[(1-\alpha)M]}^* \tag{10}$$

e o p -valor aproximado para o teste é dado por

$$\hat{p} = \frac{1}{M+1} \sum_{m=1}^M \mathbb{I}\{\hat{S}_m^* > \hat{S}\} \tag{11}$$

O resultado do teste rejeita $\mathbf{H}_0 : C \in \mathbb{C}$ quando $\hat{S} > \hat{S}_{[(1-\alpha)M]}^*$, ou de outro modo, quando $\hat{p} < \alpha$, $\forall \alpha \in \mathbf{I}$ sendo α o nível de significância do teste.

O poder e o tamanho do teste para as estatísticas \hat{S}_n e $\hat{S}_n^{(K)}$ entre outras foi analisado por Genest et. al. (2006) [20] e Genest et. al. (2008) [23] em função do tamanho amostral, grau de dependência e diferentes hipóteses nula e alternativa através de experimentos de Monte Carlo. Ambas estatísticas preservam o tamanho nominal do teste e apresentam boa performance em relação ao poder do teste.

Em estudos futuros, com a mesma idéia das estatísticas presentes na literatura, é possível a criação de uma estatística análoga as apresentadas baseada na diferença entre os elementos do Kendall Plot $H_{(i)}$ e $W_{i:n,K_0}$, ou seja, diferença entre as funções K_n^{-1} e $K_{\hat{\theta}}^{-1}$ sendo que sob hipótese nula verdadeira há convergência em probabilidade entre estas funções. Para isso é necessário o estudo do comportamento da distribuição da estatística para verificar a validade do procedimento bootstrap para o teste.

.6 Correção de Bonferroni

No Capítulo 5 o teste gráfico Kendall Plot foi interpretado como resultante de uma interseção de testes

$$\mathbf{H}_0 : \bigcap_{j \in \mathcal{I}} \mathbf{H}_{0p_j}$$

sendo $p_j \in (0, 1)$, \mathcal{I} um conjunto de índices e cada teste $\mathbf{H}_{0p_j} : K^{-1}(p_j) = K_0^{-1}(p_j)$ construído com nível de significância α^* .

A correção de Bonferroni permite determinar o nível de confiança α^* para cada teste individual \mathbf{H}_{0p_j} de modo que o nível de confiança para o teste \mathbf{H}_0 seja α . Denominando A_{p_j} como o evento que rejeita \mathbf{H}_{0p_j} quando \mathbf{H}_{0p_j} é verdadeira ($P(A_{p_j}) = \alpha^*$) e B o evento que rejeita pelo menos uma das hipóteses \mathbf{H}_{0p_j} quando \mathbf{H}_0 é verdadeira

$$B = \bigcup_{j \in \mathcal{I}} A_{p_j}$$

Desta maneira $P(B)$ é a probabilidade de rejeitar \mathbf{H}_0 quando \mathbf{H}_0 é verdadeira (probabilidade do erro tipo I) e pode ser estimada por

$$P(B) = P\left(\bigcup_{j \in \mathcal{I}} A_{p_j}\right) \leq \sum_{j \in \mathcal{I}} P(A_{p_j}) = |\mathcal{I}| \alpha^*$$

Então se $\alpha^* \equiv \alpha/|\mathcal{I}|$ o nível de confiança para o teste Kendall Plot é α . A utilização deste método para se controlar o nível de confiança α do teste implica em quanto maior o número de testes individuais envolvidos mais conservador será o teste para se rejeitar a hipótese nula quando ela é verdadeira.

.7 Teste gráfico Kendall Plot para o parâmetro θ de uma cópula Arquimediana

O Kendall Plot como um teste para o parâmetro θ é analisado sob as hipóteses nulas de cópula de Clayton e Gumbel. Considera-se θ_0 referente aos níveis de associação $\tau \approx 0.3$ e $\tau \approx 0.7$. A metodologia para a análise do poder do teste segue o padrão apresentado na

seção 5.2 sendo o teste Kendall Plot resultante da intersecção de $n - 2$ testes individuais $\mathbf{H}_{0_{p_j}} : K^{-1}(p_j) = K_{0_{p_j}}^{-1}$, $j = 1, \dots, n - 2$. O esquema de simulação bem como os resultados da análise do poder do teste são apresentados pelas tabelas e gráficos a seguir.

<i>cpula verdadeira</i>	τ	$n = 150$	$n = 500$
$\theta = 1.1$	0.01	57.9	99.6
$\theta = 1.2$	0.17	19.5	65.9
$\theta = 1.3$	0.23	5.4	11.2
$\theta = \mathbf{1.5}$	0.33	1.1	0.4
$\theta = 1.7$	0.4	1.2	1.2
$\theta = 2$	0.5	6.8	52.1
$\theta = 3$	0.66	94.4	100
$\theta = 4$	0.75	100	100
$\theta = 5$	0.8	100	100

Tabela 10: Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Gumbel $\theta = 1.5$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Gumbel com diferentes níveis de associação.

<i>cpula verdadeira</i>	τ	$n = 150$	$n = 500$
$\theta = 1.11$	0.1	100	100
$\theta = 1.5$	0.33	92.1	100
$\theta = 2$	0.5	10.8	81.6
$\theta = 2.5$	0.6	0.8	1.2
$\theta = \mathbf{3.5}$	0.7	0.02	0
$\theta = 5$	0.8	0	0
$\theta = 6.66$	0.85	0	0.58
$\theta = 10$	0.9	0	65.1
$\theta = 20$	0.95	0	100

Tabela 11: Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Gumbel $\theta = 3.5$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Gumbel com diferentes níveis de associação.

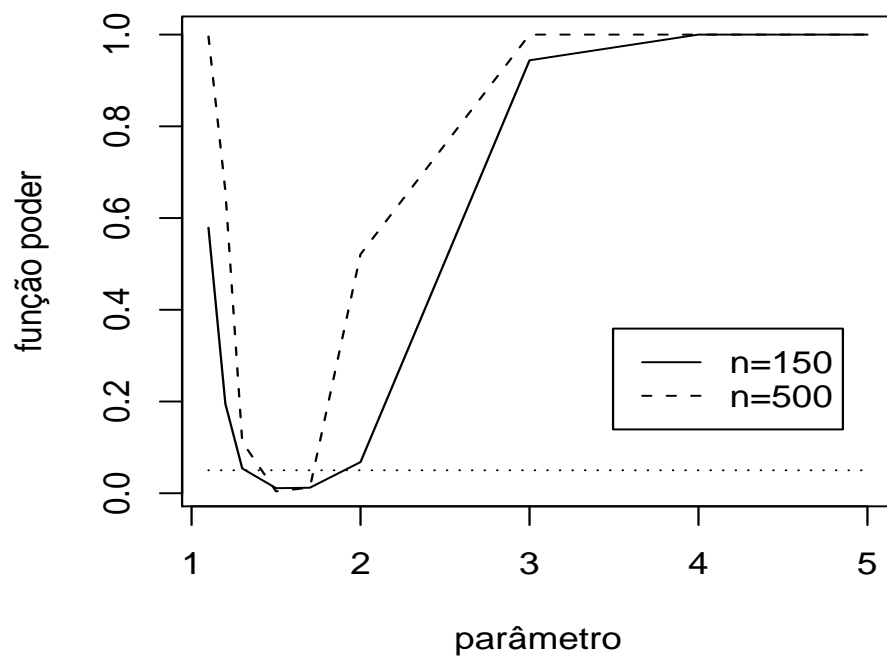


Figura 16: Função poder sob \mathbf{H}_0 : cópula de Gumbel $\theta = 1.5$ ($\tau \approx 0.33$).

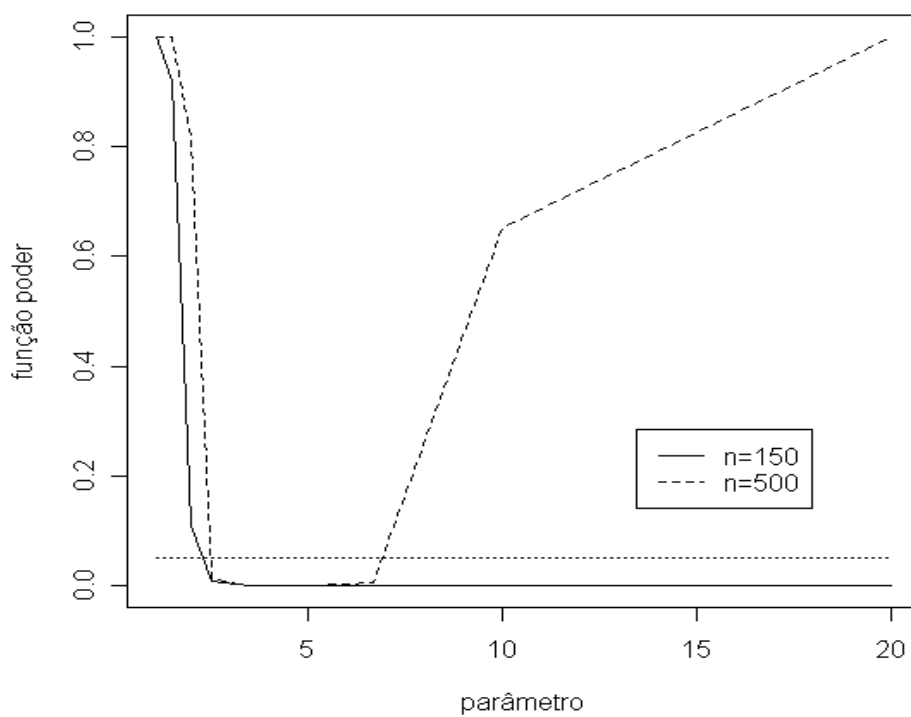


Figura 17: Função poder sob H_0 : cópula de Gumbel $\theta = 3.5$ ($\tau \approx 0.7$).

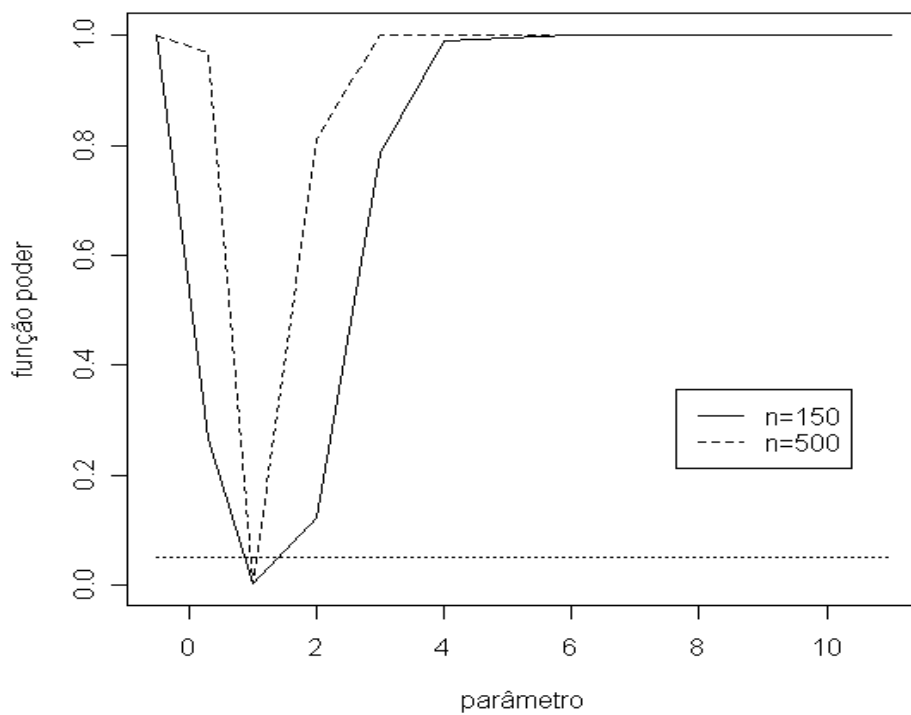


Figura 18: Função poder sob \mathbf{H}_0 : cópula de Clayton $\theta = 1$ ($\tau \approx 0.3$).

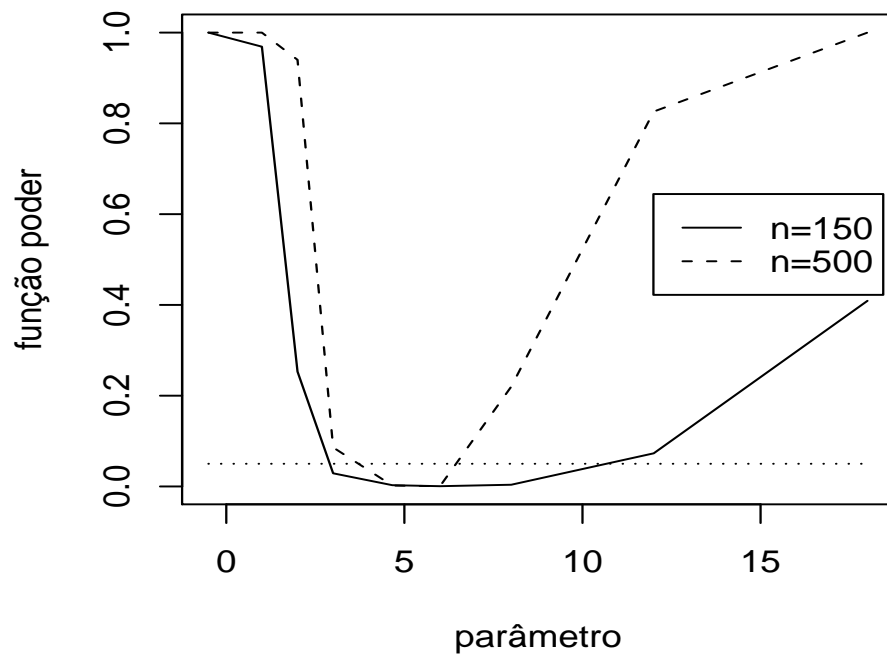


Figura 19: Função poder sob \mathbf{H}_0 : cópula de Clayton $\theta = 4.66$ ($\tau \approx 0.7$).

<i>cpula verdadeira</i>	τ	$n = 150$	$n = 500$
$\theta = -0.5$	-0.33	100	100
$\theta = 0.3$	0.1	26.5	96.9
$\theta = 1$	0.33	0.32	0.26
$\theta = 2$	0.5	12.3	81.3
$\theta = 3$	0.6	78.6	100
$\theta = 4$	0.66	99	100
$\theta = 6$	0.75	100	100
$\theta = 11$	0.85	100	100

Tabela 12: Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Clayton $\theta = 1$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Clayton com diferentes níveis de associação.

<i>cpula verdadeira</i>	τ	$n = 150$	$n = 500$
$\theta = -0.5$	-0.33	100	100
$\theta = 1$	0.33	96.9	100
$\theta = 2$	0.5	25.3	94
$\theta = 3$	0.6	2.9	8.6
$\theta = 4.66$	0.7	0.28	0.2
$\theta = 6$	0.75	0.06	0.06
$\theta = 8$	0.8	0.38	2.2
$\theta = 12$	0.85	7.3	82.6
$\theta = 18$	0.9	40.9	100

Tabela 13: Porcentagem de rejeição sob \mathbf{H}_0 : cópula de Clayton $\theta = 4.66$ de testes para 5.000 amostras de tamanho $n = 150$ e $n = 500$ seguindo cópula de Clayton com diferentes níveis de associação.

.8 Algoritmos

Abaixo são listados os algoritmos utilizados nesta dissertação escritos em linguagem *software R* para qualquer versão.

.8.1 Cálculo de $W_{i:n,K_0} \equiv E(H_{(i)}^0)$

Para integração é utilizado o comando *integrate* para quadratura adaptativa unidimensional. Os algoritmos abaixo são eficientes para tamanhos amostrais menores ou iguais a 150.

1. K_0 associada a cópula da independência:

```
#n=tamanho amostral
windep<-function (n) {
k<-function (x) {x-x*log(x)}
fdpk<-function (x) {-log(x)}
umk<-function (x) {1-k(x)}
int1<-function(x,j,m) {x*(k(x))^(j-1)*(umk(x))^(m-j)*fdpk(x)}
windp<-c(0,0)
for(i in 1:n) {
int2<-integrate(int1,0,1,j=i,m=n)$value
comb<-n*choose((n-1),(i-1))
windp[i]<-comb*int2
}
windp
}
```

2. K_0 associada a cópula de Clayton:

```
#n=tamanho amostral
#teta= parâmetro da cópula associada a K_0
winclayton<- function (n,teta) {
if (teta<(-1)) {
print ("Erro: parametro deve ser maior que -1")
break
}

kclayton<-function (x,teta) {x+(x/teta)*(1-(x^teta))}
fdpclayton<-function (x,teta) {1+(1/teta)*(1-(teta+1)*(x^teta))}
umk<-function (x,teta) {1-kclayton(x,teta)}
int1<-function(x,teta,j,m) {x*(kclayton(x,teta))^(j-1)*(umk(x,teta))^(m-j)*fdpclayton(x,teta)}
windp<-c(0,0)
for(i in 1:n) {
```

```

int2<-integrate(int1,0,1,teta=teta,j=i,m=n)$value
comb<-n*choose((n-1),(i-1))
windp[i]<-comb*int2
}
windp
}

```

3. K_0 associada a cópula de Gumbel:

```

#n=tamanho amostral
#teta= parâmetro da cópula associada a  $K_0$ 
wingumbel<- function (n,teta) {
if (teta<1) {
print ("Erro: parametro deve ser maior que 1")
break
}
kgumbel<-function (x,teta) {x-((x*log(x))/teta)}
fdpkgumbel<-function (x,teta) {1-(log(x)+1)/teta}
umk<-function (x,teta) {1-kgumbel(x,teta)}
int1<-function(x,teta,j,m) {x*(kgumbel(x,teta))^(j-1)*(umk(x,teta))^(m-j)*fdpkgumbel(x,teta)}
windp<-c(0,0)
for(i in 1:n) {
int2<-integrate(int1,0,1,teta=teta,j=i,m=n)$value
comb<-n*choose((n-1),(i-1))
windp[i]<-comb*int2
}
windp
}

```

4. K_0 associada a cópula de Frank:

```

#n=tamanho amostral
#teta= parâmetro da cópula associada a  $K_0$ 
winfrank<- function (n,teta) {
if (teta==0) {
print ("Erro: parametro deve ser diferente de 0")
break
}
kfrank<-function (x,teta) {x-((-log((exp(-teta*x)-1)/(exp(-teta)-1)))/((teta*exp(-teta*x))/(exp(-

```



```

teta*x)-1)))}
fdpkfrank<-function (x,teta) {-log((-1+exp(-teta*x))/(-1+exp(-teta)))+exp(teta*x)*(-
1+exp(-teta*x))*log((-1+exp(-teta*x))/(-1+exp(-teta)))}
umk<-function (x,teta) {1-kfrank(x,teta)}
int1<-function(x,teta,j,m) {x*(kfrank(x,teta))^(j-1)*(umk(x,teta))^(m-j)*fdpkfrank(x,teta)}
windp<-c(0,0)
for(i in 1:n) {
int2<-integrate(int1,0,1,teta=teta,j=i,m=n)$value
comb<-n*choose((n-1),(i-1))
windp[i]<-comb*int2
}
windp
}

```

5. K_0 associada a cópula 4.2.12:

```

#n=tamanho amostral
#teta= parâmetro da cópula associada a  $K_0$ 
win4212<- function (n,teta) {
if (teta<1) {
print ("Erro: parametro deve ser maior que 1")
break
}

k4212<-function (x,teta) {(x-((1/x-1)^teta/(-(teta*(-1+1/x)^(-1+teta))/x^2)))}
fdp4212<-function (x,teta) {(1-1/teta+(2*(-1+1/x1)*x1)/teta)}
umk<-function (x,teta) {1-k4212(x,teta)}
int1<-function(x,teta,j,m) {x*(k4212(x,teta))^(j-1)*(umk(x,teta))^(m-j)*fdp4212(x,teta)}
windp<-c(0,0)
for(i in 1:n) {
int2<-integrate(int1,0,1,teta=teta,j=i,m=n)$value
comb<-n*choose((n-1),(i-1))
windp[i]<-comb*int2
}
windp
}

```

Para tamanho amostral maior que 150 é aconselhável utilizar outro método de integração adaptativa para eficiência no cálculo da integral. Nesta dissertação foi utilizado

o método do trapézio aberto repetido (como apresentado no apêndice 4), abaixo segue o algoritmo em linguagem *software R* para cópulas Arquimedianas.

```

## Variáveis de entrada ##
#n=tamanho amostral
#teta=parâmetro associado a K_0
#mk=número máximo de iterações
# em cada iteração k são usados 3^(k-1) pontos de integração
#eps=precisão desejada
#kfuncao=expressão para função distribuição K_0
#fdpcopula= densidade da cópula C_0 associada a K_0

## Variáveis de saída ##
#win=vetor de estimativas para W_{i:n,K_0} i=1,2,...,n
#it=número necessário de iterações para cada estimativa de W_{i:n,K_0}
#erro=erro relativo da estimação

# Exemplo: cálculo de W_{i:n,K_0} i=1,...,500 sob hipótese nula de cópula de Clayton
com teta=4

n<-500
teta<-4
mk<-9
eps<-1e-05
kfuncao<-function (x,teta) {x+(x/teta)*(1-(x^teta))} #funcao # distribuição K para
cópula de Clayton
fdpcopula<-function (x,teta) {1+(1/teta)*(1-(teta+1)*(x^teta))} # densidade da cópu-
laa de Clayton

umk<-function (x,teta) {1-kfuncao(x,teta)}
int1<-function(x,teta,j,m) {x*(kfuncao(x,teta))^(j-1)*(umk(x,teta))^(m-j)*fdpcopula(x,teta)}
h1<-1
x<-seq(0,0,length=3^(mk-1)) #3^(mk-1) é o max de pontos para qq iteração
a33<-seq(0,0,length=3^(mk-1))
w<-c(0,0)
win<-seq(0,0,length=n)
erro<-seq(0,0,length=n)

```

```

it<-seq(0,0,length=n)

maxl2<-floor((3^(mk-1)-1)/3+1)
maxz<-3^(mk-1)

for(j in 1:n){
l2<-seq(-1,-1,length=maxl2)
z<-seq(1,1,length=maxz)
x<-seq(0,0,length=maxz)
x1<-0.5
a3<-int1(x1,teta,j,n)
w[1]<-n*choose((n-1),(j-1))*a3

soma<-a3
i<-1
while(i<mk & win[j]==0){
i<-i+1
n2<-3^(i-1)
h<-h1/n2
for(k in 1:n2){
x[k]<-(k-0.5)*h
z[n2-(k-1)]<-x[k]
}

n3<-floor((n2-1)/3+1)
t<-0
for(q in 1:n3){
l2[q]<-(n2-1)-(q-1)*3 if(l2[q]>0){
z<-z[-l2[q]]
}
if(l2[q]<=0){
t<-t+1
}
}
}

```

```

    if(t!=0){
n3<-n3-1
}

    lengtz<-n2-n3

    for(k in 1:lengtz){
a33[k]<-int1(z[k],teta,j,n)
}

    a4<-sum(a33[1:lengtz])
soma<-soma+a4
w[i]<-n*choose((n-1),(j-1))*soma*h

    if(abs(w[i]-w[i-1])<eps*abs(w[i-1])){
win[j]<-w[i]
erro[j]<-abs(w[i]-w[i-1])/abs(w[i-1])
it[j]<-i
}

    if(i==mk & w[i]!=0 & abs(w[i]-w[i-1])>=eps*abs(w[i-1])){
win[j]<-w[i]
print("Erro:nž passos insuficientes para w");print(j)
it[j]<-i
}
if(i==mk & w[i]==0){
win[j]<-1e-266 # só pra parar o while se houve indícios de win=0
it[j]<-i
}
}
}
}

```

.8.2 Teste gráfico Kendall Plot com bandas de confiança

```
## Variáveis de Entrada ##
#x= vetor de dados bidimensionais
#teta= parâmetro da copula Arquimediana sob hipótese nula
#win= vetor dos valores de  $W_{i:n}, K_{0 i=1,2,\dots,n}$ .
#alp= nível de significância do Kendall Plot (entre 0 e 1)
#quant= quantidade de quantis testados: máximo  $(n - 2)$ 
#quantil= função para expressão dos  $p_j$ -quantis
#f= função densidade  $k_0$  referente a função distribuição  $K_0$ 
#da copula Arquimediana sob hipótese nula.

## Variáveis de Saída ##
#Gráfico Kendall Plot com bandas de confiança
# h= bipit empírica ordenada:  $H_{(1)}, \dots, H_{(n)}$ 
# u= vetor dos extremos superiores da banda de confiança do Kendall Plot
#l= vetor dos extremos inferiores da banda de confiança do Kendall Plot

#Exemplo de Kendall Plot para 150 dados exponenciais bivariados sob hipótese nula
de copula de Clayton com teta=2 utilizando 20 testes conjuntos

x<-matrix(nrow=150,ncol=2)
x[,1]<-rexp(150,2)
x[,2]<-rexp(150,10)
teta<-2
alp<-0.05
quant<-20
win<-winclayton(150,2)
f<-function(x,teta){1+(1/teta)*(1-(teta+1)*(x^teta))} #densidade  $k_0$  sob hipótese nula
de copula de Clayton
quantil<-function(k){0.05*k-0.025} #expressão para  $p_j$ -quantil

hi<-c(0,0)
l<-c(0,0); u<-c(0,0); w<-c(0,0); h<-c(0,0)
```

```

if (length(x[,1])!=length(x[,2])) {
print ("Erro: os vetores devem ter o mesmo comprimento")
break
}
n<-length(x[,1])
for (i in 1:n) {
soma<-0
for (j in 1:n) {
if (j!=i & x[j,1]<=x[i,1] & x[j,2]<=x[j,2]) {
soma<-soma+1
}
}
hi[i]<-soma/(n-1)
}
hh<-hi[order(hi)]

```

```

z<-qnorm(1-alp/(quant*2))
if(quant==(n-2)){
p1<-2
pn<-n-1
}
p1<-1
pn<-quant
for(k in p1:pn){
pi<-quantil(k)
i<-floor(n*pi)+1
w[k]<-win[i] h[k]<-hh[i] dens<-f(win[i],teta)
s<-sqrt(pi*(1-pi)/dens^2)
l[k]<-h[i]-((z*s)/sqrt(n))
u[k]<-h[i]+((z*s)/sqrt(n))
}
u<-u[p1:pn]
l<-l[p1:pn]
if(pn==(n-2)){
w<-w[-n]
w<-w[-1]
h<-h[-n]
h<-h[-1]
}

```

```
}  
plot(w,h,main="Kendall Plot",xlab="W_np:n,K_0",ylab="H_(np)")  
a<-seq(0,1,length=100)  
lines(a,a)  
lines(u,h,lty=2)  
lines(l,h,lty=2)
```

Referências Bibliográficas

- [1] Abramowitz, M. & Stegun, I. E. (1972). Handbook of Mathematical Functions. New York: Dover.
- [2] Bouyé E., Durrleman V., Nikeghbali A., Riboulet G., Roncalli T. (2000). Copulas for Finance – A Reading Guide and Some Applications. Working paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.
- [3] Capéreaux, P., Fougères, A. L. & Genest, C. (1997). A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas. *Biometrika*, 84, 567:577.
- [4] Capéreaux, P., Fougères, A. L. & Genest, C. (2000). Bivariate Distributions With Given Extreme Value Attractor. *Journal of Multivariate Analysis*, 72, 30:49.
- [5] Casella, G. & Berger, L. B. (2002). *Statistical Inference*. Duxbury, 2nd. ed.
- [6] Cherubini U., Luciano E., Vecchiato W. (2004). *Copula Methods in Finance*. John Wiley & Sons.
- [7] Dobrić, J. & Schmid, F. (2007). A goodness of fit test for copulas based on Rosenblatt's transformation. *Computational Statistics & Data Analysis* 51, pp. 4633:4642.
- [8] Durrleman, V., Nikeghbali, A. & Roncali, T. (2000). Which copula is the right one? Groupe de Recherche Opérationnelle, Crédit Lyonnais.
- [9] Embrechts, P., McNeil, A. & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond*, ed. M. A. H. Dempster, Cambridge University Press, Cambridge, pp. 176:223.
- [10] Embrechts P., Lindskog F., McNeil A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In S Rachev (ed.), *Handbook of Heavy Tailed Distribution in Finance*, pp. 329:384. Elsevier.

- [11] Frees E.W., Carriere J., Valdez E.A. (1996). Annuity Valuation with Dependent Mortality. *Journal of Risk and Insurance*, 63, 229:261.
- [12] Frees E.W. & Valdez E.A. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, 2(1), 1:25.
- [13] Frees E.W., Wang P. (2005). Credibility Using Copulas. *North American Actuarial Journal*, 9(2), 31:48.
- [14] Garralda-Guillem, A. I. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées. *Comptes rendus de l'Académie des sciences de Paris, Série I*, 330, 593:596.
- [15] Genest, C. & MacKay, J. (1986a). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.*, 14, 145:159.
- [16] Genest, C. & MacKay, J. (1986b). The joy of copulas: Bivariate distributions with uniform marginals. *Amer. Statist.*, 40, 280:285.
- [17] Genest, C. & Rivest, L. P.(1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034:1043.
- [18] Genest, C. & Rivest, L.P. (2001). On the multivariate probability integral transformation. *Statistics and Probability Letters*, 53, 391:399.
- [19] Genest, C. & Boies, J.C. (2003). Detecting dependence with Kendall plots. *The American Statistician*, 57 (4), 275:284.
- [20] Genest, C., Quessy, J. F. & Rémillard, B. (2006). Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation. *Scandinavian Journal of Statistics*, 33 (2), 337:366.
- [21] Genest, C. & Frave, A. C.(2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12, 347:368.
- [22] Genest, C., Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 44 (in press).
- [23] Genest, C., B. Rémillard & Beaudoin, D. (2008). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 42 (in press).
- [24] Ghoudi, K., Khoudraji, A., Rivest, L.P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *The Canadian Journal of Statistics*, 26, 187:197.

- [25] Gonzalez-Lopez, V.A. & Tanaka, N.I. (2000). Maximal Association, Fréchet Bounds and Homotopy in Prediction. Technical report, Department of Statistics, University of São Paulo.
- [26] Gray, R. (2001). Advanced Statistical Computing Course Notes.
- [27] James, Barry R. (2004). Probabilidade: um curso em nível intermediário. Rio de Janeiro, 3.ed.: Impa.
- [28] Joe H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall Ltd.
- [29] Jun, Y. (2007). Enjoy the Joy of Copulas: With a Package copula. Journal of Statistical Software, October 2007, Volume 21, Issue 4.
- [30] Nelsen, R.B. (1999). An Introduction to Copulas. Lecture Notes in Statistics 139, Springer-Verlag, New York.
- [31] Nelsen, R.B. (1999). Dependence modeling with Archimedean copulas. Proceedings of the Second Brazilian Conference on Statistical Modelling in Insurance and Finance, (N. Kolev and P. Morettin, editors), Institute of Mathematics and Statistics, University of São Paulo (2005), 45:54.
- [32] Nelsen, R., Quesada-Molina, J., Rodrigues-Lallena, J. & Ubeda-Flores, M. (2003). Kendall distribution functions. Statistics & Probability Letters, 65, 263:268.
- [33] Piessens, R., Doncker-Kapenga, E., Überhuber, C.W. e Kahaner, D.K. (1983). Quadpack: a subroutine package for automatic integration. New York: Springer-Verlag.
- [34] Rosenblatt, M. (1952). Remarks on a multivariate transformation. Ann. Math. Statist., 23, 470:472.
- [35] Ruggiero, M.A.G & Lopes, V.L.R. (1996). Cálculo numérico: aspectos teóricos e computacionais. São Paulo: Pearson Makron Books.
- [36] Schweizer, B. & Sklar, A. (1983). Probabilistic Metric Spaces. North-Holland, New York.
- [37] Sen, P. K. & Singer, J. M. (1993). Large Sample Methods in Statistics: an introduction with applications. New York: Chapman and Hall.
- [38] Sklar, A.W. (1959). Fonctions de répartition à n - dimension et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, 8, 229:231.

- [39] Walker, A. M. (1968). A note on the asymptotic distribution of sample quantiles. *J. Roy. Statist. Soc. Ser. B*, 30 (3), 570:575.
- [40] Wang, W. & Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.*, 95, 62:72.
- [41] Wilk, M. B. & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55, 1-17.