



UFRJ

Universidade Federal do Rio de Janeiro

# Cópuas para distribuições generalizadas de valores extremos multidimensionais

Marco Aurélio dos Santos Sanfins

2009

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



UFRJ

Marco Aurélio dos Santos Sanfins

## CÓPULAS PARA DISTRIBUIÇÕES GENERALIZADAS DE VALORES EXTREMOS MULTIDIMENSIONAIS

Dissertação de Doutorado apresentada ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências Estatística.

Orientador:  
Glaucio Valle

Rio de Janeiro

Janeiro de 2009

# COPULAS PARA DISTRIBUIÇÕES GENERALIZADAS DE VALORES EXTREMOS MULTIDIMENSIONAIS

Marco Aurélio dos Santos Sanfins

Orientador: Glauco Valle

Tese de Doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

---

Presidente, Prof. Glauco Valle

IM-UFRJ

---

Prof. Luiz Koodi Hotta

IMECC-UNICAMP

---

Prof. Hélio dos Santos Migon

IM-UFRJ

---

Prof. Vladimir Belintsky

IME-USP

---

Prof. Nei Rocha

IM-UFRJ

Rio de Janeiro, janeiro de 2009.

Sanfins, Marco Aurélio dos Santos

Copulas para distribuições generalizadas de valores extremos multidimensionais/Marco Aurélio dos Santos Sanfins - Rio de Janeiro: UFRJ/IM, 2009.

xiii, 95f.: il.; 31cm.

Orientador: Glauco Valle

Tese (doutorado) - UFRJ/IM/ Programa de Pós-graduação em Estatística, 2009.

Referências Bibliográficas: f.80-82.

1. Cópulas. 2. Estatísticas de ordem. 3. Variáveis aleatórias independentes. I. Valle, Glauco II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. III. Título.

# Resumo

Nesta tese foi obtida a cópula associada a uma distribuição generalizada de valores extremos multidimensionais, chamada de cópula  $K$ -extremal. Essa é descrita através de fórmulas exatas para as suas funções de distribuição e densidade. O caso  $K = 2$  é feito primeiro, servindo de partida para os procedimentos indutivos que nos permitiram estender os resultados para  $K > 2$ . Além disso, propriedades da cópula Bi-extremal foram obtidas, entre elas o  $\tau$  de Kendall,  $\rho$  de Spearman e os coeficientes de cauda. Exemplos teóricos e simulados foram propostos. Outro resultado relevante foi provar que a cópula das  $K$ -maiores estatísticas de ordem de uma seqüência (*i.i.d.*) converge em distribuição para a cópula  $K$ -extremal. Estudo de medidas de dependência também é feito usando a cópula  $K$ -extremal e como último resultado, foi proposto um algoritmo de simulação para esta cópula.

*Palavras-Chave:* Cópulas, estatísticas de ordem, variáveis aleatórias independentes e valores extremos.

# Abstract

In this thesis, we obtain the copula related to multivariate generalized extreme values distributions, which is called the  $K$ -extremal copula. It is described by its distribution and density functions through exact expressions. The case  $K = 2$  is done first, serving as a departure for the inductive procedures that allowed us to extend the results to  $K > 2$ . In addition, properties of Bi-extremal copula were obtained, including the Kendall's  $\tau$  coefficient, Spearman's  $\rho$  coefficient and the tail dependence coefficients. Theoretical and simulated examples have been worked out. We also show that the copula of the  $K$  largest order statistics of (*i.i.d.*) sequences converges in distribution to the  $K$ -extremal copula. We also study measures of dependence. As a last result, we propose a simulation algorithm to sample from the  $K$ -extremal copula.

*Key-words:* Cópulas, order statistics, independent random variable and extreme values.

Porque sem mim nada podereis fazer

*Jesus Cristo*



# Agradecimentos

À minha querida esposa Alcina que em todos os momentos acreditou em mim.

Ao meu filho Gabriel com quem aprendi a ser mais concentrado.

Ao meu filho Lucas com quem aprendi a ser mais carinhoso.

À minha filha Clara que me ensinou ainda mais a alegria de viver.

Ao professor Glauco Valle, por toda sua dedicação, auxílio e compreensão como meu orientador. Tenho certeza de que seu auxílio foi fundamental para a conclusão deste projeto.

Ao professor Dani Gamerman que acreditou em minha capacidade; com certeza jamais esquecerei o que ele possibilitou em minha vida.

À Professora Beatriz Vaz de Melo Mendes, pela colaboração nos primeiros resultados obtidos nesta tese, os quais originaram meu primeiro artigo.

Aos professores Kaizô Iwakami Beltrão e Hélio dos Santos Migon, por terem me recomendado para este curso.

À minha amiga Ursula Faustino Nesci pela revisão deste texto por ocasião da qualificação.

Aos meus pais, estes nunca poderiam imaginar que o fruto da dedicação ao seu filho fosse proporcionar este momento. E à minha querida sogra, a quem dedico parte deste trabalho.

À FAPES - Fundação de Assistência e Pevidência Social do BNDES, pela inclusão no programa de pós-graduação desta instituição.

Agradeço, ainda, aos demais professores do Instituto pela oportunidade de ampliar meus conhecimentos e minha formação.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
	Introdução	1
<b>2</b>	<b>Preliminares</b>	<b>4</b>
2.1	Cópuas	4
2.2	Teoria dos Valores Extremos	7
2.2.1	Modelos Assintóticos	8
2.2.2	O modelo para as $K$ -maiores estatísticas de ordem	12
<b>3</b>	<b>A cópula <i>Bi</i>-extremal</b>	<b>18</b>
3.1	Especificação	18
3.2	Algoritmo para Simulação	24
3.3	Exemplos teóricos	26
3.3.1	O caso Uniforme(0,1)	26
3.3.2	O caso Exponencial(1)	27
3.4	Exemplos utilizando seleções de processos ( <i>i.i.d.</i> )	28
3.5	Verificação empírica para a <i>Bi</i> -extremal	32
<b>4</b>	<b>Propriedades da Cópula <i>Bi</i>-extremal</b>	<b>36</b>
4.1	Medidas de Dependência	36
4.1.1	O coeficiente de cauda	36
4.1.2	<i>Rho</i> de Spearman	38
4.1.3	Tau de Kendall	41
4.1.4	O coeficiente de Schweizer e Wolff	43
4.1.5	Representação contínua da dependência	45
4.2	Cópula Condicional	47

<b>5</b>	<b>A cópula K-extremal</b>	<b>51</b>
5.1	Preliminares . . . . .	51
5.2	Especificação . . . . .	56
5.3	Convergência em Distribuição . . . . .	59
5.4	Algoritmo para Simulação . . . . .	64
5.5	Simulações de processos (i.i.d.) . . . . .	66
5.6	Verificação empírica para a K-extremal . . . . .	68
5.7	Processos Estacionários . . . . .	68
5.8	Simulação de processos ARMA . . . . .	72
<b>6</b>	<b>Conclusão</b>	<b>77</b>
	<b>Bibliografia</b>	<b>79</b>
<b>A</b>	<b>Gráficos das simulações</b>	<b>82</b>

# Lista de Figuras

2.1	<i>Gráfico da função de distribuição para o caso em que as variáveis são independentes, <math>C(u, v) = uv</math>.</i>	7
2.2	<i>Função de distribuição e função de densidade limite das 5 maiores estatísticas de ordem. À esquerda temos as funções de distribuição e à direita, as funções de densidade para <math>\mu = 0</math>, <math>\sigma = 1</math> e <math>\xi = 0.2</math>.</i>	13
3.1	<i>Perspectiva da função de densidade da cópula Bi-Extremal</i>	19
3.2	<i>Gráfico da cópula Bi-Extremal.</i>	24
3.3	<i>Gráficos da cópula BI-extremal. Na parte superior, os gráficos ilustram as curvas de contorno da cópula Bi-extremal <math>C_{BIX}</math> (esquerda) e da função de densidade <math>c_{BIX}</math>. Na parte inferior, podemos ver, no gráfico da esquerda, as seções da diagonal principal das cópulas <math>C_{BIX}</math>, <math>M</math> (cópula de dependência perfeita) e <math>I</math> (produto); e no lado direito as seções diagonais das densidades correspondentes.</i>	25
3.4	<i>Histogramas do processo de simulação. Na parte superior da figura: à esquerda o histograma de 90.000 observações da distribuição <math>N(0,1)</math>, à direita o histograma dos máximos observados para os blocos de tamanho <math>n = 30</math>. Na parte inferior: à esquerda o histograma das segundas maiores observações e à direita os dados pseudo-uniforme(0,1)</i>	29
3.5	<i>Comparativo entre dados reais e simulados (<math>N = 3000; n = 30</math>). À esquerda, de forma ilustrativa, temos o valor esperado arredondado, para cada célula calculada a partir da <math>C_{BIX}</math>, já à direita, o valor observado para uma certa realização das 200 repetições do processo de simulação.</i>	30
3.6	<i>Histograma dos p-valores obtidos. P-valores oriundos das 200 repetições do processo gerados a partir da <math>N(0,1)</math>, com tamanho de amostra <math>n = 100</math>.</i>	31
3.7	<i>Evolução histórica do índice IBOVESPA e das ações da PETROBRAS.</i>	32
3.8	<i>Evolução histórica do <math>\log(\text{RETORNOS})</math> do IBOVESPA e das ações da PETROBRAS.</i>	33
3.9	<i>Histograma e respectiva função de densidade para os valores observados do máximo e segundo maior, das séries IBOVESPA e PETROBRAS.</i>	34
3.10	<i>Dados Uniformizados e Simulados.</i>	35

4.1	<i>Perspectiva da função de Spearman para cópula Bi-Extremal</i>	45
4.2	<i>Curvas de Nível da função de Spearman para cópula Bi-Extremal</i>	46
5.1	<i>Gráfico ilustrativo da convergência de <math>\psi_K^{-1}(u)</math> quando <math>K \rightarrow \infty</math></i>	58
5.2	<i>Gráficos com cruzamentos entre dados simulados da cópula K-extremal, com <math>K = 4</math></i>	66
A.1	<i>Histograma das simulações de processos (i.i.d) da cópula K-extremal com <math>K = 4</math>, para Norma(0,1) e Normal(0,4)</i>	83
A.2	<i>Histograma das simulações de processos (i.i.d) da cópula K-extremal, com <math>K = 4</math>, para a T-student(3) e T-student(4)</i>	84
A.3	<i>Histograma das simulações de processos (i.i.d) da cópula K-extremal com <math>K = 4</math>, para Gama(1,3) e Uniforme(0,1)</i>	85
A.4	<i>Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para <math>K = 4</math>.</i>	86
A.5	<i>Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para <math>K = 4</math>.</i>	87
A.6	<i>Histograma das simulações de processos fracamente estacionários MA(1), da cópula K-extremal, para <math>K = 4</math>.</i>	88
A.7	<i>Histograma das simulações de processos fracamente estacionários MA(1), da cópula K-extremal, para <math>K = 4</math>.</i>	89
A.8	<i>Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para <math>K = 4</math>.</i>	90
A.9	<i>Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para <math>K = 4</math>.</i>	91
A.10	<i>Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para <math>K = 4</math>.</i>	92
A.11	<i>Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para <math>K = 4</math> com erros <math>U(-1,1)</math>.</i>	93
A.12	<i>Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para <math>K = 4</math> com erros <math>U(-1,1)</math>.</i>	94

# Lista de Tabelas

3.1	Média e erro padrão dos p-valores obtidos nas simulações . . . . .	31
5.1	Tabela com dados simulados para a cópula 4-extremal - processos i.i.d. . . . .	67
5.2	Tabela resumo com dados simulados para a cópula 4-extremal - AR(1) . . . . .	73
5.3	Tabela resumo com dados simulados para a cópula 4-extremal - MA(1) . . . . .	74
5.4	Tabela resumo com dados simulados para a cópula 4-extremal - ARMA(1) . . . . .	76
5.5	Tabela resumo com dados simulados para a cópula 4-extremal - AR(1) . . . . .	76

# Capítulo 1

## Introdução

Freqüentemente, a mídia tem divulgado a ocorrência de eventos extremos que, em sua maioria, estão relacionados com catástrofes naturais. Diante desse fato, grandes pesquisadores vêm discutindo temas ligados a catástrofes, cujos argumentos são baseados nas grandes alterações dos níveis de ozônio, da radiação ultra violeta, dos efeitos causados pelo *El Niño*, nas conseqüências de uma guerra global e, finalmente, nos efeitos da globalização.

Exemplos sobre as conseqüências dessas catástrofes foram observados em inumeráveis situações ao longo dos anos. Entre estes podemos citar: os altos níveis do mar em Nova Orleans no ano de 2005, a ocorrência de *tsunamis* no oceano Índico em 2004, o *crash* do mercado de ações em 2002. A ocorrência desses eventos, em sua maioria, é considerada como sendo extrema, ou seja, de pequena probabilidade de ocorrência. Contudo, quando levamos em conta dados históricos, verifica-se, nesses casos, que também a segunda maior observação, ou ainda, generalizando para as  $K$ -maiores observações, que estas também são eventos extremos. Diante das conseqüências que esses eventos podem causar, tanto materialmente como financeiramente, torna-se indispensável conhecer profundamente a relação entre os  $K$ -maiores eventos. O trabalho aqui desenvolvido consiste exatamente em estudar o comportamento da dependência entre os  $K$ -maiores eventos, mais especificamente, capturar a relação de dependência não-linear entre as variáveis, utilizando o conceito de cópulas. Algumas suposições serão feitas com o propósito de formalizar e evidenciar melhor sobre que tipos de dados os resultados encontrados nos capítulos seguintes se aplicam. Inicialmente, supôs-se que essas realizações são oriundas de uma

seqüência  $\{X_t\}_{t \geq 1}$  de variáveis independentes e identicamente distribuídas (*i.i.d.*). Com a viabilidade da modelagem das  $K$ -maiores estatísticas de ordem, foi possível o cálculo da probabilidade de ocorrência de eventos extremos com uma maior precisão. Entretanto quando não é possível obter os parâmetros da GEV limite, a cópula para a distribuição generalizada de valores extremos  $K$ -dimensional nos fornece uma total identificação da dependência não linear entre essas variáveis.

A necessidade crescente de conhecer bem o comportamento de eventos extremos foi a principal motivação para o desenvolvimento e a elaboração de todos os resultados encontrados nesta tese. Inicialmente, o primeiro esforço foi estudar a estrutura de dependência para as duas maiores estatísticas de ordem, ou seja, entre o máximo e a segunda maior ( $K = 2$ ). O principal objetivo foi encontrar a cópula para esse par de estatísticas de ordem. O primeiro resultado obtido foi a expressão para a cópula  $C_{BIX}$ , a cópula Bi-extremal, pertinente à distribuição assintótica das duas maiores estatísticas de ordem de um processo (*i.i.d.*). A partir daí foi possível derivar várias medidas de dependência relacionadas a estas duas estatísticas de ordem. Entre estas estão: as expressões do tau de Kendall e do rho de Sperman, bem como os coeficientes de cauda. Baseado em um resultado de extrema importância introduzido por Charpentier (2004), que é a **LTDC** (*the lower tail dependence copula*), que caracteriza a estrutura de dependência no limite da cauda bivariada, foi possível o desenvolvimento algébrico da LTDC para a cópula Bi-extremal. Além disso, foram elaborados processos de simulações, exemplos para a cópula Bi-extremal e um algoritmo para simulação foi proposto.

O interesse pelo estudo das distribuições multivariadas vem crescendo a cada ano, grande parte em função do interesse em obter formas de medir a dependência entre variáveis aleatórias, que não sejam as já existentes na literatura. Entretanto, o estudo destas distribuições multivariadas através do conceito de cópulas torna-se bem mais prático. Diante destes argumentos e após a obtenção da cópula Bi-extremal, passou a ser de extrema importância obter um resultado genérico que contemplasse o caso **K-variado**. No capítulo 5, as expressões exatas da função de densidade e função de distribuição para a cópula da distribuição generalizada de valores extremos  $K$ -dimensional são obtidas, e chamada de cópula **K-extremal**. Também será provado que a cópula das



$K$ -maiores estatísticas de ordem de uma seqüência (*i.i.d*), sob certas condições, converge em distribuição para a cópula  $K$ -extremal. Um algoritmo de simulação para a cópula  $K$ -extremal é proposto. Além disso, foi obtido resultado de convergência exponencial. Um estudo sobre a convergência assintótica do rho de Spearman e do tau de Kendall, entre a primeira e a  $K$ -ésima maior estatística de ordem, também foi feito.

Diante desses resultados, constata-se que inúmeras aplicações poderão utilizar esta nova estrutura de dependência contida na cópula  $K$ -extremal, e, além disso, surge a possibilidade de se obter outros resultados relevantes oriundos desta.

Esta tese se apresenta dividida da seguinte forma: no capítulo 2, é feita uma introdução sobre cópulas e a teoria de valores extremos (TVE) cujos conceitos e definições, são o alicerce da grande maioria dos resultados obtidos posteriormente. No capítulo 3, foi inicialmente estudado o caso particular para  $K = 2$ , ou seja, foi obtida a cópula bi-variada para a distribuição generalizada de valores extremos, chamada de cópula **Bi-extremal**, esta foi especificada completamente através de sua função de densidade e distribuição; além disso, um algoritmo para simulação foi proposto, exemplos teóricos e simulados foram feitos e uma aplicação a dados financeiros para a cópula Bi-extremal foi proposta. No capítulo 4, foi feito um estudo detalhado sobre propriedades e medidas de dependência para cópula Bi-extremal. Por fim no capítulo 5 foi obtida a cópula para o caso  $K$ -dimensional, ou seja, a cópula associada à distribuição generalizada de valores extremos  $K$ -dimensional chamada de cópula **K-extremal**. Por fim no último capítulo apresentamos uma revisão geral das contribuições e trabalhos futuros.

# Capítulo 2

## Preliminares

### 2.1 Cópulas

Até há algum tempo, a modelagem estatística em finanças se apoiou basicamente em suposições simplificadoras. Por exemplo, séries de retornos de ativos eram assumidas como sendo oriundas de uma distribuição normal. Além disso, no aspecto multivariado, a distribuição normal multivariada era freqüentemente assumida e quase nunca contestada. Contudo, a suposição multivariada restringe o tipo de associação entre as distribuições marginais (normais) a ser linear. Essa é uma restrição drástica, e não parece razoável ser a associação linear o único tipo de dependência que poderia ser observada entre séries financeiras. Situações similares ocorrem em outras áreas, tais como em seguros, onde os dados tipicamente possuem caudas pesadas, e onde, muitas vezes, as conclusões são obtidas através do Teorema Central do Limite.

A suposição inadequada de normalidade para a modelagem multivariada desses fenômenos financeiros acarreta graves conseqüências. A principal é com relação ao coeficiente de correlação entre estas variáveis, que pouco tem a informar por conta desta suposição.

É bem conhecido que a modelagem inadequada de fenômenos financeiros pode causar grandes prejuízos, ou, em alguns casos, levar grandes instituições do mercado financeiro a um *crash*. Logo, é relevante promover um maior entendimento com relação a essas estruturas e buscar soluções adequadas para modelar esses dados, tornando-se um grande desafio para as instituições que estão diretamente convivendo com esse tipo de risco.

Com o avanço tecnológico que vem ocorrendo nos últimos anos, cada vez mais temos computadores mais velozes e a um custo cada vez menor; isso possibilitou o desenvolvimento de produtos financeiros mais complexos que permitem modelar dados multivariados com estruturas de dependência complexas.

É desta forma que o conceito de cópulas pode ser eficientemente usado para a modelagem de dados multivariados. Isso se dá devido ao número incontável de combinações possíveis de serem feitas entre os tipos de cópulas e de distribuições marginais que podemos assumir. O conceito de cópulas é relativamente fácil de ser usado para construir distribuições multivariadas oriundas, em sua maioria, na escolha de quaisquer marginais e qualquer tipo de estrutura de dependência. Uma outra motivação para se considerar o uso de cópulas para modelagem de dados financeiros é que freqüentemente a única medida de dependência usada para explicar a associação entre as variáveis é o coeficiente de correlação, o qual com a utilização do conceito de cópulas, permite conhecermos, após uma modelagem adequada, como é a estrutura de dependência não linear entre estes dados multivariados.

Nesta seção vamos definir cópulas, bem como suas propriedades. Resumindo, a função cópulas pode ser definida como sendo uma função de ligação entre as distribuições marginais univariadas que tem como resultado distribuições multivariadas. Em Nelsen (2006), cópulas podem ser definidas por dois pontos de vista: o primeiro resume cópulas como sendo funções que juntam ou acoplam distribuições multivariadas às suas distribuições marginais. A outra forma de interpretação afirma que cópulas são distribuições multivariadas cujas distribuições marginais são uniformes no intervalo  $(0,1)$ .

Todo o conceito anteriormente visto sobre cópulas é oriundo do teorema de Sklar(1959), que para o caso  $n$ -dimensional iremos agora enunciar. Inicialmente vamos definir a função:

**Definição 2.1** *Seja  $I = [0, 1]$  e  $I^k = [0, 1]^k$ , uma Cópula  $k$ -dimensional é uma função de distribuição multidimensional  $C$  definida em  $I^k$  que assume valores em  $I$ , e que possui as seguintes propriedades:*

(i) *Para todo  $u_i \in I$   $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_k) = 0$ , com  $i = 1, \dots, k$ ;*

(ii) Para todo  $u_i \in I$   $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ , com  $i = 1, \dots, k$ ;

**Teorema 2.1** Sklar (1959) Seja  $(X_1, X_2, \dots, X_k)$  um vetor aleatório em  $\mathbb{R}^k$  com função de distribuição conjunta  $F$  e marginais  $F_i, i = 1, 2, \dots, k$ . Então existe uma cópula  $C$  tal que  $\forall \mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$ .

$$F(\mathbf{x}) = C(F_1(x_1), \dots, F_k(x_k)). \quad (2.1)$$

Se  $F_1, \dots, F_k$  são contínuas, então  $C$  é única, e determinada pela fórmula

$$C(u_1, \dots, u_k) = F(F_1^{-1}(u_1), \dots, F_k^{-1}(u_k)). \quad (2.2)$$

□

Além disso, assumindo que  $F_1, \dots, F_k$  e  $C$  são diferenciáveis, então a densidade conjunta  $f(x_1, \dots, x_k)$  de  $(X_1, \dots, X_k)$  pode ser escrita como

$$f(x_1, \dots, x_k) = c(F_1(x_1), \dots, F_k(x_k)) \cdot \prod_{i=1}^k f_i(x_i), \quad (2.3)$$

onde

$$c(u_1, \dots, u_k) = \frac{\partial^k C(u_1, \dots, u_k)}{\partial u_1 \dots \partial u_k},$$

é a densidade da cópula, e  $f_i$  é a função de densidade de  $X_i, i = 1, \dots, k$ . Em (2.3) notamos a decomposição da densidade conjunta em duas partes. Onde  $c(F_1(x_1), \dots, F_k(x_k))$  descreve a estrutura de dependência e  $\prod_{i=1}^k f_i(x_i)$  descreve o comportamento marginal de cada componente. A figura (2.1) ilustra a perspectiva da cópula para o caso em que as variáveis são independentes.

Cópuas são conhecidas como funções de dependência. Estas também podem ser caracterizadas como função de distribuição acumulada multivariada cujas marginais têm distribuição  $U(0, 1)$ . Entre as cópuas podemos citar algumas importantes:

1.  $C^-(u_1, \dots, u_k) = \max(u_1 + \dots + u_k - k + 1, 0)$ , Para  $k = 2$ ,  $C^-$  representa a dependência negativa perfeita entre as variáveis, não sendo uma cópula para  $k \geq 3$ ;

2.  $C^\perp(u_1, \dots, u_k) = \prod_{i=1}^k u_i$ , é a cópula produto e representa independência entre as variáveis;
3.  $C^+(u_1, \dots, u_k) = \min(u_1, \dots, u_k)$ , e representa a dependência positiva perfeita entre as variáveis.

Uma consequência do teorema de Sklar são os limites de Fréchet-Hoeffding, nos quais se constata, para qualquer cópula  $C(u_1, \dots, u_k)$ ,

$$C^-(u_1, \dots, u_k) \prec C(u_1, \dots, u_k) \prec C^+(u_1, \dots, u_k)$$

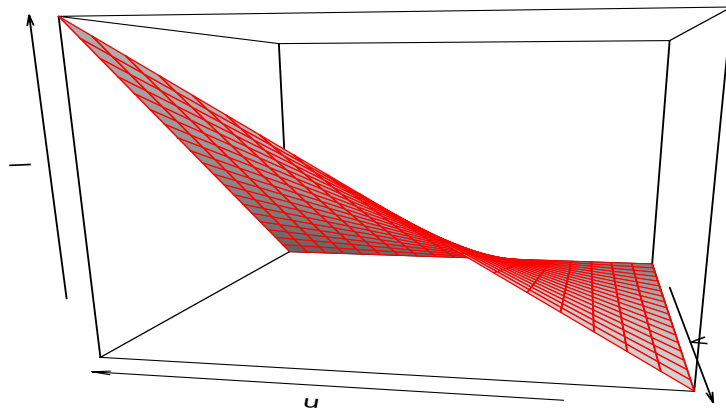


Figura 2.1: Gráfico da função de distribuição para o caso em que as variáveis são independentes,  $C(u, v) = uv$ .

## 2.2 Teoria dos Valores Extremos

A teoria de Valores Extremos (TVE) é um ramo da probabilidade que estuda o comportamento assintótico de extremos associados a uma seqüência de variáveis aleatórias. Nesta seção faremos uma breve introdução sobre os resultados da teoria de valores extremos para variáveis aleatórias (*i.i.d.*) que serão relevantes ao nosso estudo.

## 2.2.1 Modelos Assintóticos

Sejam  $X_i, i = 1, 2, \dots, n$ , variáveis aleatórias (*i.i.d.*), com função de densidade comum  $f$  e função de distribuição  $F$ , e representemos como  $M_{k,n}$  a  $k$ -ésima maior estatística de ordem de  $\{X_1, \dots, X_n\}$ . A densidade  $f_{k,n}$  e a função de distribuição de  $F_{k,n}$  de  $M_{k,n}$  podem ser encontradas em Gumbel (1958) e Embrechts, Klüppelberg e Mikosch (1997), e são dadas por

$$f_{k,n}(z) = \frac{n!}{(n-k)!(k-1)!} [F(z)]^{n-k} [1-F(z)]^{k-1} f(z), \quad (2.4)$$

e

$$F_{k,n}(z) = \sum_{j=0}^{k-1} \frac{n!}{(j)!(n-j)!} [F(z)]^{n-j} [1-F(z)]^j, \quad (2.5)$$

para  $z$  no suporte de  $F$ .

Nos capítulos seguintes será necessário conhecer a distribuição conjunta das  $K$ -maiores estatísticas de ordem, esta é fornecida pelo teorema a seguir.

**Teorema 2.2** (*Embrechts, P., Klüppelberg, C., Mikosch, T. (1997)*) *A densidade conjunta para as  $K$  maiores estatísticas de ordem, sendo  $F$  absolutamente contínua com densidade  $f$ , é:*

$$f_{M_{1,n}, \dots, M_{K,n}}(x_1, \dots, x_k) = \frac{n!}{(n-k)!} F^{n-k}(x_k) \prod_{i=1}^k f(x_i), \quad x_k < \dots < x_1. \quad (2.6)$$

É simples verificar que a função de distribuição conjunta para as  $K$ -maiores estatísticas de ordem é dada pela seguinte expressão:

$$F_{M_{1,n}, \dots, M_{K,n}}(x_1, \dots, x_k) = \sum_{j=0}^{k-1} \binom{n}{j} F(x_k)^{n-j} (F(x_j) - F(x_k))^j \quad x_k < \dots < x_1.$$

Outras duas fórmulas que serão utilizadas são a função de densidade e distribuição conjunta entre a  $k$ -ésima e  $l$ -ésima estatística de ordem, ambas são resultados decorrentes do teorema anterior ou podem ser facilmente obtidas por cálculo diferencial.

A função de densidade bivariada de um par de estatísticas de ordem  $(M_{l,n}, M_{k,n})$  para  $1 \leq l < k \leq n$ , é calculada da seguinte forma:

$$f_{M_{l,n}M_{k,n}}(z, w) = \lim_{dz, dw \rightarrow 0} \frac{P(M_{l,n} \in \Delta z, M_{k,n} \in \Delta w)}{\Delta z \cdot \Delta w},$$

mas como:

$$\begin{aligned} P(M_{l,n} \in \Delta z, M_{k,n} \in \Delta w) &= (A_{n,2}) \cdot P(M_{l,n} \in \Delta z) \cdot P(M_{k,n} \in \Delta w) \\ &\cdot \binom{n-2}{n-k} [F(w)]^{n-k} \cdot \binom{k-2}{k-l-1} [F(z) - F(w)]^{k-l-1} \cdot \binom{l-1}{l-1} [1 - F(z)]^{l-1} \approx \\ (A_{n,2}) \cdot f(z) dz \cdot f(w) dw \cdot \binom{n-2}{n-k} [F(w)]^{n-k} \cdot \binom{k-2}{k-l-1} [F(z) - F(w)]^{k-l-1} \cdot [1 - F(z)]^{l-1} \end{aligned}$$

Onde:  $A_{n,2}$  = arranjo de  $n$  dois a dois, portanto :

$$\begin{aligned} \frac{P(M_{l,n} \in \Delta z, M_{k,n} \in \Delta w)}{\Delta z \cdot \Delta w} &\approx \\ (A_{n,2}) \cdot f(z) \cdot f(w) \cdot \binom{n-2}{n-k} [F(w)]^{n-k} \cdot \binom{k-2}{k-l-1} [F(z) - F(w)]^{k-l-1} \cdot [1 - F(z)]^{l-1}, \end{aligned}$$

tomando o Limite, tem-se:

$$f_{M_{l,n}M_{k,n}}(z, w) = (A_{n,2}) \cdot f(z) \cdot f(w) \cdot \binom{n-2}{n-k} [F(w)]^{n-k} \cdot \binom{k-2}{k-l-1} [F(z) - F(w)]^{k-l-1} \cdot [1 - F(z)]^{l-1}.$$

Já a função de distribuição é dada por:

$$\begin{aligned} F_{M_{l,n}M_{k,n}}(z, w) &= P(M_{l,n} \leq z, M_{k,n} \leq w) = \\ &= P(X_{(n)} \leq w, X_{(n-1)} \leq w, \dots, X_{(k+1)} \leq w, X_{(k)} \leq z, w < X_{(k-1)} \leq z, \dots, \\ &\quad , \dots, w < X_{(l+1)} \leq z, w < X_{(l)} \leq z, X_{(l-1)} > z, \dots, X_{(1)} > z) = \\ &= \binom{n}{n-k+1} [F(w)]^{n-k+1} \binom{k-1}{k-l} [F(z) - F(w)]^{k+l} [1 - F(z)]^{l-1}. \end{aligned}$$

Para o caso particular de  $l = 1$  e  $k = 2$ , obtemos:

$$\begin{aligned}
f_{M_{1,n}M_{2,n}}(z, w) &= (A_{n,2}) \cdot f(z) \cdot f(w) \cdot \binom{n-2}{n-2} [F(w)]^{n-2} \cdot \binom{2-2}{2-1-1} [F(z) - F(w)]^{2-1-1} \\
&\quad \cdot [1 - F(z)]^{1-1} \\
f_{M_{1,n}M_{2,n}}(z, w) &= n(n-1) \cdot f(z) \cdot f(w) \cdot [F(w)]^{n-2}
\end{aligned} \tag{2.7}$$

e

$$\begin{aligned}
F_{M_{1,n}M_{2,n}}(z, w) &= \binom{n}{n-2+1} [F(w)]^{n-2+1} \binom{2-1}{2-1} [F(z) - F(w)]^{2-1} \binom{1-1}{1-1} [1 - F(z)]^{1-1} \\
&= n [F(w)]^{n-1} [F(z) - F(w)].
\end{aligned} \tag{2.8}$$

É fácil verificar que:

$$\frac{\partial^2 F_{M_{1,n}M_{2,n}}(z, w)}{\partial z \partial w} = f_{M_{1,n}M_{2,n}}(z, w) = n(n-1) \cdot f(z) \cdot f(w) \cdot [F(w)]^{n-2}$$

Entretanto, as expressões acima não são muito usadas, pois, na maioria das situações práticas,  $F$  é desconhecida. A teoria dos valores extremos fornece um resultado assintótico, a generalização do Teorema de Fisher e Tippett (Fisher e Tippett, 1928), em que a função de distribuição limite independe da função de distribuição primitiva  $F$ , veja Teorema 4.2.3 de Embrechts, Klüppelberg e Mikosch (1997). Onde  $G_{1,(\xi,\mu,\sigma)}$  é a função de distribuição que representa a família das distribuições GEV (*Generalized Extreme Value*), onde  $\xi \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ .

**Teorema 2.3** (*Fisher-Tippett (1928) e Gnedenko (1943)*) *Se existem seqüências de constantes  $a_n > 0$  e  $b_n$  tais que*

$$Pr \left\{ \frac{M_{1,n} - b_n}{a_n} \leq z \right\} \rightarrow G_1(z) \text{ quando } n \rightarrow \infty$$

*então  $G_1$  é uma função de distribuição não-degenerada pertencente a uma das seguintes famílias:*

$$I : G_1(z) = \exp \left\{ -\exp \left[ - \left( \frac{z - \mu}{\sigma} \right) \right] \right\}, -\infty < z < \infty;$$



$$\begin{aligned}
II : G_1(z) &= \begin{cases} 0, & z \leq \mu, \\ \exp \left\{ - \left( \frac{z-\mu}{\sigma} \right)^{-\xi} \right\}, & z > \mu; \end{cases} \\
III : G_1(z) &= \begin{cases} \exp \left\{ - \left[ - \left( \frac{z-\mu}{\sigma} \right)^\xi \right] \right\}, & z < \mu \\ 1, & z \geq \mu; \end{cases}
\end{aligned}$$

para parâmetros  $\sigma > 0$ ,  $\mu \in \mathbb{R}$ , e no caso das famílias II e III,  $\xi > 0$ . Sendo chamadas respectivamente como, I: Gumbel, II: Fréchet e III: Weibull.

Os três tipos de distribuição limite para os máximos que são mencionados no Teorema 2.3 têm formas distintas de comportamento com relação à função de distribuição primitiva  $F$  dos dados. Nas primeiras utilizações de valores extremos era usual, *a priori*, adotar uma das três famílias e posteriormente estimar os parâmetros relevantes da distribuição escolhida. Contudo, há duas desvantagens nesse tipo de abordagem: primeiro, é necessário empregar uma técnica para escolher qual das três famílias é a mais apropriada para os dados que se têm à mão; Segundo, assim que essa decisão é tomada, as inferências subseqüentes presumem que a escolha feita no passo anterior foi a correta e não dá margem para investigar as incertezas envolvidas na escolha, embora tais incertezas possam ser substanciais. Uma melhor análise surgiu com a reformulação dos modelos apresentados no Teorema 2.3. É extremamente simples verificar que as famílias Gumbel, Fréchet e Weibull podem ser combinadas numa única família de modelos com função de distribuição da forma:

$$G_1(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad (2.9)$$

definida no conjunto  $\{z : z > \mu - \frac{\sigma}{\xi}\}$ , onde os parâmetros satisfazem  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$ . Esta é a distribuição generalizada de valores extremos (GEV). O modelo tem três parâmetros: um parâmetro de locação,  $\mu$ ; um parâmetro de escala,  $\sigma$ ; e um parâmetro de forma,  $\xi$ . As distribuições tipo II e tipo III de valores extremos correspondem, respectivamente, aos casos  $\xi > 0$  e  $\xi < 0$  na nova parametrização. O subconjunto da família GEV em que  $\xi = 0$  é interpretado como o limite de (2.9) quando

$\xi \rightarrow 0$ , conduzindo à família Gumbel com função de distribuição

$$G_1(z) = \exp \left\{ -\exp \left[ - \left( \frac{z - \mu}{\sigma} \right) \right] \right\}, -\infty < z < \infty;$$

A unificação das três famílias originais de valores extremos simplifica bastante a implementação estatística. Através da inferência em  $\xi$  os dados por si só determinam o tipo mais apropriado de comportamento de cauda, e não há necessidade de se fazer julgamentos subjetivos *a priori* sobre qual tipo de família de valores extremos adotar. Além do mais, a incerteza associada ao valor estimado de  $\xi$  mede a falta de certeza quanto a qual dos três tipos originais é o mais apropriado para um determinado conjunto de dados. Por conveniência, vamos reformular o Teorema 2.3 para a forma generalizada.

**Teorema 2.4** (*Jenkinson (1955)*) *Se existem seqüências de constantes  $a_n > 0$  e  $b_n$  tais que*

$$Pr \left\{ \frac{M_{1,n} - b_n}{a_n} \leq z \right\} \rightarrow G_1(z) \text{ quando } n \rightarrow \infty, \quad (2.10)$$

*para uma função de distribuição  $G_1$  não-degenerada, então  $G_1$  é membro da família GEV:*

$$G_1(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\},$$

*definida em  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , onde  $-\infty < \mu < \infty, \sigma > 0$  e  $-\infty < \xi < \infty$*

## 2.2.2 O modelo para as $K$ -maiores estatísticas de ordem

Na seção anterior, concluímos que a distribuição de  $M_{1,n}$  converge para  $G_1(z)$ , mas torna-se necessário generalizar para as  $K$ -maiores estatísticas de ordem, identificando qual o comportamento limite, para  $k$  fixo, quando  $n \rightarrow \infty$ . Os resultados a seguir generalizam o Teorema 2.3.

**Teorema 2.5** (*Leadbetter, M., Lindgren, G., Rootzén, H. (1983)*) *Supondo existirem seqüências de constantes  $a_n > 0$  e  $b_n$  tais que*

$$Pr \left\{ \frac{M_{1,n} - b_n}{a_n} \leq z \right\} \rightarrow G_1(z) \text{ quando } n \rightarrow \infty,$$

onde  $G_1$  é uma função de distribuição não-degenerada, tal que  $G_1$  é uma GEV dada por (2.9), então para  $m = 1, 2, \dots$ ,

$$Pr \left\{ \frac{M_{m,n} - b_n}{a_n} \leq z \right\} \rightarrow G_m(z),$$

onde

$$G_m(z) = \begin{cases} \exp\{-\Lambda(z)\} \sum_{j=0}^{m-1} \frac{\Lambda(z)^j}{j!}, & \text{se } \xi \left(\frac{z-\mu}{\sigma}\right) > -1 \text{ para } \xi \neq 0 \text{ or } z \in \mathbb{R} \text{ para } \xi = 0 \\ 0 & , \text{ se } z < \mu - \frac{\sigma}{\xi} \text{ para } \xi > 0 \\ 1 & , \text{ se } z > \mu - \frac{\sigma}{\xi} \text{ para } \xi < 0. \end{cases} \quad (2.11)$$

Derivando a expressão (2.11), obtemos a densidade

$$g_m(z) = \begin{cases} \exp\{-\Lambda(z)\} \frac{\Lambda'(z)\Lambda(z)^{m-1}}{(m-1)!}, & \text{se } \xi \left(\frac{z-\mu}{\sigma}\right) > -1 \text{ para } \xi \neq 0 \text{ ou } z \in \mathbb{R} \text{ para } \xi = 0 \\ 0 & , \text{ caso contrario,} \end{cases} \quad (2.12)$$

onde  $\Lambda(z) = -\log(G_1(z))$ ,  $\Lambda'(z) = d\Lambda(z)/dz$ , e também

$$\Lambda(z) = \Lambda_{\xi,\mu,\sigma}(z) = \begin{cases} [1 + \xi \left(\frac{z-\mu}{\sigma}\right)]^{-\frac{1}{\xi}} & , \text{ se } \xi \neq 0 \\ \exp\left(-\frac{z-\mu}{\sigma}\right) & , \text{ se } \xi = 0, \end{cases}$$

para algum  $-\infty < \mu < \infty$ ,  $\sigma > 0$  and  $-\infty < \xi < \infty$ . A distribuição enunciada acima é chamada de distribuição *Generalized Extreme Value (GEV)*, e classificada como sendo do tipo I(Gumbel), II(Fréchet) e III(Weibull) de acordo com os respectivos valores para  $\xi = 0$ ,  $\xi > 0$  e  $\xi < 0$ , a figura (2.2) ilustra as densidades limites e as funções de distribuição das cinco maiores estatísticas de ordem para  $\mu = 0$ ,  $\sigma = 1$  e  $\xi = 0.2$ . Note que a função  $\Lambda$  é uma função positiva estritamente decrescente e satisfaz

$$\begin{aligned} \lim_{z \rightarrow -\infty} \Lambda(z) = +\infty & \quad \text{e} \quad \lim_{z \rightarrow \infty} \Lambda(z) = 0, & \text{se } \xi = 0 \\ \lim_{z \downarrow (\mu - \frac{\sigma}{\xi})} \Lambda(z) = +\infty & \quad \text{e} \quad \lim_{z \rightarrow \infty} \Lambda(z) = 0, & \text{se } \xi > 0 \\ \lim_{z \rightarrow -\infty} \Lambda(z) = +\infty & \quad \text{e} \quad \lim_{z \uparrow (\mu - \frac{\sigma}{\xi})} \Lambda(z) = 0, & \text{se } \xi < 0. \end{aligned} \quad (2.13)$$

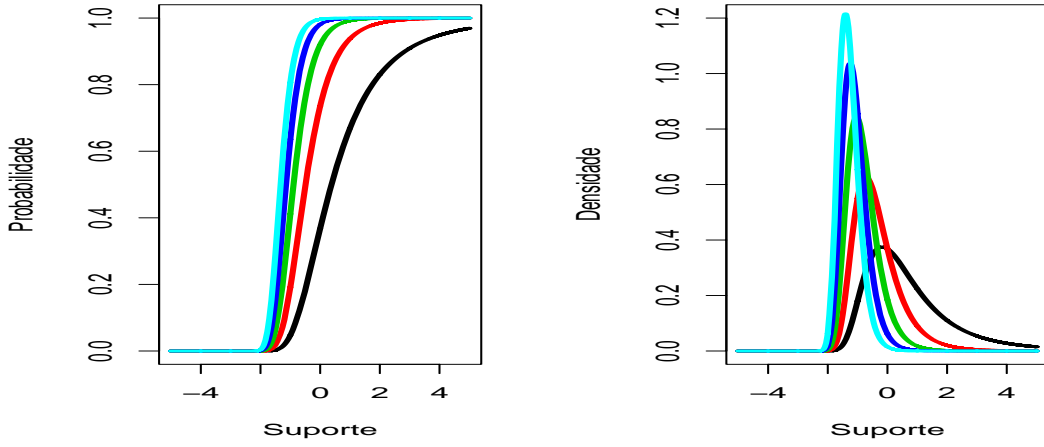


Figura 2.2: *Função de distribuição e função de densidade limite das 5 maiores estatísticas de ordem. À esquerda temos as funções de distribuição e à direita, as funções de densidade para  $\mu = 0$ ,  $\sigma = 1$  e  $\xi = 0.2$ .*

A distribuição conjunta exata de  $(M_{1,n}, M_{2,n}, \dots, M_{K,n})$  também não pode ser usada para inferências, pois, como no caso anterior, esta depende da distribuição primitiva  $F$ . Com a mudança de escala apropriada do vetor aleatório, uma distribuição limite pode ser obtida (veja Weissman (1978), Smith (1986), Tawn (1988)). Se  $F$  está no domínio de atração de alguma distribuição de valores extremos  $G_1(\xi, \mu, \sigma)$  com  $\xi \neq 0$  e com seqüências de constantes normalizadoras  $\{a_n > 0\}$  e  $\{b_n\}$ , então a distribuição limite quando  $n \rightarrow \infty$  de

$$\left( \frac{M_{1,n} - b_n}{a_n}, \dots, \frac{M_{K,n} - b_n}{a_n} \right), \quad (2.14)$$

tem a seguinte função de densidade conjunta (Smith,1986)

$$\tilde{g}(z_1, \dots, z_K) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z_K - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \prod_{i=1}^K \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi} - 1}, \quad (2.15)$$

no conjunto dos valores de  $z_i, i = 1, \dots, K$ , tal que  $1 + \xi(z_i - \mu)/\sigma > 0$  e  $z_K \leq \dots \leq z_1$ .

Em (2.15), as constantes de escalas desconhecidas são absorvidas pelo parâmetro de forma  $\xi$ , já que esta fornece a expressão da distribuição limite conjunta para os três tipos de GEV. Para uma prova de (2.15) ver Coles (2001).

O caso  $\xi = 0$  é fornecido em Weissman (1978), sendo interpretado como uma forma limite de (2.15). Quando  $\xi \rightarrow 0$ ,

$$\tilde{g}(z_1, \dots, z_K) = \exp \left\{ -\exp \left\{ -\left( \frac{z_K - \mu}{\sigma} \right) \right\} \right\} \times \prod_{i=1}^K \frac{1}{\sigma} \exp \left\{ -\left( \frac{z_i - \mu}{\sigma} \right) \right\}, \quad (2.16)$$

onde  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  e  $z_K \leq \dots \leq z_1$ . Em (2.16), o caso  $K = 1$  se reduz na densidade da família Gumbel. Uma outra expressão mais geral, que inclua os três casos é:

$$\tilde{g}_K(z_1, \dots, z_K) = \begin{cases} (-1)^K \exp\{-\Lambda(z_K)\} \prod_{j=1}^K \Lambda'(z_j) & , \text{ se } (z_1, \dots, z_K) \in \Omega_\xi \\ 0 & , \text{ caso contrario.} \end{cases} \quad (2.17)$$

onde

$$\Omega_\xi = \begin{cases} \mathbb{R}^K & , \text{ se } \xi = 0 \\ \{(z_1, \dots, z_K) \in \mathbb{R}^K : z_1 > \dots > z_K > \mu - \frac{\sigma}{\xi}\} & , \text{ se } \xi > 0 \\ \{(z_1, \dots, z_K) \in \mathbb{R}^K : \mu - \frac{\sigma}{\xi} > z_1 > \dots > z_K\} & , \text{ se } \xi < 0. \end{cases}$$

A distribuição cuja densidade é fornecida em (2.17) com parâmetros  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$  é chamada de distribuição *Multivariate Generalized Extreme Value (MGEV)*.

**Observação 2.1** *Uma vasta classe de seqüências de variáveis aleatórias estacionárias, tem como distribuição conjunta assintótica das K-maiores estatísticas de ordem, a distribuição MGEV.*

A função de distribuição das K-maiores estatísticas de ordem pode ser obtida quando consideramos o número de excessos obtidos em relação ao nível  $u_n = a_n x + b_n$  por  $X_1, \dots, X_n$ . Vamos definir  $S_n^{(i)}$  como o número de excessos em relação ao nível  $u_n^{(i)}$  por  $X_1, \dots, X_n$ , onde  $u_n^{(1)} \geq \dots \geq u_n^{(k)}$ .

**Teorema 2.6** (Leadbetter, M., Lindgren, G., Rootzén, H. (1983)) *Supondo que  $\{X_i\}$  é uma seqüência independente e identicamente distribuída e  $\{u_n^{(i)}\}$  satisfaça a  $n(1 -$*

$F(u_n^{(i)}) \rightarrow \Lambda(x_i)$ , para  $i = 1, \dots, K$ . Então para  $l_1 \geq \dots \geq l_K$ ,

$$\begin{aligned} & P\{S_n^{(1)} = l_1, S_n^{(2)} = l_1 + l_2, \dots, S_n^{(K)} = l_1 + \dots + l_K\} \\ & \rightarrow \frac{[\Lambda(x_1)]^{l_1}}{l_1!} \cdot \frac{[\Lambda(x_2) - \Lambda(x_1)]^{l_2}}{l_2!} \dots \frac{[\Lambda(x_K) - \Lambda(x_{K-1})]^{l_K}}{l_K!} \exp[-\Lambda(x_K)]. \square \end{aligned} \quad (2.18)$$

Fica evidente que a distribuição conjunta assintótica das  $K$ -maiores estatísticas de ordem pode ser obtida diretamente através do (2.6), pelo seguinte cálculo:

$$P\{M_{1,n} \leq u_n^{(1)}, \dots, M_{K,n} \leq u_n^{(K)}\} = P\{S_n^{(1)} = 0, S_n^{(1)} \leq 1, \dots, S_n^{(K)} \leq K - 1\} \quad (2.19)$$

Além disso, Leadbetter, M., Lindgren, G., Rootzén, H. (1983) afirmam que, se a distribuição de  $a_n(M_{1,n} - b_n)$  converge, então não somente a  $(M_{i,n} - b_n)/a_n$  converge em distribuição para  $i = 1, \dots, K$ , resultado já visto anteriormente, como também a distribuição conjunta de  $[(M_{1,n} - b_n)/a_n, \dots, (M_{K,n} - b_n)/a_n]$  converge. Este resultado é obtido diretamente de (2.19), mas, devido à forma limite da distribuição ser complicada de se obter quando  $K > 2$ , mostraremos, a seguir, o resultado somente para as duas maiores estatísticas de ordem.

**Teorema 2.7** (Leadbetter, M., Lindgren, G., Rootzén, H. (1983)) *Supondo existirem seqüências de constantes  $a_n > 0$  e  $b_n$  tais que*

$$Pr \left\{ \frac{M_{1,n} - b_n}{a_n} \leq z \right\} \rightarrow G_1(z) \text{ quando } n \rightarrow \infty,$$

onde  $G_1$  é uma função de distribuição não-degenerada, tal que  $G_1$  é uma GEV dada por (2.9). Então, para  $x_1 > x_2$ ,

$$P\{(M_{1,n} - b_n)/a_n \leq x_1, (M_{2,n} - b_n)/a_n \leq x_2\} \rightarrow \exp[-\Lambda(x_2)] [1 + \Lambda(x_2) - \Lambda(x_1)] \quad (2.20)$$

onde  $\Lambda(x_1) = G_1(x_1)$  e  $\Lambda(x_2) = G_1(x_2)$ .

**Prova:**

$$\begin{aligned} & P\{M_{1,n} \leq u_n^{(1)}, M_{2,n} \leq u_n^{(2)}\} = P\{S_n^{(1)} = 0, S_n^{(2)} \leq 1\} \\ & = P\{S_n^{(1)} = 0, S_n^{(2)} = 0\} + P\{S_n^{(1)} = 0, S_n^{(2)} = 1\} \\ & = \frac{[\Lambda(x_1)]^0}{0!} \cdot \frac{[\Lambda(x_2) - \Lambda(x_1)]^0}{0!} \exp[-\Lambda(x_2)] + \frac{[\Lambda(x_1)]^0}{0!} \cdot \frac{[\Lambda(x_2) - \Lambda(x_1)]^1}{1!} \exp[-\Lambda(x_2)] \\ & = \exp[-\Lambda(x_2)] [1 + \Lambda(x_2) - \Lambda(x_1)] \end{aligned}$$

onde  $\Lambda(x_1) = G_1(x_1)$  e  $\Lambda(x_2) = G_1(x_2)$ .  $\square$

Existem muitas situações em que um conjunto multivariado de dados mostra uma forte associação positiva, se assemelhando à estrutura de dependência pertencente ao modelo das  $K$ -maiores estatísticas (2.15), incluindo as características particulares do suporte da distribuição multivariada. Exemplos destas situações podem ser casos de inundações ou níveis elevados do mar onde, usualmente, o foco é na estimação dos riscos associados a alguns cenários de baixa probabilidade. Entretanto, o modelo em (2.15) pode não fornecer um bom ajuste para algumas das componentes marginais. Em tais situações pode-se desejar ter a opção de encontrar o melhor ajuste marginal possível, o qual poderia, então, ser ligado pela função da cópula pertencente ao modelo das  $K$ -maiores. Esta construção é apoiada pelo Teorema de Sklar (Sklar, 1959), sendo a motivação para a presente tese.

Portanto, nesta tese inicialmente deriva-se a expressão para a cópula  $C_{BIX}$ , a cópula Bi-extremal pertinente à distribuição assintótica das 2-maiores estatísticas de ordem de um processo (*i.i.d.*). O leitor pode consultar Leadbetter et al. (1983) sobre o desenvolvimento das condições teóricas preenchidas para os modelos de valores extremos, no caso de  $X_i, i = 1, 2, \dots$  ser oriunda de um processo (*i.i.d.*). Esta cópula é não-paramétrica e depende somente de  $F$  através das suposições feitas pelos teoremas limite.

Dois trabalhos relacionados são Schmitz (2004) e Avérous, Genest e Kochar (2005). Schmitz (2004) obtém a estrutura de dependência do mínimo e do máximo de  $n$  variáveis aleatórias (*i.i.d.*), que é determinada também através da cópula. Avérous, Genest and Kochar (2005) comparam o grau de associação apresentado por dois pares de estatísticas de ordem para a mesma distribuição contínua. Para finalizar, eles usam o conceito *bivariate monotone regression dependence*. É demonstrado, também, que a cópula associada ao par de estatísticas de ordem não depende da função de distribuição  $F$ , porém uma expressão para a cópula não foi elaborada. Um outro trabalho relacionado é Anjos, Kolev e Tanaka (2005), no qual é dada uma representação da cópula da função de distribuição conjunta das  $r$ -ésima e  $s$ -ésima estatísticas de ordem correspondentes aos pares de variáveis aleatórias.

# Capítulo 3

## A cópula *Bi*-extremal

O estudo da estrutura de dependência entre eventos extremos é cada vez mais pesquisado em todo mundo, principalmente aos dados relacionados ao mercado financeiro, ou geológicos. Neste capítulo iremos obter a expressão da cópula pertinente às duas maiores estatísticas de ordem, que será chamada de cópula *Bi*-extremal.

### 3.1 Especificação

Consideremos, inicialmente, o caso  $K = 2$ . Sem perda de generalidade, sejam  $\mu = 0$  e  $\sigma = 1$ . A distribuição marginal limite do máximo  $M_{1,n}$  e da segunda maior  $M_{2,n}$  são dadas por

$$G_1(z) = \exp\{-\Lambda(z)\} \text{ e } G_2(z) = \exp\{-\Lambda(z)\}[1 + \Lambda(z)], \quad (3.1)$$

Onde, define-se previamente,  $\Lambda(z) = -\log(G_1(z))$ . As respectivas densidades são

$$g_1(z) = -\exp\{-\Lambda(z)\}\Lambda'(z), \quad (3.2)$$

$$g_2(z) = -\exp\{-\Lambda(z)\}\Lambda'(z)\Lambda(z). \quad (3.3)$$

Utilizando a expressão (2.17), o caso bivariado tem a seguinte densidade conjunta assintótica

$$\tilde{g}_2(z_1, z_2) = \exp\{-\Lambda(z_2)\}\Lambda'(z_1)\Lambda'(z_2), \quad (3.4)$$

com  $z_2 \leq z_1$  e  $z_k : 1 + \xi z_k > 0$ , para o caso  $\xi \neq 0$  e  $z_k \in \mathbb{R}$  quando  $\xi = 0$ , para  $k = 1, 2$ .



A partir deste ponto e utilizando-se os resultados obtidos e ilustrados anteriormente, será possível obter a densidade  $c_{BIX}$  da cópula Bi-extremal.

**Teorema 3.1** (*Cópula Bi-extremal - função de densidade*). *A função de densidade da cópula Bi-extremal, pertinente à distribuição limite das duas maiores estatísticas de ordem oriundas de uma seqüência de variáveis aleatórias (i.i.d.), é dada por:*

$$c_{BIX}(u, v) = \frac{\psi(v)}{u(v - \psi(v))}, \quad (3.5)$$

sendo  $0 \leq u, v, \psi(v) \leq 1$ ,  $\psi(v) \leq u$ ,  $\psi(v) \leq v$  e  $\psi(v)$  é implicitamente definida por  $v = \psi(v)(1 - \log(\psi(v)))$ , e  $c(u, v) = 0$  caso contrário.

*Prova:* Substituindo as expressões (3.2), (3.3), (3.4) em (2.3), obtemos  $c_{BIX} = 1/(u\Lambda(z_2))$ . Seja  $\psi(v) = \exp\{-\Lambda(z_2)\}$ . Então,  $\Lambda(z_2) = -\log(\psi(v))$  e  $c_{BIX} = (-1)/(u\log(\psi(v)))$ . Note-se que  $z_2 \leq z_1 \Rightarrow \psi(v) \leq u$ , de (3.1), e, desde que  $G_1(z_2) < G_2(z_2)$ , implica  $\psi(v) < v$ . Como  $v = \exp\{-\Lambda(z_2)\}[1 + \Lambda(z_2)]$ , temos  $v = \psi(v)(1 - \log(\psi(v)))$  e o resultado se segue. Na figura (3.1) verificar-se o gráfico da densidade da cópula Bi-extremal.

□

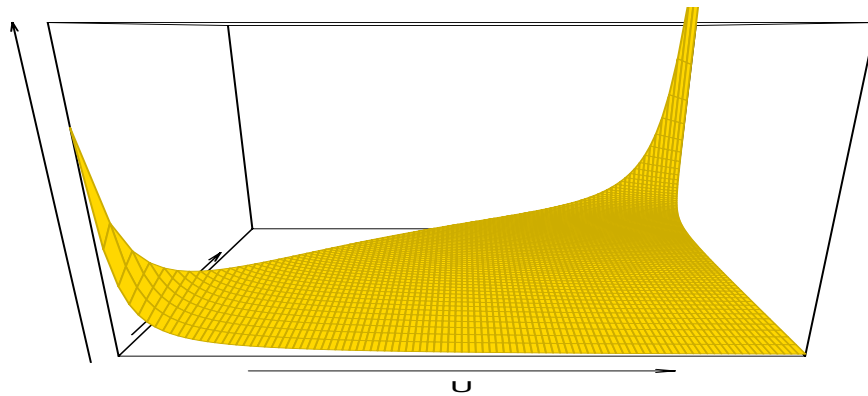


Figura 3.1: *Perspectiva da função de densidade da cópula Bi-Extremal*

**Observação 3.1** *Por definição, a função  $\psi$  é a inversa da função  $x \mapsto x(1 - \log(x))$ . Para uma avaliação numérica da função  $\psi$ , que será necessária para simulações futuras, utilizamos a software R-project.*

**Observação 3.2** Da Proposição 10 em Avérous, Genest e Kochar (2005), temos que a cópula Bi-extremal é, também, a cópula de  $(-M_{n,n}, -M_{n-1,n})$  (a menor e a segunda menor estatística de ordem) associada à amostra aleatória de tamanho  $n$  de alguma distribuição contínua.

É fácil verificar que a função de distribuição bivariada limite das duas maiores estatísticas de ordem, é dada por

$$\tilde{G}_2(z_1, z_2) = \begin{cases} \exp[-\Lambda(z_2)] [1 + \Lambda(z_2) - \Lambda(z_1)] & \text{se } z_2 \leq z_1 \\ \exp[-\Lambda(z_1)] & \text{se } z_2 > z_1 \end{cases} \quad (3.6)$$

**Prova:** Quando  $z_2 \leq z_1$ , sabe-se que

$$\tilde{G}_2(z_1, z_2) = 1 - P(M_1 > z_1) - P(M_2 > z_2) + P(M_1 > z_1, M_2 > z_2),$$

por outro lado,

$$\begin{aligned} P(M_1 > z_1, M_2 > z_2) &= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} \exp[-\Lambda(z_2)] [\Lambda'(z_1)] [\Lambda'(z_2)] dz_2 dz_1 = \\ &= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} \exp[-\Lambda(z_2)] d[-\Lambda(z_2)] d[-\Lambda(z_1)] \\ &= \int_{z_1}^{+\infty} \{ \exp[-\Lambda(z_2)] \}_{z_2}^{z_1} d[-\Lambda(z_1)] \\ &= \left\{ \int_{z_1}^{+\infty} \exp[-\Lambda(z_1)] d[-\Lambda(z_1)] \right\} - \left\{ \exp[-\Lambda(z_2)] \int_{z_1}^{+\infty} d[-\Lambda(z_1)] \right\} \\ &= \{ \exp[-\Lambda(z_1)] \}_{z_1}^{+\infty} - \exp[-\Lambda(z_2)] \cdot \{ -\Lambda(z_1) \}_{z_1}^{+\infty} \\ &= 1 - \exp[-\Lambda(z_1)] - \exp[-\Lambda(z_2)] [\Lambda(z_1)]. \end{aligned}$$

Também sabe-se que:

$$P(M_1 > z_1) = 1 - P(M_1 \leq z_1) = 1 - \exp[-\Lambda(z_1)]$$

$$P(M_2 > z_2) = 1 - P(M_2 \leq z_2) = 1 - \exp[-\Lambda(z_2)] [1 + \Lambda(z_2)],$$

utilizando as expressões obtidas anteriormente tem-se:

$$\begin{aligned}
\tilde{G}_2(z_1, z_2) &= 1 - \{1 - \exp[-\Lambda(z_1)]\} \\
&\quad - \{1 - \exp[-\Lambda(z_2)] - \exp[-\Lambda(z_2)][\Lambda(z_2)]\} \\
&\quad + \{1 - \exp[-\Lambda(z_1)] - \exp[-\Lambda(z_2)][\Lambda(z_1)]\} \\
&= \exp[-\Lambda(z_2)] [1 + \Lambda(z_2) - \Lambda(z_1)].
\end{aligned} \tag{3.7}$$

Para o caso em que  $z_2 > z_1$ , temos que:

$$\begin{aligned}
\tilde{G}_2(z_1, z_2) &= P(M_1 \leq z_1, M_2 \leq z_2) \\
&= P(M_1 \leq z_1, M_2 \leq z_1) \\
&= \exp[-\Lambda(z_1)] [1 + \Lambda(z_1) - \Lambda(z_1)] \\
&= \exp[-\Lambda(z_1)],
\end{aligned}$$

resultado este idêntico ao do teorema 2.7.

**Observação 3.3** *Como simples verificação deste resultado, pode-se aplicar o limite na expressão acima, de onde obtêm-se as distribuições marginais encontradas em (3.1). Vejamos:*

$$\begin{aligned}
\lim_{z_2 \rightarrow z_1} \tilde{G}_2(z_1, z_2) &= G_1(z_1) = \exp[-\Lambda(z_1)] \\
\lim_{z_1 \rightarrow +\infty} \tilde{G}_2(z_1, z_2) &= G_2(z_2) = \exp[-\Lambda(z_2)] (1 + \Lambda(z_2))
\end{aligned} \tag{3.8}$$

do teorema de Sklar, temos

$$C_{BIX} = \tilde{G}_2(G_1^{-1}(u), G_2^{-1}(v)), \tag{3.9}$$

Onde  $G_i^{-1}(\cdot)$  representa a função inversa de  $G_i, i = 1, 2$ .

Para se obter a cópula é necessário, inicialmente, calcular as funções inversas  $G_i^{-1}(\cdot), i = 1, 2$ . Note que  $G_1(z_1) = u = \exp\{-\Lambda(z_1)\}$ , implicando  $\Lambda(z_1) = -\log(u)$ . Então a inversa de  $G_1$  é  $z_1 = ((-\log(u))^{-\xi} - 1)/\xi$ . Como  $\exp[-\Lambda(z_2)][1 + \Lambda(z_2)] = v$  e, por definição,  $v = \psi(v)(1 - \log(\psi(v)))$ , tem-se  $\psi(v) = G_1(z_2) \Rightarrow -\log(\psi(v)) = \Lambda(z_2)$ . A inversa de  $G_2$  é  $z_2 = \frac{(\frac{v-\psi(v)}{\psi(v)})^{-\xi}-1}{\xi}$  e este resultado segue de que  $v = \psi(v)(1 + \Lambda(z_2)) \Rightarrow \frac{v-\psi(v)}{\psi(v)} =$

$(1 + \xi z_2)^{-1/\xi}$ . Observe que  $\Lambda(z_2) = \left(1 + \xi \frac{\left(\frac{v-\psi(v)}{\psi(v)}\right)^{-\xi} - 1}{\xi}\right)^{-1/\xi} = \frac{v-\psi(v)}{\psi(v)}$ . Com estes resultados estamos em posição de obter a cópula  $C_{BIX}$ .

**Teorema 3.2** (A cópula  $C_{BIX}$ ). A cópula Bi-extremal  $C_{BIX}$ , pertinente à distribuição limite das duas maiores estatísticas de ordem oriundas de uma seqüência de variáveis aleatórias (i.i.d.), é dada por

$$C_{BIX} = \begin{cases} v + \psi(v)\log(u), & \text{se } v \leq u(1 - \log(u)) \\ u, & \text{se } v > u(1 - \log(u)) \end{cases} \quad (3.10)$$

*Prova:*

(i) Primeiro, note que  $z_2 \leq z_1 \Rightarrow \Lambda(z_2) \geq \Lambda(z_1)$  o que implica  $u(1 - \log(\psi(v))) \leq u(1 - \log(u))$ . Como  $\psi(v) \leq u$  e  $v = \psi(v)(1 - \log(\psi(v)))$ ,  $v \leq u(1 - \log(u))$ . Para  $v \leq u(1 - \log(u))$ ,  $\tilde{G}_2(G_1^{-1}(u), G_2^{-1}(v)) = \exp\left\{-\frac{v-\psi(v)}{\psi(v)}\right\}\left(1 + \frac{v-\psi(v)}{\psi(v)} + \log(u)\right) = \exp\left\{-\frac{v-\psi(v)}{\psi(v)}\right\}\left(\frac{v}{\psi(v)} + \log(u)\right)$ . De  $v = \psi(v)(1 - \log(\psi(v)))$ , obtemos  $\tilde{G}_2(G_1^{-1}(u), G_2^{-1}(v)) = \psi(v)(1 - \log(\psi(v))) + \psi(v)\log(u)$ , de onde se segue o resultado.

(ii) Quando  $z_1 < z_2$  e para  $\xi > 0$ , onde  $\Lambda(z)$  é uma função decrescente em  $z$ , temos:  $\exp\{\Lambda(z_1)\} = G_1(z_1) = u < \psi(v) = G_1(z_2) < G_2(z_2) = v$ , e então  $C_{BIX} = u = \min(u, v)$ .  $\square$

Entretanto, segundo Nelsen (2006), a cópula Bi-extremal tem que satisfazer a três propriedades, que são:

$$\mathbf{1-} C(u, 0) = 0 = C(0, v) \quad \forall u, v;$$

$$\mathbf{2-} C(u, 1) = u \quad \text{e} \quad C(1, v) = v;$$

$$\mathbf{3-} C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 \quad \text{tal que} \quad u_1 \leq u_2, v_1 \leq v_2.$$

Estas propriedades para a cópula Bi-extremal podem ser facilmente verificadas.

**Prova:**

(i) Temos que o maior valor que  $v$  assume é  $v = u(1 - \ln u)$  logo:

$$C(u, 0) = 0 + 0 \ln u = 0$$

$$C(0, v) = \lim_{u \rightarrow 0} (v + \psi(v) \ln u) \leq \lim_{u \rightarrow 0} [u(1 - \ln u) + u \ln u] = \lim_{u \rightarrow 0} u = 0$$

(ii)

$$C(u, 1) = u(1 - \ln u) + u \ln u = u$$

$$C(1, v) = v + \psi(v) \ln 1 = v$$

(iii) Sejam  $a, b, c, d \in (0, 1)$  com  $a \leq b$  e  $c \leq d$ , logo :

$$\begin{aligned} & C(b, d) - C(b, c) - C(a, d) + C(a, c) = \\ &= (d + \psi(d) \ln b) - (c + \psi(c) \ln b) - (d + \psi(d) \ln a) + (c + \psi(c) \ln a) = \\ &= \psi(c)(\ln a - \ln b) + \psi(d)(\ln b - \ln a) - \psi(c)(\ln b - \ln a) + \psi(d)(\ln b - \ln a) = \\ &= \psi(c)(\ln a - \ln b) + \psi(d)(\ln b - \ln a) = \\ &= -\psi(c)(\ln b - \ln a) + \psi(d)(\ln b - \ln a) = \\ &= (\psi(d) - \psi(c))(\ln b - \ln a) \end{aligned}$$

Como  $(\psi(d) - \psi(c)) \geq 0$  e  $(\ln b - \ln a) \geq 0$ , implica que:

$$(\psi(d) - \psi(c))(\ln b - \ln a) \geq 0$$

Com esta prova fica claro que a cópula Bi-extremal pertence à classe das cópulas.  $\square$

**Observação 3.4** A cópula Bi-Extremal  $C_{BIX}$  é não permutável, ou seja,

$$C_{BIX}(u, v) \neq C_{BIX}(v, u).$$

Para ajudar a entender melhor a cópula Bi-extremal, a Figura (3.3) ilustra, na parte superior, as linhas de contorno da  $C_{BIX}$  (esquerda) e as linhas de contorno da função de densidade  $c_{BIX}$  (direita). Na parte inferior da figura, temos as seções diagonais da  $C_{BIX}$ ,  $M$  e  $I$  (esquerda), e a seção diagonal das respectivas densidade (direita).

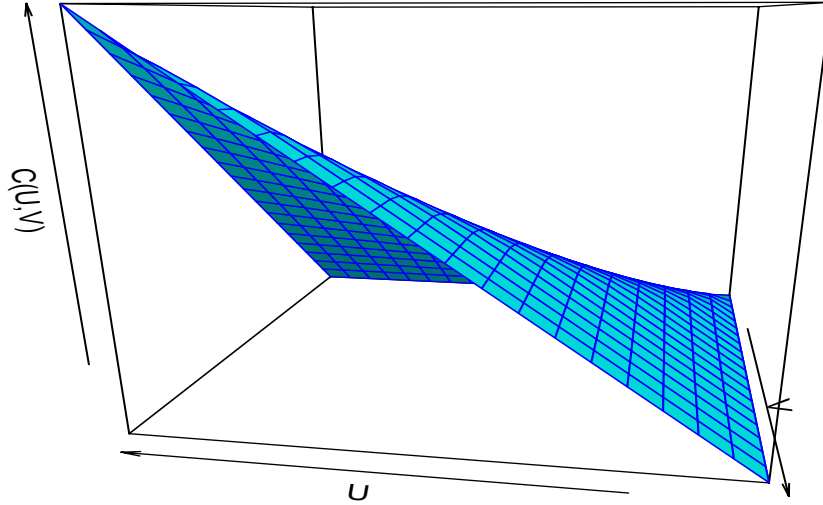


Figura 3.2: Gráfico da cópula *Bi-Extremal*.

## 3.2 Algoritmo para Simulação

Uma vez que recentemente se tenha proposto uma nova cópula, torna-se desejável gerar pseudo-observações desta. O algoritmo clássico para simulação (Frees and Valdez, 1998) extrai  $U$ , uniformemente distribuído no intervalo  $[0, 1]$ , e depois gera  $V$  a partir da distribuição condicional  $C_{BIX}(u|v) = \partial C_{BIX}(u, v)/\partial v$ . A expressão de  $C_{BIX}(u|v)$  e sua inversa são dadas na **proposição 3.1** e **observação 3.5**.

**Proposição 3.1** *Seja  $(U, V)$  um par de variáveis aleatórias com função de distribuição  $C_{BIX}$  e supondo todas as definições e notações encontradas nos Teorema 3.1 e Teorema 3.2. A distribuição condicional  $C_v(u|v) = Pr\{U \leq u|V = v\} = 1 - \frac{\log(u)}{\log(\psi(v))}$  para todo  $u, v \leq 1$  e  $v \leq u(1 - \log(u))$  e  $C_v(u|v) = 0$  caso contrario.*

*Prova:*

Note, inicialmente, que quando  $v > u(1 - \log(u))$ ,  $C_{BIX} = u$ , então  $\partial C_{BIX}(u, v)/\partial v = 0$ . Mas, para  $v \leq u(1 - \log(u))$ ,  $C_v(u|v) = Pr\{U \leq u|V = v\} = \partial C_{BIX}(u, v)/\partial v =$

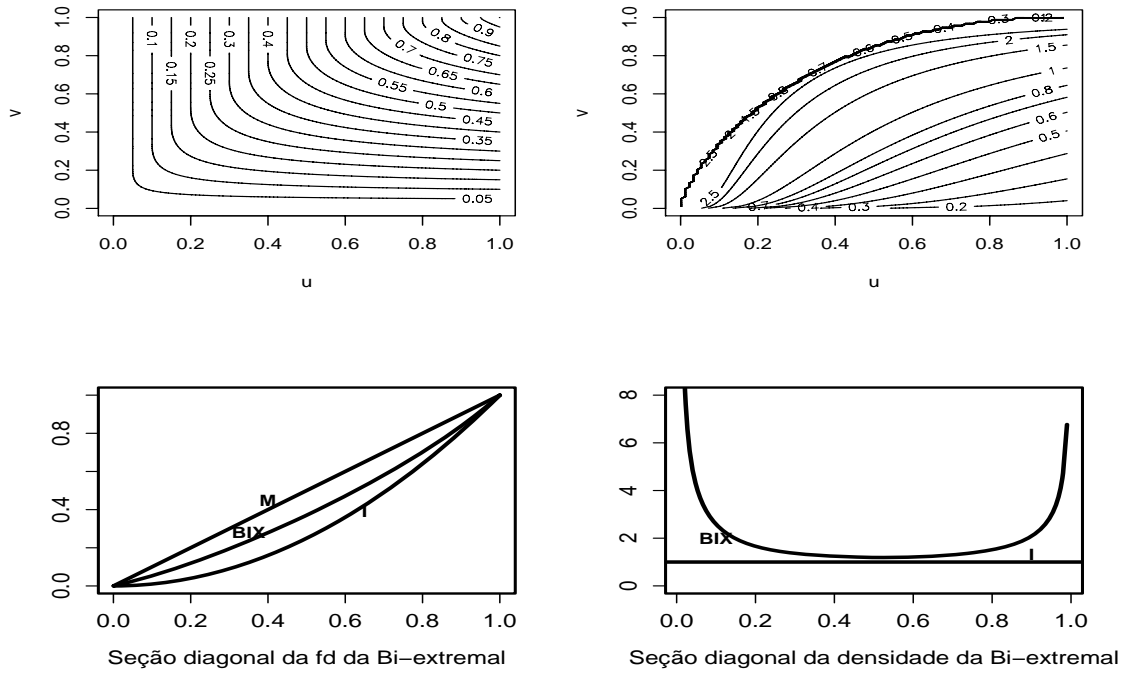


Figura 3.3: Gráficos da cópula BI-extremal. Na parte superior, os gráficos ilustram as curvas de contorno da cópula Bi-extremal  $C_{BIX}$  (esquerda) e da função de densidade  $c_{BIX}$ . Na parte inferior, podemos ver, no gráfico da esquerda, as seções da diagonal principal das cópulas  $C_{BIX}$ ,  $M$  (cópula de dependência perfeita) e  $I$  (produto); e no lado direito as seções diagonais das densidades correspondentes.

$$\frac{\partial(v+\psi(v)\log(u))}{\partial v} = 1 + \frac{\partial\psi(v)}{\partial v}\log(u) \square$$

**Observação 3.5** A função inversa de  $C_v(u|v)$  é dada por  $C_v^{-1}(q|v) = \exp\{(1-q)\log(\psi(v))\}$ . Para se gerar uma observação  $(u, v)$  da  $C_{BIX}$ , pode-se proceder da seguinte maneira:

1. Simule  $v$  de uma  $Uniforme(0, 1)$ ;
2. Calcule numericamente,  $\psi(v)$  definido implicitamente por  $v = \psi(v)(1 - \log(\psi(v)))$ ;
3. Simule  $q$  de uma  $Uniforme(0, 1)$ ;
4. Calcule  $u = \exp\{(1 - q)\log(\psi(v))\}$ .

### 3.3 Exemplos teóricos

Agora iremos provar dois exemplos, nos quais a  $C_{BIX}$  pode ser obtida analiticamente. Seja  $X_1, X_2, \dots, X_n$  uma seqüência de variáveis aleatórias (*i.i.d.*), com distribuição  $F$  e densidade  $f$  comuns. Assumindo todas as notações fornecidas anteriormente, então, das expressões (2.4) e (2.5), obtemos  $F_{1,n} = (F(y))^n$ ,  $F_{2,n} = (F(x))^n + n(F(x))^{n-1}(1 - F(x))$ , tendo como respectivas densidades  $f_{1,n}(y) = n(F(y))^{n-1}f(y)$ ,  $f_{2,n} = n(n - 1)f(x)(F(x))^{n-2}(1 - F(x))$ , para  $x < y$ . A função de distribuição conjunta  $F_{M_{1,n}M_{2,n}}(y, x)$  é obtida pela expressão (2.8) e dada por

$$F_{M_{1,n}M_{2,n}}(y, x) = n(F(x))^{n-1}(F(y) - F(x)),$$

já densidade conjunta dada pela expressão (2.7) é

$$f_{M_{1,n}M_{2,n}}(y, x) = n(n - 1)(F(x))^{n-2}f(y)f(x).$$

Sejam  $a_n$  e  $b_n$  constantes normalizadoras. As estatísticas de ordem padronizadas são  $M_{i,n}^* = (M_{i,n} - b_n)/a_n$ ,  $i = 1, 2$ , com distribuição  $F_{M_{i,n}^*} = F_{i,n}(a_n z_i + b_n)$ , função de densidade  $f_{M_{i,n}^*}(z_i) = a_n f_{i,n}(a_n z_i + b_n)$ ,  $i = 1, 2$ , função de distribuição conjunta  $F_{M_{1,n}^*, M_{2,n}^*}(z_1, z_2) = F_{M_{1,n}M_{2,n}}(a_n z_1 + b_n, a_n z_2 + b_n)$  e função de densidade conjunta  $f_{M_{1,n}^*, M_{2,n}^*}(z_1, z_2) = (a_n)^2 \cdot f_{M_{1,n}M_{2,n}}(a_n z_1 + b_n, a_n z_2 + b_n)$ .

#### 3.3.1 O caso Uniforme(0,1)

Seja  $F$  uma *Uniforme(0,1)*. Então,  $F_{1,n} = y^n$ ,  $F_{2,n}(x) = nx^{n-1}(1 - x) + x^n$ ,  $f_{1,n}(y) = ny^{n-1}$ ,  $f_{2,n}(x) = n(n - 1)x^{n-2}(1 - x)$ , para  $0 \leq x \leq y \leq 1$ . A função de distribuição conjunta é dada por

$$F_{M_{1,n}M_{2,n}}(y, x) = nx^{n-1}(y - x) \quad \text{se } 0 < x < y < 1$$

e a densidade conjunta é

$$f_{M_{1,n}M_{2,n}}(y, x) = n(n - 1)x^{n-2}, \quad \text{se } 0 < x < y < 1$$

sejam  $a_n = \frac{1}{n}$  e  $b_n = 1$ . Então

$$f_{M_{1,n}^*}(z_1) = \left(1 + \frac{z_1}{n}\right)^{n-1}$$



e

$$f_{M_{2,n}^*}(z_2) = \left(\frac{n-1}{n}\right) (-z_2) \left(1 + \frac{z_2}{n}\right)^{n-2}.$$

A função de distribuição correspondente a

$$F_{M_{1,n}^*, M_{2,n}^*}(z_1, z_2) = n \left(\frac{z_2}{n} + 1\right) (z_1 - z_2)$$

com a seguinte densidade

$$f_{M_{1,n}^*, M_{2,n}^*}(z_1, z_2) = \left(\frac{n-1}{n}\right) \left(\frac{z_2}{n} + 1\right)^{n-2},$$

substituindo estas expressões na equação (2.3), obtemos

$$c(F_{M_{1,n}^*}(z_1), F_{M_{2,n}^*}(z_2)) \longrightarrow \lim_{n \rightarrow \infty} \frac{1}{(-z_2) \left(1 + \frac{z_1}{n}\right)^{n-1}} = \frac{-1}{z_2 \exp\{z_1\}}.$$

Agora, no caso Weibull com  $\xi = -1$ , temos  $G_1(z) = \exp\{z\}$ . Se  $u = G_1(z_1)$  e  $\psi(v) = G_1(z_2)$ , nós obtemos  $c(u, v) = \frac{-1}{u \log(\psi(v))} \square$

### 3.3.2 O caso Exponencial(1)

Seja  $F$  exponencial(1). Então  $f_{1,n} = n(1 - \exp\{-y\})^{n-1} \exp\{-y\}$ ,  $f_{2,n}(x) = n(n-1) \exp\{-2x\} (1 - \exp\{-x\})^{n-2}$ , para  $x > 0$  e  $y > 0$ . A densidade conjunta é

$$f_{M_{1,n}, M_{2,n}}(y, x) = n(n-1) (1 - \exp\{-x\})^{n-2} \exp\{-x\} \exp\{-y\}, \quad \text{se } 0 \leq x < y.$$

Sejam  $a_n = 1$  e  $b_n = \log(n)$ . Então

$$f_{M_{1,n}^*}(z_1) = \exp\{-z_1\} \left(1 - \frac{\exp\{-z_1\}}{n}\right)^{n-1}$$

e

$$f_{M_{2,n}^*}(z_2) = \left(\frac{n-1}{n}\right) \exp\{-2z_2\} \left(1 - \frac{\exp\{-z_2\}}{n}\right)^{n-2},$$

com função de densidade dada por:

$$f_{M_{1,n}^*, M_{2,n}^*}(z_1, z_2) = \left(\frac{n-1}{n}\right) \exp\{-z_1\} \exp\{-z_2\} \left(1 - \frac{\exp\{-z_2\}}{n}\right)^{n-2}.$$

Substituindo estas expressões na equação (2.3), obtemos

$$c(F_{M_{1,n}^*}(z_1), F_{M_{2,n}^*}(z_2)) \longrightarrow \lim_{n \rightarrow +\infty} \frac{1}{\exp\{-z_2\} \left(1 - \frac{\exp\{-z_1\}}{n}\right)^{n-1}} = \frac{-1}{-\exp\{-z_2\} \exp\{-\exp\{-z_1\}\}}.$$

Agora, no caso Gumbel com  $\xi = 0$ , temos  $G_1(z) = \exp\{-\exp\{-z\}\}$ . Se  $u = G_1(z_1)$  e  $\psi(v) = G_1(z_2)$ , obtemos  $c(u, v) = \frac{-1}{u \log(\psi(v))} \square$

### 3.4 Exemplos utilizando seleções de processos (*i.i.d.*)

Nosso propósito é verificar, neste estágio, os resultados assintóticos para amostras finitas. Foram construídos 8 experimentos de simulação e as amostras simuladas para cada experimento foram das seguintes distribuições: (*i*) Normal(0,1), (*ii*) Normal(0,4), (*iii*) t-student padrão (3 graus de liberdade) e (*iv*) t-student padrão (6 graus de liberdade). Inicialmente,  $N$  observações (*i.i.d.*) do processo selecionado foram geradas e, em seguida, as séries foram divididas em blocos de tamanho  $n$  e as duas maiores observações foram coletadas. Foram feitos blocos de tamanho  $n = 30$  e, para cada processo, consideramos dois tamanhos de série,  $N = 900$  e  $N = 3000$ , a distribuição bivariada de valores extremos (2.15) foi ajustada ao par de observações e as estimativas dos parâmetros foram usadas em (2.11) para obter as observações pseudo-Uniforme(0,1). Na figura (3.4), como exemplo ilustrativo, podemos verificar o histograma dos dados originais, da maior e da segunda maior observação coletadas para cada bloco, sendo o tamanho da série igual a  $N = 90.000$  observações geradas da distribuição  $N(0, 1)$ .

Para checar se os dados serão corretamente modelados pela cópula Bi-extremal, adotamos um teste de qualidade de ajuste baseado no teste qui-quadrado descrito em Patton e Andrew (2006)(veja também Fermanian (2005)), que é uma extensão bivariada do teste usual de Pearson. O teste consiste em dividir o quadrado unitário em  $m \times m$  quadrados e, coletar  $O_{i,j}$  e  $E_{i,j}$ , definidos como sendo o número observado e a frequência esperada para cada célula  $(i, j); i, j = 1, \dots, m$ . E seja  $D$  a estatística de teste, definida a seguir,

$$D = \sum_i^m \sum_j^m \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}},$$

esta segue uma distribuição qui-quadrado com  $v$  graus de liberdade, onde  $v = (m \times m) - 1$ .

Construímos um grid sobre  $[0, 1]^2$ , com  $m = 10$ , resultando em 100 células, conforme ilustra a figura (3.5), onde, para uma certa realização das 200 iterações, pode ser visto o valor observado, bem como o valor esperado para cada uma das 100 células construídas sobre o grid  $[0, 1]^2$ . Em DeGroot (2002) é mencionado que a estatística de teste  $D$  tem uma boa aproximação pela distribuição  $\chi^2$  quando o valor esperado de cada célula é superior a 1.5. Logo, como forma de evitar distorções no valor calculado da estatística

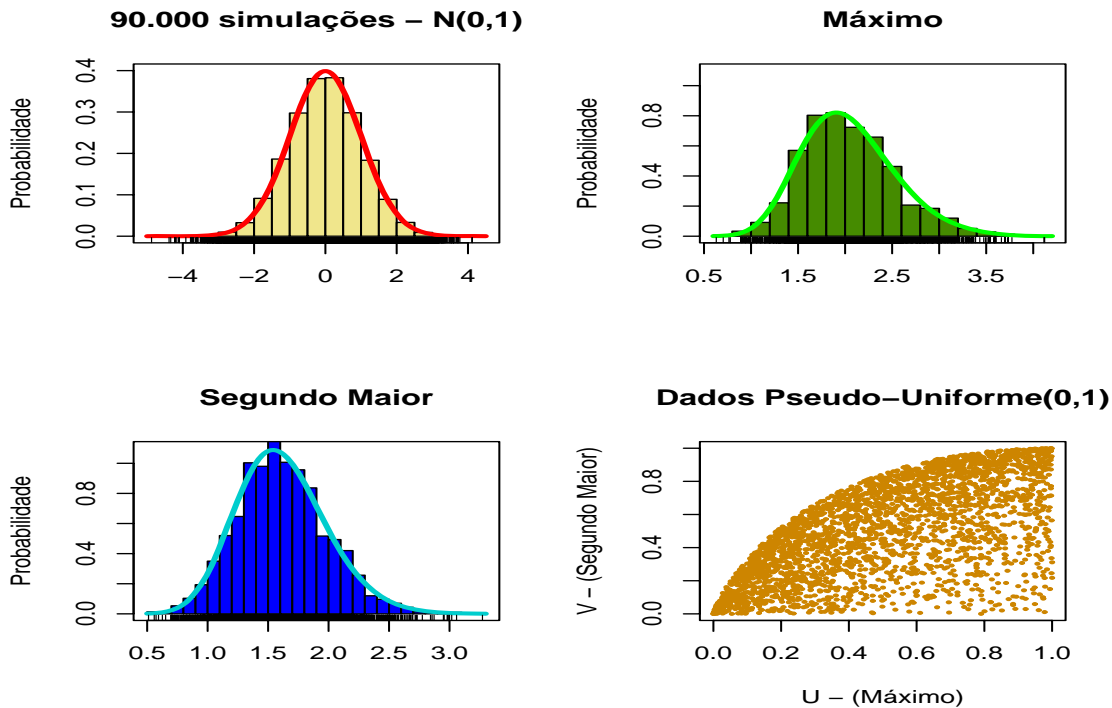


Figura 3.4: *Histogramas do processo de simulação. Na parte superior da figura: à esquerda o histograma de 90.000 observações da distribuição  $N(0,1)$ , à direita o histograma dos máximos observados para os blocos de tamanho  $n = 30$ . Na parte inferior: à esquerda o histograma das segundas maiores observações e à direita os dados pseudo-uniforme(0,1)*

de teste, todas as células que possuíam valor esperado inferior a um certo  $\epsilon$ , aqui neste caso igual a 1.5, tiveram os valores esperados e observados agrupados em uma célula vizinha, e assim sucessivamente até que todas as células apresentassem valor esperado superior ao  $\epsilon$  determinado. Para cada repetição do processo de simulação, após o cálculo da estatística de teste  $D$ , foi calculado o  $p$ -valor na distribuição qui-quadrado, onde foi atribuído a  $v$  o número de células com probabilidade positiva menos 1, na figura (3.6) podemos verificar para uma certa realização do processo de simulação a frequência dos  $p$ -valores obtidos. A partir daí, como forma de verificar a aderência dos dados à cópula Bi-Extremal, foram calculados a média e o erro padrão do  $p$ -valor para as 200 repetições de cada experimento realizado.

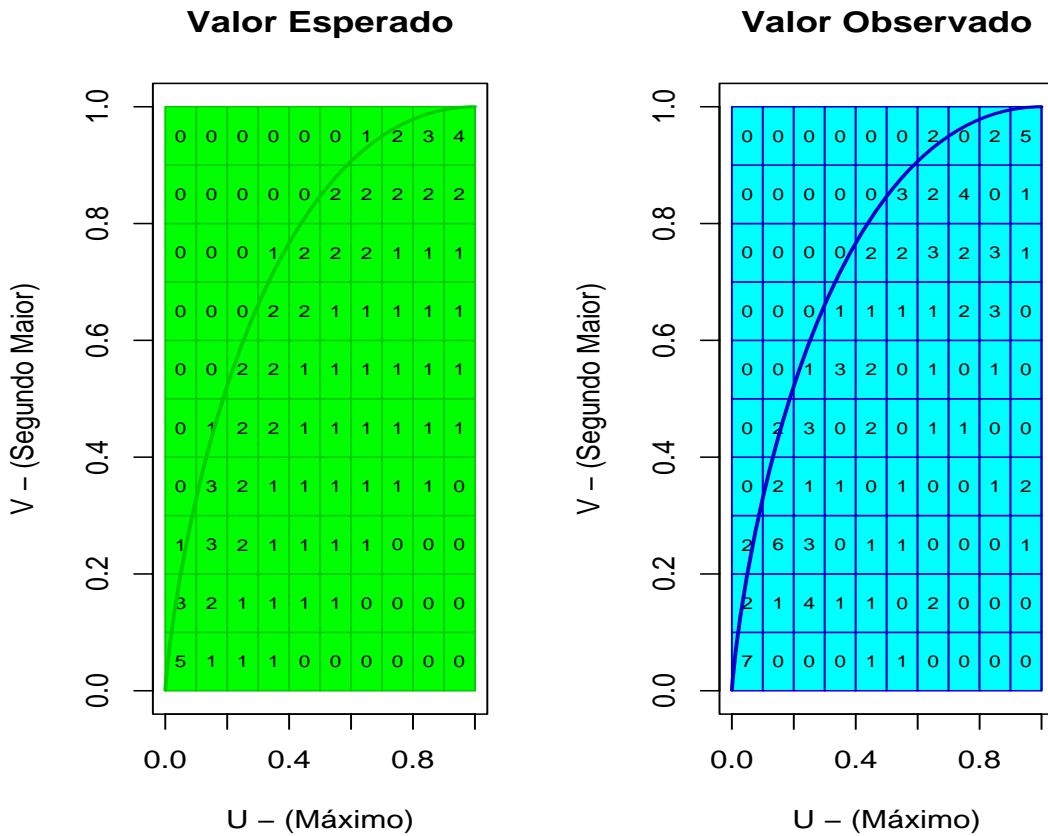


Figura 3.5: Comparativo entre dados reais e simulados ( $N = 3000; n = 30$ ). À esquerda, de forma ilustrativa, temos o valor esperado arredondado, para cada célula calculada a partir da  $C_{BIX}$ , já à direita, o valor observado para uma certa realização das 200 repetições do processo de simulação.

Os resultados foram apresentados na tabela 3.1. Como esperado, a média dos  $p$ -valores ficou próxima de 0.50, aceitando, portanto, a hipótese nula de que os dados podem ser modelados pela cópula Bi-extremal.

Tabela 3.1: Média e erro padrão dos p-valores obtidos nas simulações

	$N(0, 1)$	$N(0, 4)$	$t - st(3)$	$t - st(4)$
Pequena amostra de tamanho (30)				
média	0.5692	0.6050	0.5477	0.5638
erro padrão	0.3355	0.3283	0.3475	0.3215
Moderada amostra de tamanho (100)				
média	0.5296	0.5504	0.5708	0.5568
erro padrão	0.3112	0.3093	0.3077	0.3082

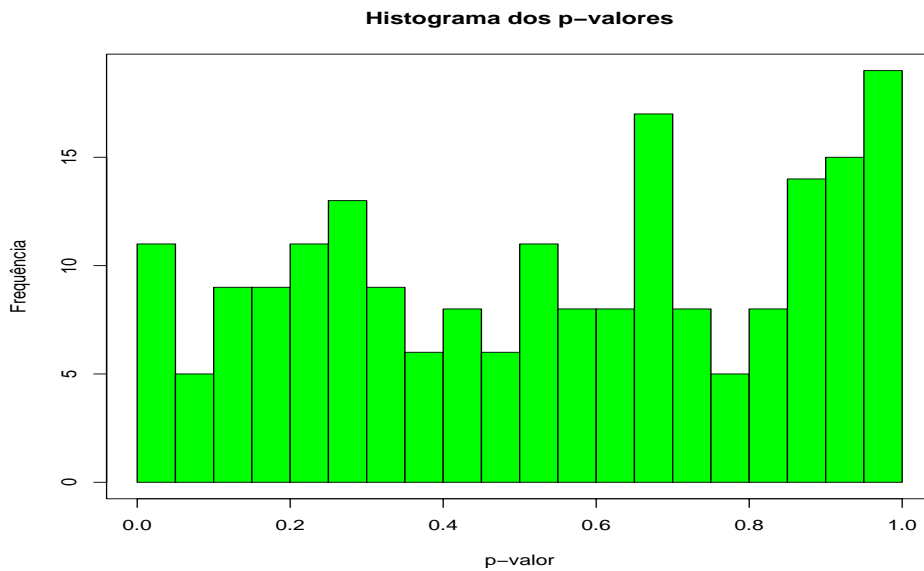


Figura 3.6: *Histograma dos p-valores obtidos. P-valores oriundos das 200 repetições do processo gerados a partir da  $N(0,1)$ , com tamanho de amostra  $n = 100$ .*

### 3.5 Verificação empírica para a Bi-extremal

Como forma de verificar a possibilidade da modelagem de dados reais com a utilização da estrutura de dependência da cópula Bi-extremal, foram coletadas observações históricas de duas séries financeiras, sendo estas: a cotação de fechamento do IBOVESPA, que atualmente é o principal índice de negociações de ações no Brasil, e também a cotação de fechamento das ações da PETROBRAS no período de 02/01/1994 a 17/06/2008. Os dados são provenientes do site *www.bovespa.com.br* totalizando 3325 observações coletadas e ambas as séries podem ser visualizadas na figura (3.7). Como esperado, as séries possuem uma grande dependência temporal, logo, como forma de obter séries de dados no mínimo estacionárias, iremos trabalhar com o logaritmo dos retornos destas séries históricas; estes dados transformados podem ser visualizados na figura (3.8).

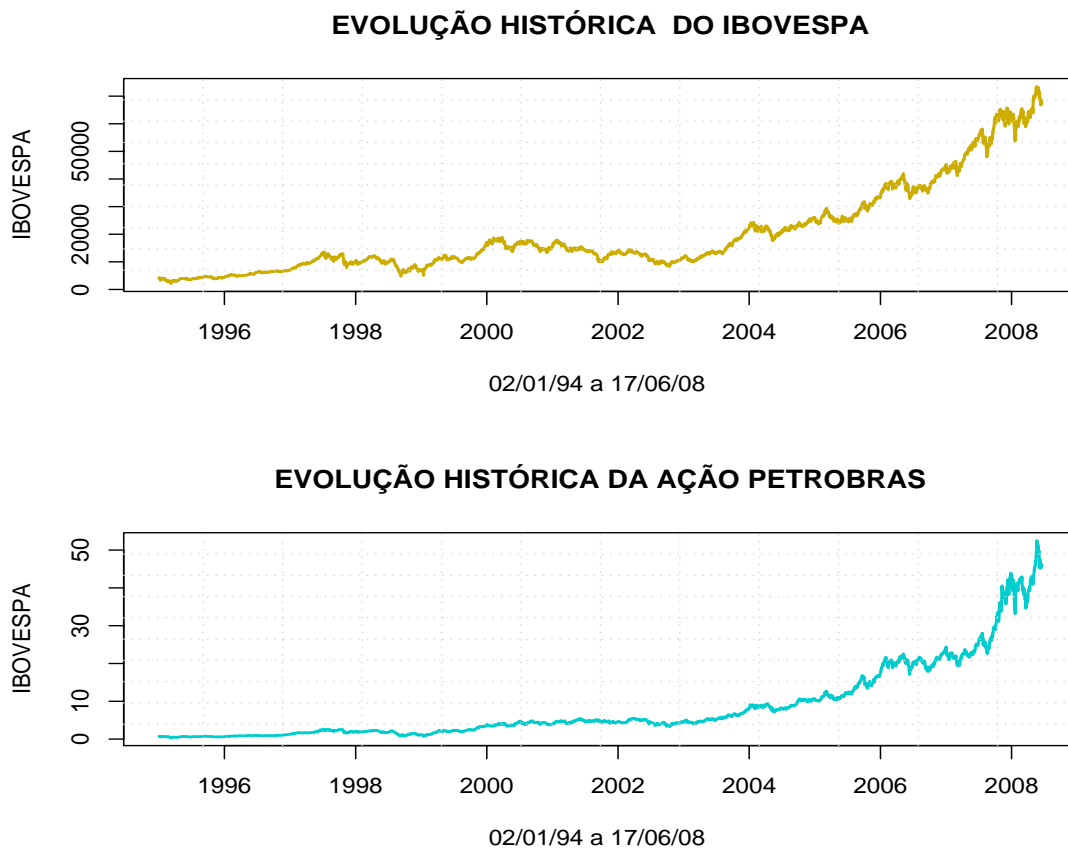


Figura 3.7: *Evolução histórica do índice IBOVESPA e das ações da PETROBRAS.*

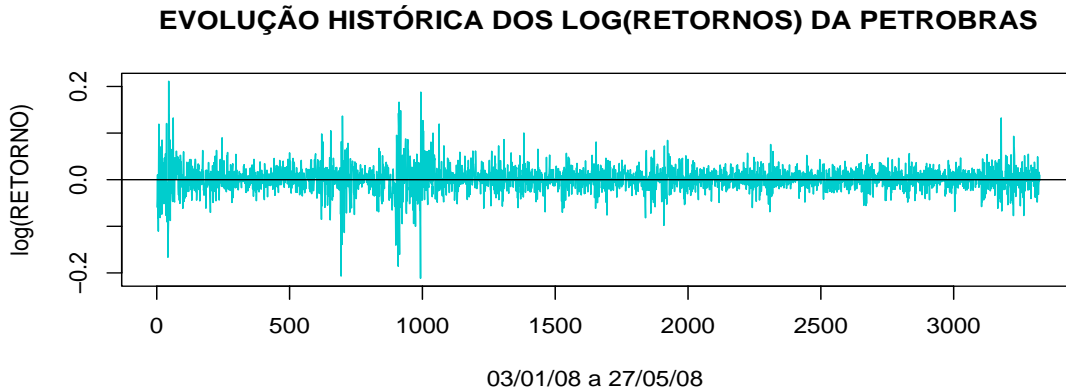
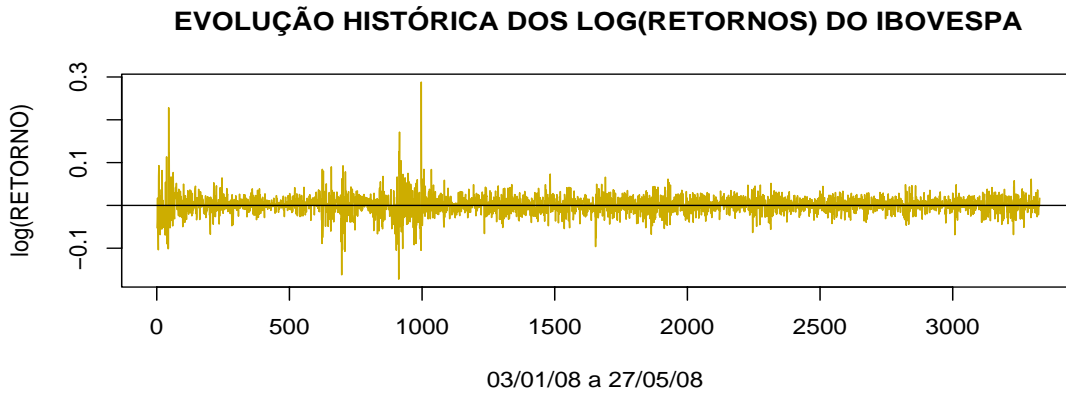


Figura 3.8: *Evolução histórica do  $\log(\text{RETORNOS})$  do IBOVESPA e das ações da PETROBRAS.*

Inicialmente os dados de ambas as séries foram divididos em blocos de tamanho  $n$ . Neste estudo o valor escolhido para  $n$  foi igual ao número médio de dias úteis que um mês contém, ou seja,  $n = 22$ . Posteriormente, para cada bloco, foi coletado o par das duas maiores estatísticas de ordem, o que permitiu coletar 151 observações do máximo e do segundo maior valor para cada um dos blocos, em ambas as séries. Além disso, como os parâmetros a serem estimados para o máximo e para a segunda maior estatísticas de ordem padronizadas são os mesmos, optamos por ajustar a distribuição conjunta das 2-maiores estatísticas de ordem, seção (2.2.2), à cada uma das série financeira. Para maiores detalhes ver estudo em Rubem, A.P.S. (2006), onde se verifica que as estimativas provenientes deste ajuste possuem erro padrão menor do que as estimativas provenientes de ajustes separados para o máximo e segundo maior. Na figura (3.9) podemos obser-

var os histogramas das duas maiores estatísticas de ordem coletadas e suas respectivas densidades ajustadas.

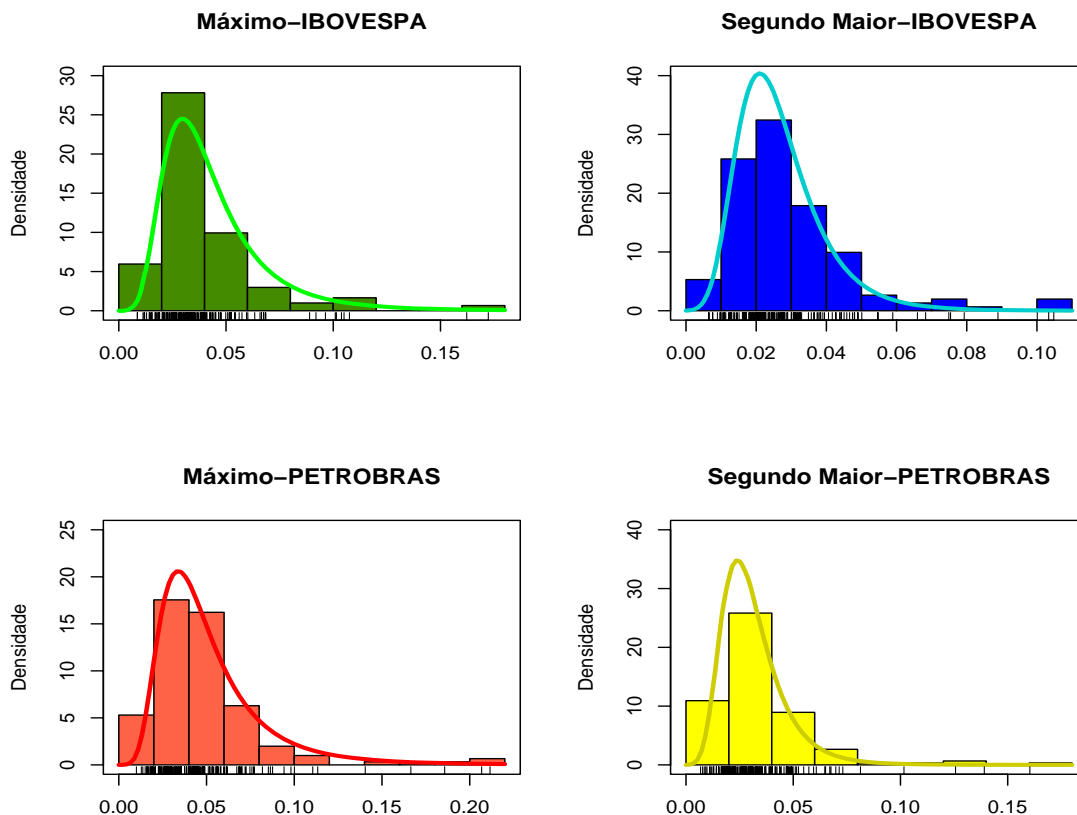


Figura 3.9: *Histograma e respectiva função de densidade para os valores observados do máximo e segundo maior, das séries IBOVESPA e PETROBRAS.*

Para ambas as séries, após o ajuste, foram obtidos os valores pseudo-uniformes para o máximo e para a segunda maior observação. A figura (3.10) ilustra os dados no quadrado unitário, tanto para o índice IBOVESPA, quanto para as cotações da PETROBRAS. Além disso, temos, neste gráfico, dados simulados da cópula Bi-extremal. O que podemos constatar visualmente é que, em ambos os casos, existe uma forte evidência de que a cópula Bi-extremal, é apropriada para representar a estrutura de dependência da distribuição conjunta, entre o máximo e a segunda maior observação em ambas as séries. Somente a análise gráfica não é suficiente para verificar a aderência dos dados à cópula Bi-extremal; é necessário verificar esta modelagem através de um teste de hipótese; o



teste descrito na seção (3.4) foi realizado para cada série. Para o IBOVESPA, foi encontrado um  $p$ -valor igual a 0.5002, não rejeitando a hipótese de que os dados uniformizados são oriundos da cópula Bi-extremal; o mesmo ocorreu para as cotações da PETROBRAS, onde foi obtido um  $p$ -valor igual a 0.4305.

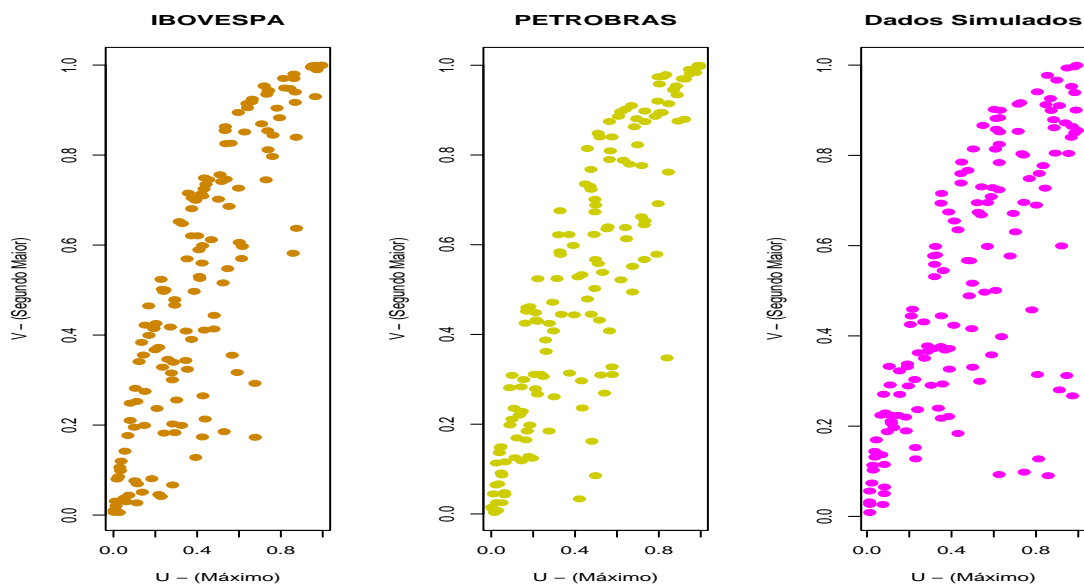


Figura 3.10: *Dados Uniformizados e Simulados.*

# Capítulo 4

## Propriedades da Cópula Bi-extremal

A estrutura de dependência revelada pela cópula Bi-extremal é de grande importância para a modelagem de dados bivariados oriundos de uma GEV bivariada. Esta estrutura pode ser melhor identificada através de algumas medidas de dependência e de associação. Neste capítulo, nosso intuito é apresentar as propriedades oriundas da cópula Bi-extremal. Para tanto, iremos obter algumas destas medidas de dependência. As medidas de dependência e associação, cujos valores iremos calcular, são muito conhecidas e sintetizam a estrutura de dependência capturada pela cópula Bi-extremal.

### 4.1 Medidas de Dependência

#### 4.1.1 O coeficiente de cauda

O coeficiente de cauda tem como objetivo medir a dependência na cauda superior e inferior no quadrante  $\mathbf{I}^2$ . O conceito de dependência de cauda em sua grande maioria é do interesse de profissionais que estejam preocupados em modelar, e conseqüentemente prever, eventos relacionados com as caudas das distribuições. Embora a probabilidade destes seja baixa, a ocorrência destes eventos pode representar grandes perdas, em especial aos relacionados ao mercado financeiro. Logo é de extrema importância estudarmos qual é o comportamento limite destas probabilidades.

**Definição 4.1** (*Nelsen (2006)*) *Sejam  $X$  e  $Y$  variáveis aleatórias contínuas com funções*

de distribuição  $F$  e  $G$ , respectivamente. O coeficiente de dependência de cauda superior denominado  $\lambda_U$  é o limite (se existir) da probabilidade condicional de  $Y$  ser maior que o  $t$ -ésimo percentil de  $G$  dado que  $X$  é maior que o  $t$ -ésimo percentil de  $F$  quando  $t$  tende para 1, ou seja,

$$\lambda_U = \lim_{t \rightarrow 1^-} P[Y > G^{(-1)}(t) | X > F^{(-1)}(t)],$$

similarmente, o coeficiente de dependência de cauda inferior denominado  $\lambda_L$  é o limite (se existir) da probabilidade condicional de  $Y$  ser menor ou igual ao o  $t$ -ésimo percentil de  $G$  dado que  $X$  é menor ou igual ao  $t$ -ésimo percentil de  $F$  quando  $t$  tende para 0, ou seja,

$$\lambda_L = \lim_{t \rightarrow 0^+} P[Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t)],$$

A partir da definição acima, nosso intuito é verificar quais são os valores de  $\lambda_U$  e  $\lambda_L$  para a cópula Bi-extremal.

**Teorema 4.1** *O coeficiente de dependência de cauda: O coeficiente de dependência de cauda superior e inferior da cópula Bi-extremal são iguais a zero.*

**Prova:** Seja a seção diagonal da cópula Bi-extremal representada por  $\delta(t)$  e definida da forma  $\delta(t) = C(t, t)$ . Os coeficientes de dependência  $\lambda_U$  e  $\lambda_L$  para uma cópula  $C$  qualquer são calculados usando as seguintes expressões Nelsen (2006)

$$\lambda_U = 2 - \lim_{t \rightarrow 1} \frac{\partial \delta(t)}{\partial t} \quad \text{e} \quad \lambda_L = \lim_{t \rightarrow 0} \frac{\partial \delta(t)}{\partial t}$$

No caso da cópula Bi-extremal, a função da seção diagonal é:

$$\delta_C(t) = C(t, t) = t + \psi(t) \ln t$$

Onde  $\psi(t)$  satisfaz  $t = \psi(t)(1 - \ln \psi(t))$ . Temos que

$$\frac{\partial \psi(t)}{\partial t} = -\frac{1}{\ln \psi(t)}$$

e portanto

$$\frac{\partial \delta(t)}{\partial t} = 1 + \frac{\psi(t)}{t} - \frac{\ln t}{\ln \psi(t)}.$$

Logo, para  $\lambda_U$  tem-se:

$$\frac{\partial\delta(1^-)}{\partial t} = \lim_{t \rightarrow 1^-} \left( 1 + \frac{\psi(t)}{t} - \frac{\ln t}{\ln \psi(t)} \right) = \lim_{t \rightarrow 1^-} 1 + \lim_{t \rightarrow 1^-} \frac{\psi(t)}{t} - \lim_{t \rightarrow 1^-} \frac{\ln t}{\ln \psi(t)} = 2 - \lim_{t \rightarrow 1^-} \frac{\ln t}{\ln \psi(t)},$$

pois  $\lim_{t \rightarrow 1^-} \psi(t) = 1$ . Empregando a regra de L'Hôpital:

$$\frac{\partial\delta(1^-)}{\partial t} = 2 - \lim_{t \rightarrow 1^-} \frac{\frac{1}{t}}{\frac{1}{\psi(t)} \frac{\partial\psi(t)}{\partial t}} = 2 + \lim_{t \rightarrow 1^-} \frac{\psi(t) \log \psi(t)}{t}.$$

Novamente usando o fato que  $\lim_{t \rightarrow 1^-} \psi(t) = 1$ , temos que

$$\frac{\partial\delta'(1^-)}{\partial t} = 2.$$

Finalmente, conclui-se que  $\lambda_U = 2 - 2 = 0$ .

Para  $\lambda_L$ ;

$$\frac{\partial\delta(0)}{\partial t} = \lim_{t \rightarrow 0} 1 + \frac{\psi(t)}{t} - \frac{\ln t}{\ln \psi(t)} = \lim_{t \rightarrow 0^+} 1 + \lim_{t \rightarrow 0^+} \frac{\psi(t)}{t} - \lim_{t \rightarrow 0^+} \frac{\ln t}{\ln \psi(t)}$$

Empregando a regra de L'Hôpital:

$$\frac{\partial\delta(0)}{\partial t} = 1 - \lim_{t \rightarrow 0} \frac{1}{\ln \psi(t)} + \lim_{t \rightarrow 0^+} \frac{\psi(t) \ln \psi(t)}{t} = 1 + \lim_{t \rightarrow 0} \frac{\psi(t) \ln \psi(t)}{t}.$$

Empregando novamente a regra de L'Hôpital;

$$\frac{\partial\delta(0)}{\partial t} = 1 + \lim_{t \rightarrow 0} \frac{1 \cdot \frac{-1}{\ln \psi(t)} \ln \psi(t) + \psi(t) \frac{1}{\psi(t)} \frac{-1}{\ln \psi(t)}}{1} = 1 - \lim_{t \rightarrow 0^+} 1 - \lim_{t \rightarrow 0^+} \frac{1}{\ln \psi(t)} = 1 - 1 = 0$$

Concluindo que:  $\lambda_L = 0$ .  $\square$

Os resultados obtidos para  $\lambda_U = 0$  e  $\lambda_L = 0$  indicam que os extremos da distribuição conjunta de  $(M_1, M_2)$  em ambas as caudas são assintoticamente independentes.

### 4.1.2 *Rho de Spearman*

A medida conhecida como  $\rho$  de Spearman na sua versão populacional é baseada na concordância e discordância. Seja  $(X_1, Y_1), (X_2, Y_2)$  e  $(X_3, Y_3)$  três vetores aleatórios independentes, com função de distribuição conjunta  $H$  comum, marginais  $F$  e  $G$  e cópula  $C$ . A versão populacional  $\rho_{X,Y}$  de Spearman é definida como sendo a proporção da probabilidade de concordância menos a probabilidade de discordância para os vetores  $(X_1, Y_1)$  e  $(X_2, Y_3)$ , ou seja:

$$\rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

Esta medida pode ser obtida também , através da cópula  $C$  associada ao vetor  $(X, Y)$ , sua definição é encontrada no teorema 5.1.6, Nelsen(2006). Logo, para qualquer cópula  $C$ , o  $\rho$  de Spearman é calculado da seguinte forma

$$\begin{aligned}\rho_C &= 12 \int \int_{I^2} [C(u, v) - uv] dudv \\ &= 12 \int \int_{I^2} C(u, v) dudv - 3 \\ &= 12 \int \int_{I^2} uv dC(u, v) - 3\end{aligned}\tag{4.1}$$

**Proposição 4.1** *O coeficiente de correlação de Spearman,  $\rho_{S,BIX}$ , da cópula Bi-Extremal é igual a 2/3.*

**Prova:** A definição do  $\rho_{S,BIX}$ , encontrada em Nelsen (2006), é:

$$\begin{aligned}\rho_{S,BIX} &= 12 \int \int_{I^2} [C_{BIX}(u, v) - uv] dudv = \\ &= 12 \int_0^1 \int_0^{u(1-\ln u)} (v + \psi(v) \ln u - uv) dv du + 12 \int_0^1 \int_{u(1-\ln u)}^1 (u - uv) dv du\end{aligned}$$

Para facilitar o cálculo, podemos separar em duas partes, ou seja:

$$\rho_{S,BIX} = \rho_{S,BIX}(1) + \rho_{S,BIX}(2).$$

Desta forma vamos calcular inicialmente o valor de  $\rho_{S,BIX}(1)$ ,

$$\begin{aligned}\rho_{S,BIX}(1) &= 12 \int_0^1 \int_0^{u(1-\ln u)} (v + \psi(v) \ln u - uv) dv du \\ &= 12 \int_0^1 \int_0^{u(1-\ln u)} [v(1-u) + \psi(v) \ln u] dv du \\ &= 12 \int_0^1 \left[ \int_0^{u(1-\ln u)} v(1-u) dv + \int_0^{u(1-\ln u)} \psi(v) (\ln u) dv \right] du \\ &= 12 \int_0^1 \left[ (1-u) \int_0^{u(1-\ln u)} v dv + (\ln u) \int_0^{u(1-\ln u)} \psi(v) dv \right] du.\end{aligned}$$

Como por definição  $v = \psi(v)(1 - \ln \psi(v))$ , implica que  $dv = -\ln \psi(v) d\psi(v)$ , logo continuando:

$$= 12 \int_0^1 \left[ (1-u) \left( \frac{v^2}{2} \right) \right]_0^{u(1-\ln u)} + (\ln u) \int_0^u \psi(v) (-\ln \psi(v)) d\psi(v) du$$

$$\begin{aligned}
&= 12 \int_0^1 \left[ (1-u) \frac{u^2(1-\ln u)^2}{2} - (\ln u) \int_0^u (\ln \psi(v)) \psi(v) d\psi(v) \right] du \\
&= 12 \int_0^1 \left\{ (1-u) \frac{u^2(1-\ln u)^2}{2} - (\ln u) \left[ \left( \ln \psi(v) \frac{\psi(v)^2}{2} \right)_0^u - \int_0^u \frac{\psi(v)^2}{2} \frac{1}{\psi(v)} d\psi(v) \right] \right\} du \\
&= 12 \int_0^1 \left[ (1-u) \frac{u^2(1-\ln u)^2}{2} - (\ln u) \left( \ln u \frac{u^2}{2} - \frac{1}{2} \frac{u^2}{2} \right) \right] du \\
&= 12 \int_0^1 \left[ (1-u) \frac{u^2(1-\ln u)^2}{2} - \frac{(\ln u)^2 u^2}{2} + \frac{u^2(\ln u)}{4} \right] du \\
&= 3 \int_0^1 [2u^2(1-\ln u)^2 - 2u^3(1-\ln u)^2 - 2u^2(\ln u)^2 + u^2(\ln u)] du \\
&= 3 \int_0^1 [2u^2 + 2u^2(\ln u)^2 - 4u^2 \ln u - 2u^3 + \\
&\quad -2u^3(\ln u)^2 + 4u^3 \ln u - 2u^2(\ln u)^2 + u^2 \ln u] du \\
&= 3 \int_0^1 [2u^2 - 3u^2 \ln u - 2u^3 - 2u^3(\ln u)^2 + 4u^3 \ln u] du.
\end{aligned}$$

Calculando separadamente, tem-se:

$$\begin{aligned}
&\int_0^1 2u^2 du = 2 \left( \frac{u^3}{3} \right)_0^1 = \frac{2}{3} \\
&-3 \int_0^1 u^2 \ln u du = -3 \left[ \left( \ln u \frac{u^3}{3} \right)_0^1 - \int_0^1 \frac{u^3}{3} \frac{1}{u} du \right] = - (u^3 \ln u)_0^1 + \left( \frac{u^3}{3} \right)_0^1 \\
&= (0 + \lim_{u \rightarrow 0} u^3 \ln u) + \left( \frac{1}{3} - 0 \right) = (\lim_{u \rightarrow 0} -u^3) + \frac{1}{3} = \frac{1}{3} \\
&-2 \int_0^1 u^3 du = -2 \left( \frac{u^4}{4} \right)_0^1 = -2 \frac{1}{4} = -\frac{1}{2} \\
&-2 \int_0^1 u^3 (\ln u)^2 du = -2 \left[ \left( (\ln u)^2 \frac{u^4}{4} \right)_0^1 - \int_0^1 \frac{u^4}{4} 2(\ln u) \frac{1}{u} du \right] \\
&= \int_0^1 u^3 (\ln u) du = \left[ \left( (\ln u) \frac{u^4}{4} \right)_0^1 - \int_0^1 \frac{u^4}{4} \frac{1}{u} du \right] = -\frac{1}{4} \left( \frac{u^4}{4} \right)_0^1 = -\frac{1}{4} \frac{1}{4} = -\frac{1}{16} \\
&4 \int_0^1 u^3 (\ln u) du = 4 \left[ \left( (\ln u) \frac{u^4}{4} \right)_0^1 - \int_0^1 \frac{u^4}{4} \frac{1}{u} du \right] = -4 \int_0^1 u^3 du = -4 \left( \frac{u^4}{4} \right)_0^1 = -\frac{1}{4},
\end{aligned}$$

logo:

$$\rho_{S,BIX}(1) = 3 \left( \frac{2}{3} + \frac{1}{3} - \frac{1}{2} - \frac{1}{16} - \frac{1}{4} \right) = \frac{9}{16}.$$

Por outro lado:

$$\begin{aligned}
\rho_{S,BIX}(2) &= 12 \int_0^1 \int_{u(1-\ln u)}^1 (u - uv) dv du = 12 \int_0^1 \int_{u(1-\ln u)}^1 u(1-v) dv du \\
&= 12 \int_0^1 \left[ u \left( v - \frac{v^2}{2} \right) \right]_{u(1-\ln u)}^1 du \\
&= 12 \int_0^1 \left\{ u \left[ \frac{1}{2} - u(1-\ln u) + \frac{u^2(1-\ln u)^2}{2} \right] \right\} du \\
&= 12 \int_0^1 \left( \frac{u}{2} - u^2(1-\ln u) + \frac{u^3(1-\ln u)^2}{2} \right) du.
\end{aligned}$$

Calculando, novamente, cada membro interno da integral,

$$\begin{aligned}
\int_0^1 \frac{u}{2} &= \frac{1}{2} \left[ \frac{u^2}{2} \right]_0^1 = \frac{1}{4} \\
\int_0^1 u^2(1-\ln u) du &= \left[ (1-\ln u) \frac{u^3}{3} \right]_0^1 - \int_0^1 \frac{u^3-1}{3} \frac{1}{u} du = \frac{1}{3} + \frac{1}{3} \left[ \frac{u^3}{3} \right]_0^1 = \frac{1}{3} + \frac{1}{9} = \frac{4}{9} \\
\frac{1}{2} \int_0^1 u^3(1-\ln u)^2 du &= \frac{1}{2} \int_0^1 u^3(1-2\ln u + (\ln u)^2) du = \\
&= \frac{1}{2} \int_0^1 (u^3 - 2u^3 \ln u + u^3(\ln u)^2) du = \frac{1}{2} \left[ \frac{1}{4} + \frac{2}{16} + \frac{1}{32} \right] = \frac{13}{64},
\end{aligned}$$

chega-se:

$$\rho_{S,BIX}(2) = 12 \left( \frac{1}{4} - \frac{4}{9} + \frac{13}{64} \right) = \frac{5}{48}.$$

Finalmente, o valor resultante é:

$$\rho_{S,BIX} = \rho_{S,BIX}(1) + \rho_{S,BIX}(2) = \frac{9}{16} + \frac{5}{48} = \frac{2}{3}. \square$$

### 4.1.3 Tau de Kendall

Uma das medidas de dependência de extrema importância é o Tau de Kendall. A versão amostral desta medida de associação é definida em termos de concordância. Como forma de ilustrar o conceito, seja  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  uma amostra aleatória de tamanho  $n$  do vetor  $(X, Y)$  de variáveis aleatórias contínuas. Quando introduzimos os conceito de concordância, equivale dizer que o par de observações birariadas  $(x_i, y_i)$  e  $(x_j, y_j)$  é concordante se  $x_i < x_j$  e  $y_i < y_j$  ou se  $x_i > x_j$  e  $y_i > y_j$ , da mesma forma dizemos que  $(x_i, y_i)$  e  $(x_j, y_j)$  são discordantes se  $x_i < x_j$  e  $y_i > y_j$  ou  $x_i > x_j$  e  $y_i < y_j$ . É fácil verificar que existem  $\binom{n}{2}$  pares distintos da amostra observada, e definido:

$c$  = número de pares concordantes

$d$  = número de pares discordantes

Por definição, o tau de Kendall amostral é:

$$t_n = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}} \quad (4.2)$$

desta forma,  $t_n$  é a probabilidade de concordância menos a probabilidade de discordância para cada par de observações  $(x_i, y_i)$  e  $(x_j, y_j)$  escolhido aleatoriamente da amostra. A versão populacional para o cálculo do tau de Kendall é dada por:

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \quad (4.3)$$

Logo, como forma de descobrir todas as propriedades da cópula Bi-Extremal, também iremos calcular algebricamente qual é o tau de Kendall para a cópula Bi-Extremal.

**Proposição 4.2** *O  $\tau$  de Kendall: O coeficiente de correlação de Kendall,  $\tau_{S,BIX}$ , da cópula Bi-Extremal é igual a 1/2.*

*Prova:* Do teorema 5.1.3, Nelsen (2006), o  $\tau_{S,BIX}$  é dado por:

$$\tau_{S,BIX} = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \quad (4.4)$$

Agora,

$$\begin{aligned} \tau_{S,BIX} &= \int_0^1 \int_0^{u(1-\ln u)} (v + \psi(v) \ln u) \frac{-1}{u \ln \psi(v)} dv du = \\ &= \int_0^1 \int_0^{u(1-\ln u)} \frac{-v}{u \ln \psi(v)} dv du + \int_0^1 \int_0^{u(1-\ln u)} \frac{-\psi(v) \ln u}{u \ln \psi(v)} dv du. \end{aligned}$$

Como, por definição,  $v = \psi(v)(1 - \ln \psi(v))$ , então:

$$\begin{aligned} &\int_0^1 \int_0^u \frac{-\psi(v)(1 - \ln \psi(v))}{u \ln \psi(v)} (-\ln \psi(v)) d\psi(v) du + \int_0^1 \int_0^u \frac{-\psi(v) \ln u}{u \ln \psi(v)} (-\ln \psi(v)) d\psi(v) du = \\ &= \int_0^1 \frac{1}{u} \left\{ \left[ \frac{\psi(v)^2}{2} (1 - \ln \psi(v)) \right]_0^u + \int_0^u \frac{\psi(v)}{2} d\psi(v) \right\} du + \int_0^1 \frac{\ln u}{u} \left[ \frac{\psi(v)^2}{2} \right]_0^u du = \\ &= \int_0^1 \frac{1}{u} \left[ \frac{u^2}{2} (1 - \ln u) + \frac{u^2}{4} \right] du + \frac{1}{2} \int_0^1 (u \ln u) du = \\ &= \int_0^1 \frac{u(1 - \ln u)}{2} du + \int_0^1 \frac{u}{4} du + \frac{1}{2} \left\{ \left[ \frac{u^2 \ln u}{2} \right]_0^1 - \int_0^1 \frac{u}{2} du \right\} = \end{aligned}$$



$$\begin{aligned}
\frac{1}{2} \left\{ \left[ \frac{u^2(1 - \ln u)}{2} \right]_0^1 + \int_0^1 \frac{u}{2} du \right\} + \frac{1}{4} \left[ \frac{u^2}{2} \right]_0^1 + \frac{1}{2} \left\{ 0 - 0 - \frac{1}{2} \left[ \frac{u^2}{2} \right]_0^1 \right\} = \\
\frac{1}{2} \left\{ \frac{1}{2} - 0 + \frac{1}{2} \int_0^1 u du \right\} + \frac{1}{8} - \frac{1}{8} = \\
\frac{1}{4} \left\{ 1 + \left[ \frac{u^2}{2} \right]_0^1 \right\} = \frac{3}{8}.
\end{aligned}$$

Logo, temos que:

$$\tau_{S,BIX} = 4 \frac{3}{8} - 1 = \frac{1}{2}. \square$$

#### 4.1.4 O coeficiente de Schweizer e Wolff

Em certas situações onde os valores obtidos para as medidas  $\tau$  de Kendall e  $\rho$  de Spearman sejam próximas de zero, é recomendado calcular a medida de dependência  $\sigma$  de Schweizer e Wolff entre duas variáveis aleatórias contínuas. A definição de  $\sigma_C$  encontrada em Schweizer e Wolff (1981) é dada pela seguinte expressão:

$$\sigma_C = \int \int_{I^2} |C(u, v) - uv| dudv.$$

$\sigma_C$  pode ser visto como a norma em  $L_1$  entre uma certa cópula e a cópula produto, além disso seu valor está contido no intervalo  $[0, 1]$ . Uma importante propriedade é:  $\sigma_C = 0$  se e somente se as variáveis correspondentes são mutuamente independentes, ou seja,  $C = \Pi$ .

**Proposição 4.3** *O coeficiente de Schweizer e Wolff: A medida de associação de Schweizer e Wolff,  $\sigma_{BIX}$ , da cópula Bi-extremal é igual a  $1/18$ .*

*Prova:* Agora,  $C_{BIX}(u, v) \geq uv$ , para todo  $(u, v) \in [0, 1]^2$ . Logo,

$$\sigma_{BIX} = \int \int_{I^2} C_{BIX}(u, v) dudv - \int \int_{I^2} uv dudv,$$

como

$$\int \int_{I^2} C_{BIX}(u, v) dudv = \int_0^1 \int_0^{u(1-\log(u))} (v + \psi(v)\log(u)) dv du + \int_0^1 \int_{u(1-\log(u))}^1 uv dv du.$$

Além disso

$$\int_0^1 \int_{u(1-\log(u))}^1 uv dv du = \int_0^1 u[1 - u(1 - \log(u))] du = \frac{1}{18}.$$

Fazendo agora a mudança de variável,

$$v = \psi(v)(1 - \log(\psi(v))) \Rightarrow dv = -\log(\psi(v))d\psi(v),$$

obtemos

$$\int_0^1 \int_0^{u(1-\log(u))} (v + \psi(v)\log(u))dvdu =$$

$$\int_0^1 \int_0^u (\psi(v) - \psi(v)\log(\psi(v)) + \psi(v)\log(\psi(u))(-\log(\psi(v))))d\psi(v) = \frac{1}{4},$$

o que implica

$$\int \int_{I^2} C_{BIX}(u, v)dudv = \frac{11}{36}.$$

Por outro lado, é fácil verificar que

$$\int \int_{I^2} uvdudv = \frac{1}{4},$$

o que resulta  $\sigma_{BIX} = 1/18$  .□

**Observação 4.1** *Este resultado também pode ser obtido através do limite da expressão  $1 - \binom{n}{n-1}^2 / \binom{2n}{2(n-1)}$  quando  $n \rightarrow \infty$ , que é derivado da proposição 9 em Avérous, Genest e Kochar (2005).*

**Observação 4.2** *Segue da Observação 3.2 que o coeficiente de correlação tau de Kendall da distribuição limite da menor e da segunda menor estatística de ordem é igual a 1/2.*

**Observação 4.3** *Note que  $\rho_{S,BIX} > \tau_{S,BIX}$ , o que está de acordo com o resultado (The pair of order statistics is positive likelihood ratio dependent), Nelsen (1992). Fredricks e Nelsen (2007) obtiveram o limite de 3/2, quando  $n \rightarrow \infty$  para a razão  $\rho/\tau$  entre as duas menores estatísticas de ordem. Para este caso, obtemos o valor de 4/3.*

**Observação 4.4** *O coeficiente de Gini da cópula Bi-extremal, que é dado por  $\gamma_{BIX} = 4 \left[ \int_0^1 C_{BIX}(u, 1-u)du - \int_0^1 [u - C_{BIX}(u, u)]du \right]$ , é aproximadamente igual a 2/5. Este resultado foi obtido por integração numérica.*

### 4.1.5 Representação contínua da dependência

Uma outra maneira de visualizar a estrutura de dependência da cópula Bi-extremal é através da função de Spearman, introduzida por Anjos e Kolev, (2005), cuja afirmação é que, para qualquer cópula  $C$  bivariada, existe uma função contínua definida por  $\rho_c$  para todo  $(u, v) \in [0, 1]^2$ , que é dada por:

$$\rho_c(u, v) = \frac{C(u, v) - uv}{\sqrt{uv(1-u)(1-v)}} \quad (4.5)$$

A função  $\rho_c$  permite verificar, para todo par  $(u, v)$ , como é o comportamento de dependência quando comparada com a cópula produto. Para o caso da cópula Bi-extremal a função  $\rho_c$  é:

$$\rho_c(u, v) = \begin{cases} \frac{[v + \phi(v) \ln(u)] - uv}{\sqrt{uv(1-u)(1-v)}}, & \text{se } v \leq u(1 - \log(u)) \\ \frac{u(1-v)}{\sqrt{uv(1-u)(1-v)}}, & \text{se } v > u(1 - \log(u)) \end{cases} \quad (4.6)$$

As figuras (4.1) e (4.2) ilustram a perspectiva e suas curvas de Nível para esta função.

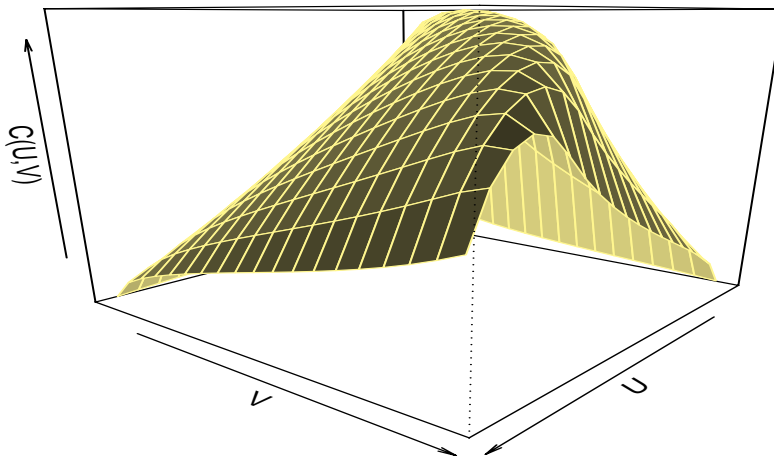


Figura 4.1: *Perspectiva da função de Spearman para cópula Bi-Extremal*

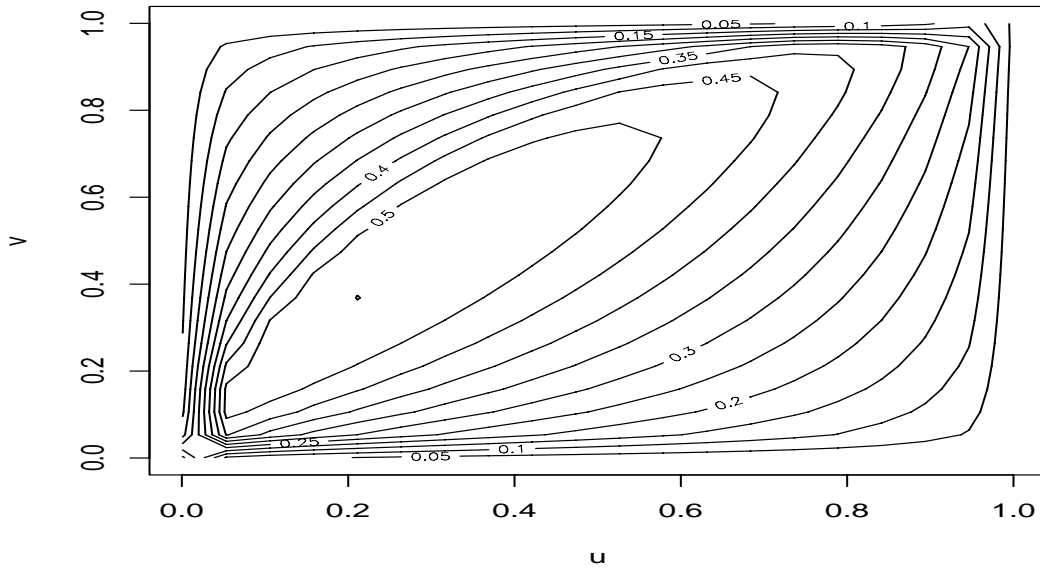


Figura 4.2: *Curvas de Nível da função de Spearman para cópula Bi-Extremal*

Esta função auxilia verificarmos como é o comportamento da dependência entre as variáveis. Para o nosso caso em particular, qual é a relação de dependência entre as duas maiores estatísticas de ordem padronizadas. O que constatamos graficamente é que a maior dependência ocorre quando os valores do par  $(u, v)$ , no quadrado unitário, estão próximos da restrição  $v = u(1 - \log(u))$ . Além disso, podemos validar através das curvas de nível, os valores obtidos para os coeficientes de cauda  $\lambda_U = 0$  e  $\lambda_L = 0$ .

## 4.2 Cópula Condicional

Em muitos casos, é interessante estudar a estrutura de dependência da cópula nas regiões superior ou inferior, ou seja, próximo à região limite  $(1, 1)$  ou  $(0, 0)$ . Este fato torna-se interessante, pois, em muitos casos, esta estrutura limite é completamente diferente do todo. Visando a um melhor entendimento sobre esta estrutura, Charpentier (2004) introduziu o conceito da **LTDC** (*The lower tail dependence copula*), que nada mais é do que a cópula condicional de  $(U, V)$  dada a ocorrência de  $(U \leq u, V \leq v)$ . Além desta nova cópula que reflete a estrutura de dependência condicional, também podemos estudar o limite desta distribuição quando  $u$  e  $v \rightarrow 1$  ou  $0$ , o que permite conhecer exatamente o comportamento das caudas.

Seja  $(U, V)$  o vetor aleatório em  $[0, 1]^2$  com função de distribuição  $C$  e definindo para todo  $(u, v) \in [0, 1]^2$  o evento  $\Xi = \{U \leq u\} \cap \{V \leq v\}$ . Temos que a distribuição condicional de  $(U, V)$  dado  $\Xi$ , denotada por  $F_{C|\Xi}$ , é dada pela expressão:

$$F_{C|\Xi}(x, y) = P(U \leq x, V \leq y | \Xi) = \frac{C(x, y)}{C(u, v)} \quad \text{onde} \quad 0 \leq x \leq u, 0 \leq y \leq v.$$

É importante lembrar que as distribuições marginais de  $U$  e  $V$  dado  $\Xi$  não são uniformes e  $F_{C|\Xi}$  não é uma cópula. As distribuições marginais são dadas por:

$$F_{U|\Xi}(x) = F_{C|\Xi}(x, v) = \frac{C(x, v)}{C(u, v)} \quad \text{e} \quad F_{V|\Xi}(y) = F_{C|\Xi}(u, y) = \frac{C(u, y)}{C(u, v)}$$

A **LTDC**, *The lower tail dependence copula*, relativa a  $C$ , é denotada  $\Phi_{C|\Xi}$ , sendo definida como:

$$\Phi_{C|\Xi}(z, w) = F_{C|\Xi}(F_{U|\Xi}^{-1}(z), F_{V|\Xi}^{-1}(w)) = \frac{C(F_{U|\Xi}^{-1}(z), F_{V|\Xi}^{-1}(w))}{C(u, v)}$$

A definição acima é de extrema importância, pois possibilita encontrar o comportamento limite da estrutura de dependência entre as variáveis. Para o caso da Bi-extremal, esta definição torna-se ainda mais interessante, visto que irá possibilitar entender o comportamento dos valores extremos.

**Teorema 4.2** A LTDC para a cópula Bi-Extremal é dada pela seguinte expressão:

$$\Phi_{C|\Xi}(z, w) = \begin{cases} z.w & F_{V|\Xi}^{-1}(w) \leq F_{U|\Xi}^{-1}(z)(1 - \ln(F_{U|\Xi}^{-1}(z))) \\ z & F_{V|\Xi}^{-1}(w) > F_{U|\Xi}^{-1}(z)(1 - \ln(F_{U|\Xi}^{-1}(z))) \end{cases}.$$

**Prova:** Para o cálculo da LTDC da cópula Bi-Extremal, quando  $v \leq u(1 - \ln u)$ , a distribuição condicional é:

$$F_{C|\Xi}(x, y) = \begin{cases} \frac{y+\psi(y) \ln x}{v+\psi(v) \ln u} & y \leq x(1 - \ln x) \\ \frac{x}{v+\psi(v) \ln u} & y > x(1 - \ln x) \end{cases} \quad (4.7)$$

que tem distribuições marginais:

$$F_{U|\Xi}(x) = \begin{cases} \frac{v+\psi(v) \ln x}{v+\psi(v) \ln u} & v \leq x(1 - \ln x) \\ \frac{x}{v+\psi(v) \ln u} & v > x(1 - \ln x) \end{cases} \quad (4.8)$$

$$F_{V|\Xi}(y) = \begin{cases} \frac{y+\psi(y) \ln u}{v+\psi(v) \ln u} & y \leq u(1 - \ln u) \\ \frac{u}{v+\psi(v) \ln u} & y > u(1 - \ln u) \end{cases} \quad (4.9)$$

Para obter a LTDC da Bi-extremal é necessário, em um primeiro momento, calcular as inversas relativas às distribuições marginais de  $U$  e  $V$ . Inicialmente, vamos calcular para o caso  $y \leq x(1 - \ln x)$ . A partir da distribuição marginal condicional  $F_{U|\Xi}$  em (4.8), temos:

$$z = F_{U|\Xi}(x) \Rightarrow x = F_{U|\Xi}^{-1}(z) = \exp\left(\frac{z(v+\psi(v) \ln u) - v}{\psi(v)}\right). \quad (4.10)$$

Em relação à distribuição  $F_{V|\Xi}$ , constata-se que não há inversa analítica. Entretanto, da expressão (4.9) conclui-se a seguinte igualdade:

$$y + \psi(y) \ln u = (v + \psi(v) \ln u)w,$$

esta vale para todo  $u \in [0, 1]$ . Logo, também vale para  $x = F_{U|\Xi}^{-1}(z)$ , que, por definição, está contido no intervalo  $[0, u]$ . Portanto:

$$y + \psi(y) \ln x = (v + \psi(v) \ln x)w, \quad \forall x \in [0, u] \quad (4.11)$$

substituindo os resultados encontrados em (4.10),(4.11) em (4.7), tem-se:

$$\begin{aligned}\Phi_{C|\Xi}(z, w) &= F_{C|\Xi}(F_{U|\Xi}^{-1}(z), F_{V|\Xi}^{-1}(w)) = \frac{y + \psi(y) \ln \left[ \exp \left( \frac{z(v + \psi(v) \ln u) - v}{\psi(v)} \right) \right]}{v + \psi(v) \ln u} = \\ &= \frac{\left\{ v + \psi(v) \ln \left[ \exp \left( \frac{z(v + \psi(v) \ln u) - v}{\psi(v)} \right) \right] \right\} w}{v + \psi(v) \ln u} = \frac{z(v + \psi(v) \ln u)w}{v + \psi(v) \ln u} = zw\end{aligned}$$

Concluimos, portanto, que a **LTDC** da cópula Bi-extremal é a cópula produto quando  $y \leq x(1 - \ln x)$ . É importante ressaltar que o resultado final não é uma função de  $u$  e  $v$ . Então, para todo par bivariado  $(u, v)$  dado, a **LTDC** é a cópula produto, o que caracteriza que condicionalmente as variáveis são independentes.

Para o caso em que  $y > x(1 - \ln x)$ ,

$$z = F_{U|\Xi}(x) \Rightarrow x = F_{U|\Xi}^{-1}(z) = z(v + \psi(v) \ln u), \quad (4.12)$$

substituindo em (4.7), conclui-se:

$$\Phi_{C|\Xi}(z, w) = z$$

Por outro lado, quando  $v > u(1 - \ln u)$ , a distribuição condicional da cópula Bi-extremal é:

$$F_{C|\Xi}(x, y) = \begin{cases} \frac{y + \psi(y) \ln x}{u} & y \leq x(1 - \ln x) \\ \frac{x}{u} & y > x(1 - \ln x) \end{cases} \quad (4.13)$$

Possuindo as respectivas distribuições marginais:

$$F_{U|\Xi}(x) = \frac{x}{u} \quad (4.14)$$

$$F_{V|\Xi}(y) = \begin{cases} \frac{y + \psi(y) \ln u}{u} & y \leq u(1 - \ln u) \\ 1 & y > u(1 - \ln u) \end{cases} \quad (4.15)$$

Novamente, para obter a **LTDC** da Bi-extremal é necessário, em um primeiro momento, calcular as inversas relativas às distribuições marginais de  $U$  e  $V$ . Inicialmente, vamos calcular para o caso  $y \leq x(1 - \ln x)$ . A partir da distribuição marginal condicional  $F_{U|\Xi}$  em (4.14) temos:

$$z = F_{U|\Xi}(x) \Rightarrow x = F_{U|\Xi}^{-1}(z) = z.u \quad (4.16)$$

Em relação à distribuição  $F_{V|\Xi}$ , constata-se que não há inversa analítica. Entretanto, da expressão (4.15) conclui-se a seguinte igualdade:

$$y + \psi(y) \ln u = w.u$$

Tal igualdade vale para todo  $u \in [0, 1]$ . Logo, também vale para  $x = F_{U|\Xi}^{-1}(z)$ , que, por definição, está contido no intervalo  $[0, u]$ . Portanto:

$$y + \psi(y) \ln x = w.x, \quad \forall x \in [0, u]. \quad (4.17)$$

Substituindo os resultados encontrados em (4.16) e (4.17) em (4.13), tem-se:

$$\begin{aligned} \Phi_{C|\Xi}(z, w) &= F_{C|\Xi}(F_{U|\Xi}^{-1}(z), F_{V|\Xi}^{-1}(w)) = \frac{w.x}{u} = \\ &= \frac{w.u.z}{u} = z.w \end{aligned}$$

Para este caso, a **LTDC** da cópula Bi-extremal também é a cópula produto para  $y \leq x(1 - \ln x)$ .

Para o caso em que  $y > x(1 - \ln x)$ , basta substituir (4.16) em (4.13), de onde se conclui:

$$\Phi_{C|\Xi}(z, w) = z.\square$$

A **LTDC** da cópula Bi-extremal reforça os resultados obtidos para os coeficientes de cauda  $\lambda_U = 0$  e  $\lambda_L = 0$ , estes valores indicam independência entre as variáveis quando  $u$  e  $v \rightarrow 0$  ou 1, visto que o resultado obtido para cópula condicional, quando sujeita à restrição  $F_{V|C}^{-1}(w) \leq F_{U|C}^{-1}(z)(1 - \ln(F_{U|C}^{-1}(z)))$  é a cópula produto, ou seja, que as variáveis são independentes quando  $u$  e  $v$  tendem para 0 ou para 1.



# Capítulo 5

## A cópula K-extremal

### 5.1 Preliminares

Um grande desafio foi encontrar uma generalização da cópula Bi-extremal, ou seja, obter a cópula para as K-maiores estatísticas de ordem para amostras (*i.i.d.*). O primeiro passo imprescindível, foi conhecer a forma exata da função de distribuição das K-maiores estatísticas de ordem para uma amostra (*i.i.d.*) de uma variável aleatória contínua.

**Teorema 5.1** *A função de distribuição  $\tilde{G}_K$  de uma distribuição MGEV K-dimensional, tem a seguinte representação,*

$$\tilde{G}_K(z_1, \dots, z_K) = H_K(z_1, \min(z_1, z_2), \min(z_1, z_2, z_3), \dots, \min(z_1, \dots, z_K)), \quad (5.1)$$

$\forall (z_1, \dots, z_K) \in \mathbb{R}^K$ , onde

$$H_K(z_1, \dots, z_K) = \exp\{-\Lambda(z_K)\} J_K(\Lambda(z_1), \dots, \Lambda(z_K)), \quad (5.2)$$

para  $\min(z_1, \dots, z_K) > \mu - \frac{\sigma}{\xi}$ , se  $\xi > 0$ , ou para  $\min(z_1, \dots, z_K) < \mu - \frac{\sigma}{\xi}$  se  $\xi < 0$ , ou para  $(z_1, \dots, z_K) \in \mathbb{R}^K$  se  $\xi = 0$ , caso contrário  $H_K(z_1, \dots, z_K) = 0$ . A função  $J_K : \mathbb{R}_+^K \rightarrow \mathbb{R}_+$  é um polinômio em K variáveis definido por indução, de tal forma que  $J_1 \equiv 1$  e

$$J_m(x_1, \dots, x_m) = \sum_{j=0}^{m-1} \frac{x_m^j}{j!} - \sum_{j=1}^{m-1} \frac{x_j^j}{j!} J_{m-j}(x_{j+1}, \dots, x_m), \quad \text{para } m \geq 1 \quad (5.3)$$

**Prova:** Inicialmente é necessário provar que  $\tilde{G}_K$  é uma função de distribuição K-dimensional. Para tanto, esta deve satisfazer certas propriedades; um maior detalhamento pode ser encontrado em James B.R. (1981).

**Definição 5.1** (James B.R. (1981)) Uma função  $\mathbf{F}$  definida de  $\mathbb{R}^K \rightarrow \mathbb{R}$  que satisfaça as propriedades P1, P2, P3 e P4 relacionadas abaixo, é chamada **função de distribuição K-dimensional (ou K-variada)**.

P1 -  $F(x_1, \dots, x_k)$  é não decrescente em cada uma das variáveis. Ou seja, se  $r_i < s_i$ , então

$$F(x_1, \dots, r_i, \dots, x_k) \leq F(x_1, \dots, s_i, \dots, x_k) \quad \forall i = 1, \dots, k.$$

P2 -  $F(x_1, \dots, x_k)$  é contínua à direita em cada uma das variáveis. Por exemplo, se  $y_m \downarrow x_1$  quando  $m \rightarrow \infty$ , então

$$F(y_m, x_2, \dots, x_k) \downarrow F(x_1, x_2, \dots, x_k) \quad m \rightarrow \infty.$$

valendo os resultados análogos quando  $y_m \downarrow x_2, y_m \downarrow x_3$ , etc. P3 - Para todo  $i$ ,

$$\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_k) = 0.$$

Também,

$$\lim_{\forall i, x_i \rightarrow +\infty} F(x_1, \dots, x_k) = 1.$$

P4 - Se  $F(x_1, \dots, x_k)$  é uma função de distribuição, então:

$$\Delta_{I_1} \dots \Delta_{I_k} F(x_1, \dots, x_k) \geq 0, \quad \forall I_i = (a_i, b_i], a_i < b_i, i = 1, \dots, k.$$

Por outro lado, é suficiente provar, que a solução da integral múltipla a seguir é igual expressão fornecida em (5.1)

$$\int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_K} \tilde{g}_K(y_1, \dots, y_K) dz_K \dots dz_1,$$

entretanto essa é igual a

$$\int_{-\infty}^{z_1} \int_{-\infty}^{\min(z_1, z_2)} \dots \int_{-\infty}^{\min(z_1, \dots, z_K)} \tilde{g}_K(y_1, \dots, y_K) dz_K \dots dz_1.$$

Portanto  $\tilde{G}_K(z_1, \dots, z_K) = \tilde{G}_K(z_1, \min(z_1, z_2), \min(z_1, z_2, z_3), \dots, \min(z_1, \dots, z_K))$  e supondo em seguida que  $z_1 > z_2 > \dots > z_K$ , então

$$\tilde{G}_K(z_1, \dots, z_K) = (-1)^K \int_{A_\xi}^{z_K} \int_{y_K}^{z_{K-1}} \dots \int_{y_3}^{z_2} \int_{y_2}^{z_1} \exp\{-\Lambda(y_K)\} \prod_{j=1}^K \Lambda'(y_j) dy_1 \dots dy_K.$$

Onde  $A_{\xi=0} = -\infty$ ,  $A_{\xi>0} = \mu - \frac{\sigma}{\xi}$  e  $A_{\xi<0} = -\infty$ . Fazendo a seguinte mudança de variável na última integral,  $x_j = \Lambda(y_j)$ , para  $1 \leq j \leq K$ , obtemos a seguinte integral

$$I_K(w_1, \dots, w_K) := \int_{w_K}^{+\infty} \int_{w_{K-1}}^{z_K} \dots \int_{w_2}^{z_3} \int_{w_1}^{z_2} e^{-x_K} dx_1 \dots dx_K,$$

onde  $w_j = \Lambda(z_j)$ . Vamos provar por indução que

$$I_K(w_1, \dots, w_K) = e^{-w_K} J_K(w_1, \dots, w_K).$$

Para  $K = 1$ , é fácil verificarmos:

$$I_1(w_1) = -1 \int_{w_1}^{\infty} e^{-x_1} dx_1 = e^{-x_1} \Big|_{w_1}^{\infty} = 1.$$

Agora supondo que vale para  $1 \leq K \leq L-1$  então para  $K = L$  temos que  $I_K(w_1, \dots, w_K)$  é igual a

$$(-1)^{K-1} \int_{w_K}^{\infty} \int_{w_{K-1}}^{w_K} \dots \int_{w_2}^{x_3} x_2 e^{-x_K} dx_2 \dots dx_K - w_1 I_{K-1}(w_2, \dots, w_K),$$

que é igual a

$$(-1)^{K-2} \int_{w_K}^{\infty} \int_{w_{K-1}}^{w_K} \dots \int_{w_3}^{x_4} \frac{x_3}{2} e^{-x_K} dx_3 \dots dx_K - \frac{w_2}{2} I_{K-2}(w_3, \dots, w_K) - w_1 I_{K-1}(w_2, \dots, w_K).$$

Integrando recursivamente a expressão anterior

$$I_K(w_1, \dots, w_K) = e^{-w_K} \sum_{j=0}^{L-1} \frac{w_K^j}{j!} - \sum_{j=1}^{L-1} \frac{w_K^j}{j!} I_{L-j}(w_{j+1}, \dots, w_m).$$

Pela definição de  $J_K$  e a hipótese de indução, completa-se a prova.  $\square$

Uma análise de casos particulares pode ser feita, primeiramente fazendo  $K = 2$ , e utilizando a expressão fornecida pelo Teorema 5.1, temos:

$$\tilde{G}_2(z_1, z_2) = \exp[-\Lambda(z_2)] J_2(\Lambda(z_1), \Lambda(z_2)) =$$

$$\begin{aligned}
&= \exp[-\Lambda(z_2)] \left[ \sum_{j=0}^{2-1} \frac{\Lambda(z_2)^j}{j!} - \sum_{j=1}^{2-1} \frac{\Lambda(z_j)^j}{j!} J_{2-j}(\Lambda(z_{j+1}), \dots, \Lambda(z_2)) \right] = \\
&= \exp[-\Lambda(z_2)] [1 + \Lambda(z_2) - \Lambda(z_1)J_1] = \\
&= \exp[-\Lambda(z_2)] [1 + \Lambda(z_2) - \Lambda(z_1)].
\end{aligned}$$

Resultado este que coincide com o obtido na expressão (3.6). Para o caso  $K = 3$ , temos:

$$\begin{aligned}
\tilde{G}_3(z_1, z_2, z_3) &= \exp[-\Lambda(z_3)] J_3(\Lambda(z_1), \Lambda(z_2), \Lambda(z_3)) = \\
&= \exp[-\Lambda(z_3)] \left[ \sum_{j=0}^{3-1} \frac{\Lambda(z_3)^j}{j!} - \sum_{j=1}^{3-1} \frac{\Lambda(z_j)^j}{j!} J_{3-j}(\Lambda(z_{j+1}), \dots, \Lambda(z_3)) \right] = \\
&= \exp[-\Lambda(z_3)] \left[ 1 + \Lambda(z_3) + \frac{\Lambda(z_3)^2}{2!} - \Lambda(z_1)J_2(\Lambda(z_2), \Lambda(z_3)) - \frac{\Lambda(z_2)^2}{2!}J_1 \right] = \\
&= \exp[-\Lambda(z_3)] \left[ 1 + \Lambda(z_3) + \frac{\Lambda(z_3)^2}{2!} - \Lambda(z_1) [1 + \Lambda(z_3) - \Lambda(z_2)J_1] - \frac{\Lambda(z_2)^2}{2!}J_1 \right] = \\
&= \exp[-\Lambda(z_3)] \left[ 1 + \Lambda(z_3) + \frac{\Lambda(z_3)^2}{2!} - \Lambda(z_1) - \frac{\Lambda(z_2)^2}{2!} - \Lambda(z_1)\Lambda(z_3) + \Lambda(z_1)\Lambda(z_2) \right] =
\end{aligned}$$

Mas este resultado também pode ser obtido por métodos de integração, da mesma forma que no caso  $K = 2$ , ou seja:

$$\tilde{G}_3(z_1, z_2, z_3) = P(M_1 > z_1, M_2 > z_2, M_3 > z_3)$$

Sabendo que:

$$\begin{aligned}
&G_3(z_1, z_2, z_3) = \\
&= 1 - [\bar{G}_1(z_1) + \bar{G}_2(z_2) + \bar{G}_3(z_3) - \bar{G}_{12}(z_1, z_2) - \bar{G}_{13}(z_1, z_3) - \bar{G}_{23}(z_2, z_3) + \bar{G}_3(z_1, z_2, z_3)]
\end{aligned}$$

Calculando inicialmente:

$$\begin{aligned}
\tilde{G}_3(z_1, z_2, z_3) &= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} \int_{z_3}^{z_2} \exp[-\Lambda(z_3)] [-\Lambda'(z_1)][-\Lambda'(z_2)][-\Lambda'(z_3)] dz_3 dz_2 dz_1 = \\
&= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} \int_{z_3}^{z_2} \exp[-\Lambda(z_3)] d[-\Lambda(z_3)] d[-\Lambda(z_2)] d[-\Lambda(z_1)] = \\
&= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} [\exp[-\Lambda(z_3)]]_{z_3}^{z_2} d[-\Lambda(z_2)] d[-\Lambda(z_1)] = \\
&= \int_{z_1}^{+\infty} \int_{z_2}^{z_1} \{ \exp[-\Lambda(z_2)] - \exp[-\Lambda(z_3)] \} d[-\Lambda(z_2)] d[-\Lambda(z_1)] =
\end{aligned}$$

$$\begin{aligned}
&= \int_{z_1}^{+\infty} \{ \{ \exp[-\Lambda(z_2)] \}_{z_2}^{z_1} - \exp[-\Lambda(z_3)] \{ -\Lambda(z_2) \}_{z_2}^{z_1} \} d[-\Lambda(z_1)] = \\
&= \int_{z_1}^{+\infty} \{ \exp[-\Lambda(z_1)] - \exp[-\Lambda(z_2)] - [-\Lambda(z_1)] \exp[-\Lambda(z_3)] + \\
&\quad + [-\Lambda(z_2)] \exp[-\Lambda(z_3)] \} d[-\Lambda(z_1)] = \\
&= \{ \exp[-\Lambda(z_1)] \}_{z_1}^{+\infty} - \exp[-\Lambda(z_2)] \{ -\Lambda(z_1) \}_{z_1}^{+\infty} - \exp[-\Lambda(z_3)] \left\{ \frac{[-\Lambda(z_1)]^2}{2} \right\}_{z_1}^{+\infty} + \\
&\quad + [-\Lambda(z_2)] \exp[-\Lambda(z_3)] \{ -\Lambda(z_1) \}_{z_1}^{+\infty} = \\
&= 1 - \exp[-\Lambda(z_1)] + [-\Lambda(z_1)] \exp[-\Lambda(z_2)] + \\
&\quad + \frac{[-\Lambda(z_1)]^2}{2} \exp[-\Lambda(z_3)] - [-\Lambda(z_1)] [-\Lambda(z_2)] \exp[-\Lambda(z_3)]
\end{aligned}$$

O próximo passo será calcular todas as distribuições marginais, que são:

$$\begin{aligned}
\tilde{G}_{12}(z_1, z_2) &= \tilde{G}_3(z_1, z_2, -\infty) = \\
&= 1 - \exp[-\Lambda(z_1)] + [-\Lambda(z_1)] \exp[-\Lambda(z_2)] \\
\tilde{G}_{13}(z_1, z_3) &= \tilde{G}_3(z_1, z_3, z_3) = \\
&= 1 - \exp[-\Lambda(z_1)] + [-\Lambda(z_1)] \exp[-\Lambda(z_3)] + \\
&\quad + \frac{[-\Lambda(z_1)]^2}{2} \exp[-\Lambda(z_3)] - [-\Lambda(z_1)] [-\Lambda(z_3)] \exp[-\Lambda(z_3)] \\
\tilde{G}_{23}(z_2, z_3) &= \tilde{G}_3(z_2, z_2, z_3) = \\
&= 1 - \exp[-\Lambda(z_2)] + [-\Lambda(z_2)] \exp[-\Lambda(z_2)] + \frac{[-\Lambda(z_2)]^2}{2} \exp[-\Lambda(z_3)] - [-\Lambda(z_2)]^2 \exp[-\Lambda(z_3)] \\
\tilde{G}_1(z_1) &= \tilde{G}_{12}(z_1, -\infty) = 1 - \exp[-\Lambda(z_1)] \\
\tilde{G}_2(z_2) &= \tilde{G}_{12}(z_2, z_2) = 1 - \exp[-\Lambda(z_2)] + [-\Lambda(z_2)] \exp[-\Lambda(z_2)] \\
\tilde{G}_3(z_3) &= \tilde{G}_{23}(z_3, z_3) = \\
&= 1 - \exp[-\Lambda(z_3)] + [-\Lambda(z_3)] \exp[-\Lambda(z_3)] - \frac{[-\Lambda(z_3)]^2}{2} \exp[-\Lambda(z_3)]
\end{aligned}$$

com isso, conclui-se que:

$$\begin{aligned}
&\tilde{G}_3(z_1, z_2, z_3) = \\
&\exp[-\Lambda(z_3)] \left[ 1 + \Lambda(z_3) + \frac{\Lambda(z_3)^2}{2} - \Lambda(z_1) - \frac{\Lambda(z_2)^2}{2} - \Lambda(z_1)\Lambda(z_3) + \Lambda(z_1)\Lambda(z_2) \right].
\end{aligned}$$

**Observação 5.1** Se  $\tilde{G}_K$  é a função de distribuição limite das  $K$ -maiores estatísticas de ordem para uma amostra (i.i.d) de uma variável aleatória contínua, então  $\tilde{G}_K$  tem como função de densidade  $\tilde{g}_K$  fornecida em (2.17).

## 5.2 Especificação

Nesta seção a densidade da cópula  $K$ -extremal será obtida utilizando-se a expressão encontrada em (2.3), além disso, os resultados da seção anterior serão empregados para se obter a cópula  $K$ -extremal. Como primeiro passo vamos definir uma função implícita que será de extrema importância na determinação dos resultados.

**Definição 5.2** *Seja  $\psi_m : (0, 1) \rightarrow (0, 1)$  definida como uma função crescente que satisfaz a seguinte equação implícita*

$$u = \psi_m(u) \sum_{j=0}^{m-1} (-1)^j \frac{(\log \psi_m(u))^j}{j!} \quad (5.4)$$

em particular,  $\psi_1(u) = u$ . Esta definição é consequência direta da fórmula explícita para a função de distribuição  $G_m$  fornecida em (2.11). Além disso, se derivarmos os dois lados de (5.4), iremos obter os seguintes resultados:

$$1 = \left( \sum_{j=0}^{m-1} (-1)^j \frac{(\log \psi_m)^j}{j!} - \sum_{j=0}^{m-2} (-1)^j \frac{(\log \psi_m)^j}{j!} \right) \frac{d\psi_m}{du} = (-1)^{m-1} \frac{(\log \psi_m)^{m-1}}{(m-1)!} \frac{d\psi_m}{du},$$

isso implica

$$\frac{d\psi_m}{du} = (-1)^{m-1} \left( \frac{(\log \psi_m)^{m-1}}{(m-1)!} \right)^{-1} \quad e \quad \frac{d \log \psi_m}{du} = (-1)^{m-1} \left( \psi_m \frac{(\log \psi_m)^{m-1}}{(m-1)!} \right)^{-1}.$$

Um resultado obtido foi conhecer qual a convergência da função  $\psi_K^{-1}(u)$  quando  $K \rightarrow 1$ .

**Proposição 5.1** *A função  $\psi_K^{-1}(u) : (0, 1) \rightarrow (0, 1)$  converge para 1 exponencialmente rápido, quando  $K \rightarrow \infty$ .*

**PROVA:** Temos que a função  $\psi_K^{-1}(u) : (0, 1) \rightarrow (0, 1)$ , é:

$$u = \psi_K^{-1}(u) = u \sum_{j=0}^{K-1} (-1)^j \frac{(\log u)^j}{j!}. \quad (5.5)$$

Entretanto sabemos que:

$$u \sum_{j=0}^{+\infty} (-1)^j \frac{(\log u)^j}{j!} = u \cdot e^{\log(\frac{1}{u})} = 1,$$

portanto:

$$|\psi_K^{-1}(u) - 1| = u \sum_{j=K}^{+\infty} (-1)^j \frac{(\log u)^j}{j!},$$

além disso, existe  $\lambda > 0$ , tal que  $u = e^{-\lambda}$ , com isso,

$$u \sum_{j=K}^{+\infty} (-1)^j \frac{(\log u)^j}{j!} = e^{-\lambda} \sum_{j=K}^{+\infty} \frac{\lambda^j}{j!} \leq \frac{\lambda^K}{K!},$$

utilizando-se a aproximação de stirling, onde  $K! \approx \sqrt{2\pi K} \cdot K^K \cdot e^{-K}$ , temos:

$$\frac{\lambda^K}{K!} \approx \frac{(\lambda \cdot e)^K}{\sqrt{2\pi K} K^{K+\frac{1}{2}}},$$

fazendo  $\lambda = K/(2e)$

$$= C \cdot e^{-K \log 2} \leq e^{-K \log 2},$$

onde  $C$  é uma constante igual a  $1/\sqrt{2\pi \cdot K}$ , tal que  $C \leq 1$ . Logo para  $0 < \lambda \leq \frac{K}{2e} \Rightarrow u \in (e^{-\frac{K}{2e}}, 1)$ , daí para qualquer valor de  $u$  pertencente a este intervalo:

$$|\psi_K^{-1}(u) - 1| \leq e^{-K \log 2}. \square \quad (5.6)$$

**Teorema 5.2** *A densidade da cópula de uma distribuição GEV  $K$ -dimensional, e dada pela expressão*

$$c_K(u_1, \dots, u_K) = \left( \prod_{j=1}^{K-1} \frac{d \log(\psi_j(u_j))}{du_j} \right) \frac{d\psi_K(u_K)}{du_K} \quad (5.7)$$

$$= \left( \prod_{j=1}^{K-1} (-1)^{j-1} \psi_j(u_j) \frac{(\log \psi_j(u_j))^{j-1}}{(j-1)!} \right)^{-1} \left( \frac{(-\log \psi_K(u_K))^{K-1}}{(K-1)!} \right)^{-1} \quad (5.8)$$

para  $(u_1, \dots, u_K) \in (0, 1)^K$ , tal que  $u_1 > \psi_2(u_2) > \dots > \psi_K(u_K)$  onde  $\psi_m : (0, 1) \rightarrow (0, 1)$  é uma função crescente definida em (5.4).

**Prova:** Sabe-se que

$$c(u_1, \dots, u_K) = \frac{\tilde{g}_K(G_1^{-1}(u_1), \dots, G_K^{-1}(u_K))}{\prod_{j=1}^K g_j^{-1}(G_j^{-1}(u_j))}.$$

Entretanto, aplicando as fórmulas (2.11) e (2.17) obtemos

$$c(u_1, \dots, u_K) = \left( \prod_{j=1}^{K-1} \exp\{-\Lambda(G_j^{-1}(u_j))\} \frac{\Lambda(G_j^{-1}(u_j))^{j-1}}{(j-1)!} \right)^{-1} \left( \frac{\Lambda(G_K^{-1}(u_K))^{K-1}}{(K-1)!} \right)^{-1}$$

## SIMULAÇÃO

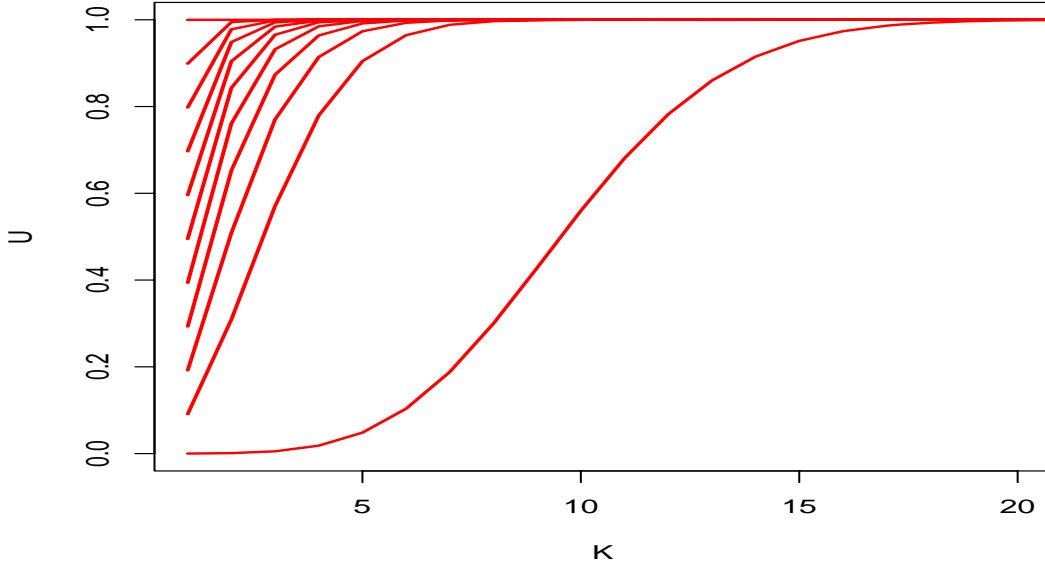


Figura 5.1: Gráfico ilustrativo da convergência de  $\psi_K^{-1}(u)$  quando  $K \rightarrow \infty$

Da definição de  $\psi_m$  encontrada em (5.4), obtemos (5.7) de imediato. E, ainda mais, utilizando as expressões das derivadas de  $\psi_m$  e  $\log \psi_m$  obtemos (5.8).  $\square$

**Observação 5.2** A função  $\psi_m$  que aparece na expressão da densidade da cópula  $K$ -extremal pode ser calculada explicitamente através da função de distribuição MGEV como  $\psi_m(u) = \exp\{-\Lambda(G_m^{-1}(u))\}$  para todo  $u \in (0, 1)$  e  $m \geq 1$ .

Utilizando todos os resultados até aqui vistos, iremos, no próximo teorema, obter a cópula  $K$ -extremal.

**Teorema 5.3** A cópula de uma distribuição GEV  $K$ -dimensional, é dada pela expressão

$$C_K(u_1, \dots, u_K) = \mathcal{H}_K(u_1, r_1(u_1, u_2), r_2(u_1, u_2, u_3), \dots, r_{K-1}(u_1, \dots, u_K)), \quad (5.9)$$

para todo  $(u_1, \dots, u_K) \in [0, 1]^K$ , onde

$$r_{m-1}(u_1, \dots, u_m) = \psi_m^{-1}(\psi_l(u_l)) = \psi_l(u_l) \sum_{j=0}^{m-1} (-1)^j \frac{(\log \psi_l(u_l))^j}{j!},$$



se  $\psi_l(u_l) = \min(\psi_1(u_1), \dots, \psi_m(u_m))$  e para todo  $(u_1, \dots, u_K)$  desde que  $u_1 = \psi_1(u_1) \geq \psi_2(u_2) \geq \dots \geq \psi_K(u_K)$

$$\begin{aligned} \mathcal{H}_K(u_1, \dots, u_K) &= \psi_K(u_K) J_K(-\log u_1, -\log \psi_2(u_2), \dots, -\log \psi_K(u_K)), \\ &= u_K - \psi_K(u_K) \sum_{j=1}^{K-1} \frac{(-\log \psi_j(u_j))^j}{j!} J_{K-j}(-\log \psi_{j+1}(u_{j+1}), \dots, -\log \psi_K(u_K)) \end{aligned}$$

com  $J_m$  definido previamente no teorema 5.1

**Prova:** Seja  $\tilde{G}_K$  a função de distribuição limite de valores extremos. Então a função de distribuição da cópula K-extremal é dada por

$$C_K(u_1, \dots, u_K) = \tilde{G}_K(G_1^{-1}(u_1), \dots, G_K^{-1}(u_K))$$

para todo  $(u_1, \dots, u_K) \in [0, 1]^K$  que pelo teorema 5.1 é igual a

$$H_K(G_1^{-1}(u_1), \min(G_1^{-1}(u_1), G_2^{-1}(u_2)), \dots, \min(G_1^{-1}(u_1), \dots, G_K^{-1}(u_K))).$$

Pela definição de  $H_K$ , monotonicidade e a expressão para  $\psi_m$  na definição (5.2), chega-se

$$\min_{1 \leq l \leq K} (\psi_l(u_l)) J_K \left( -\log u_1, -\log \min_{l=1,2} (\psi_l(u_l)), \dots, -\log \min_{1 \leq l \leq K} (\psi_l(u_l)) \right).$$

Usando a definição de  $r_m$ , podemos escrever

$$\psi_K(r_K(u_1, \dots, u_m)) J_K(-\log u_1, -\log \psi_2(r_2(u_1, u_2)), \dots, -\log \psi_K(r_K(u_1, \dots, u_m))),$$

com isso completamos a prova.  $\square$

### 5.3 Convergência em Distribuição

Nesta seção nosso objetivo é verificar qual é o comportamento limite da cópula das K-maiores estatísticas de ordem, para qualquer seqüência de variáveis aleatórias contínuas (*i.i.d.*). O teorema seguinte é um resultado de convergência, no qual prova-se que, para uma grande classe de distribuições absolutamente contínuas, a estrutura de dependência não-linear para as K-maiores estatísticas de ordem de uma amostra (*i.i.d.*) é aproximadamente capturada pela cópula K-extremal. Por uma simples generalização do Lema

6 encontrado em Avérous, J., Genest, C., Kochar, S. (2005), tem-se que a cópula multivariada para as K-maiores estatísticas de ordem para uma amostra (i.i.d.) não depende da distribuição contínua primitiva da amostra. Esta cópula será denotada por  $\tilde{C}_K^{(n)}$ , onde  $n$  denota o tamanho da amostra.

**Teorema 5.4** *A cópula  $\tilde{C}_K^{(n)}$  converge em distribuição para  $C_K$  quando  $n \rightarrow \infty$ .*

**Prova:**

Seja  $M_{1,n}, \dots, M_{K,n}$  uma amostra de tamanho  $n$  das K-maiores estatísticas de ordem, de uma certa função de distribuição  $F$ , que pertence ao domínio de atração da distribuição GEV. Isto significa que existem seqüências de números reais  $(a_n)_{n=1}^{+\infty}$  and  $(b_n)_{n=1}^{+\infty}$  tais que

$$\left( \frac{M_{1,n} - b_n}{a_n}, \dots, \frac{M_{K,n} - b_n}{a_n} \right)$$

converge em distribuição para  $\tilde{G}_K$ , que é a distribuição MGEV. Pelo princípio de invariância relativa, a cópula associada a  $(M_{1,n}, \dots, M_{K,n})$  e  $((M_{1,n} - b_n)/a_n, \dots, (M_{K,n} - b_n)/a_n)$  é  $\tilde{C}_K^{(n)}$ , e  $\tilde{C}_K^{(n)}$  independe de  $F$ .

Seja  $F_{n,j}$  a função de distribuição de  $(a_n M_{n,j} + b_n)$ . Portanto, se definirmos a função

$$V_n(x_1, \dots, x_K) = (F_{1,n}(x_1), \dots, F_{K,n}(x_K)) \quad (x_1, \dots, x_K) \in \mathbb{R}^n,$$

então

$$V_n \left( \frac{M_{1,n} - b_n}{a_n}, \dots, \frac{M_{K,n} - b_n}{a_n} \right), \quad (5.10)$$

tem distribuição à cópula  $\tilde{C}_K^{(n)}$ . A cópula K-extremal tem a distribuição de  $V(Y_1, \dots, Y_K)$ , onde

$$V(x_1, \dots, x_K) = (G_1(x_1), \dots, G_K(x_K)) \quad (x_1, \dots, x_K) \in \mathbb{R}^n.$$

Pelo teorema 5.1 página 30, encontrado em Billingsley, P. (1968), a expressão (5.10) converge em distribuição para a cópula K-extremal, se  $V_n$  converge uniformemente para  $V$  em intervalos compactos. Mas isto é uma consequência do Teorema de Polya encontrado em Magalhães, M. N. (2006), que implica convergência uniforme de  $F_{n,j}$  para  $G_j$  desde que esta seja absolutamente contínua.  $\square$

**Observação 5.3** *Através da cópula K-extremal, pode-se obter a cópula limite entre a l-ésima maior e a m-ésima maior estatísticas de ordem para qualquer escolha de l e m.*

Pode-se utilizar as cópulas bivariadas mencionadas na observação (5.3), para obter as medidas de dependência rho de Spearman e tau de Kendall. Para uma cópula bivariada  $C$  qualquer, o coeficiente rho de Spearman e tau de Kendall são calculados respectivamente pelas expressões (4.1) e (4.4). Um estudo sobre a convergência destas medidas, entre a primeira e a  $K$ -ésima estatística de ordem quando  $K \rightarrow \infty$  foi feito, iremos denotar estas medidas respectivamente por  $\rho_K$  e  $\tau_K$ ,  $K \geq 2$ . Como já demonstrado anteriormente, temos que  $\rho_2 = 2/3$  e  $\tau_2 = 1/2$ . Além disso, para maiores detalhes sobre estas medidas de dependência, veja Avérous, J., Genest, C., Kochar, S. (2005) e Chen, Y. (2007). Com isso temos o seguinte teorema de convergência.

**Teorema 5.5** *Ambas as seqüências  $(\rho_K)$  e  $(\tau_K)$  convergem para zero quando  $K \rightarrow \infty$ .*

Iremos provar, através da estimativa da expressão exata, que  $(\rho_K) \rightarrow 0$ . Por analogia, o resultado também pode ser aplicado à  $(\tau_K)$ , dado que  $\rho_K \geq \tau_K \geq 0$ , este resultado pode ser verificado através do teorema 5.1 em Fredricks e Nelsen (2007).

Aplicando diretamente a definição podemos reescrever  $(\rho_K + 3)/12$  como

$$\int_0^1 \int_{\psi_{K-1}^{-1}(\psi_K(u_K))}^1 \dots \int_{\psi_2^{-1}(\psi_3(u_3))}^1 \int_{\psi_2(u_2)}^1 u_1 u_K c_K(u_1, \dots, u_K) du_1 \dots du_K. \quad (5.11)$$

Iremos provar que esta converge para 1/4 quando  $K \rightarrow \infty$  resultando em  $\rho_K \rightarrow 0$ .

Utilizando a expressão (5.7), a integral anterior pode ser escrita como

$$\int_0^1 \int_{\psi_{K-1}^{-1}(\psi_K(u_K))}^1 \dots \int_{\psi_2^{-1}(\psi_3(u_3))}^1 \int_{\psi_2(u_2)}^1 u_1 u_K \left( \prod_{j=1}^{K-1} \frac{d \log(\psi_j(u_j))}{du_j} \right) \frac{d\psi_K(u_K)}{du_K} du_1 \dots du_K.$$

Por indução em  $1 \leq m \leq K - 1$ , provaremos que

$$\int_{\psi_m^{-1}(\psi_{m+1}(u_{m+1}))}^1 \dots \int_{\psi_2^{-1}(\psi_3(u_3))}^1 \int_{\psi_2(u_2)}^1 u_1 \prod_{j=1}^m \frac{d \log(\psi_j(u_j))}{du_j} du_1 \dots du_m.$$

é igual a

$$(-1)^m \left[ \psi_{m+1}(u_{m+1}) - \sum_{j=0}^{m-1} \frac{(\log \psi_{m+1}(u_{m+1}))^j}{j!} \right]. \quad (5.12)$$

Na verdade,  $\psi_1$  é a função identidade em  $(0, 1)$ , e portanto

$$\int_{\psi_2(u_2)}^1 u_1 \frac{d \log(\psi_1(u_1))}{du_1} du_1 = \int_{\psi_2(u_2)}^1 u_1 \frac{1}{u_1} du_1 = (-1)[\psi_2(u_2) - 1].$$

Agora supondo 5.12 verdadeira para algum  $1 \leq m = l \leq K - 2$  então

$$(-1)^l \left[ \psi_{l+1}(u_{l+1}) - \sum_{j=0}^{l-1} \frac{(\log \psi_{l+1}(u_{l+1}))^j}{j!} \right] \frac{d \log(\psi_{l+1}(u_{l+1}))}{du_{l+1}},$$

esta é igual a

$$(-1)^l \frac{d}{du_{l+1}} \left( \psi_{l+1}(u_{l+1}) - \sum_{j=1}^l \frac{(\log \psi_{l+1}(u_{l+1}))^j}{j!} \right)$$

como  $\psi_{l+1}(1) = 1$ , e integrando em  $u_{l+1}$ , no intervalo  $(\psi_{l+1}^{-1}(\psi_{l+2}(u_{l+2})), 1)$ , obtemos (5.11)

válida para  $m = l + 1$ . Entretanto a integral (5.11) é igual a

$$\int_0^1 u \frac{d\psi_K}{du}(u) (-1)^{K-1} \left[ \psi_K(u) - \sum_{j=0}^{K-2} \frac{(\log \psi_K(u))^j}{j!} \right] du.$$

Fazendo  $v = \psi_K(u)$ ,  $u \in (0, 1)$  e usando a expansão em séries de potência

$$v = \sum_{j=0}^{\infty} \frac{[\log(v)]^j}{j!}$$

logo, podemos reescrever a integral anterior da seguinte forma

$$(-1)^{K-1} \int_0^1 \psi_K^{-1}(v) \left( \sum_{K-1}^{\infty} \frac{[\log(v)]^j}{j!} \right) dv$$

Fazendo uma outra mudança de variável e com a expressão (5.4), podemos escrever a integral em (5.11) como

$$\begin{aligned} & (-1)^{K-1} \sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} \frac{(-1)^{l-1}}{j! l!} \int_0^{+\infty} y^{l+j} e^{-2y} dy = \\ & = (-1)^{K-1} \sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} (-1)^{l-1} \binom{l+j}{l} \frac{1}{2^{l+j+1}} = \frac{1}{2} \cdot (-1)^{K-1} \sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} (-1)^{l-1} \binom{l+j}{l} \frac{1}{2^{l+j}} \end{aligned}$$

onde

$$\int_0^{+\infty} y^{l+j} e^{-2y} dy = \frac{(l+j)!}{2^{l+j+1}}.$$

Terminamos a prova mostrando que

$$(-1)^{K-1} \sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} (-1)^{l-1} \binom{l+j}{l} \frac{1}{2^{l+j}} \rightarrow \frac{1}{2}$$

A partir deste ponto supomos  $K$  ímpar (para  $K$  par a prova é semelhante com poucas mudanças sinal). O lado esquerdo da expressão anterior de convergência é igual a

$$\sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} \binom{l+j}{l} \frac{1}{2^{l+j}} - \sum_{l=0}^{K-1} \sum_{j=\frac{K-1}{2}}^{\infty} \binom{l+2j+1}{l} \frac{1}{2^{l+2j}}, \quad (5.13)$$

Agora aplicando as seguintes identidades

$$\binom{l+2j}{0} = 1 \quad \text{e} \quad \binom{l+2j+1}{l} = \binom{l+2j}{l-1} + \binom{l+2j}{l}, \quad \text{para } l \geq 1,$$

para escrever o segundo termo de (5.13) como

$$\sum_{j=K-1}^{\infty} \binom{2j+1}{K-1} \frac{1}{2^{2j+1}} - \sum_{l=0}^{K-1} \sum_{j=K-1}^{\infty} \binom{l+j}{l} \frac{1}{2^{l+j}}.$$

Portanto (5.13) é igual a

$$\sum_{j=K-1}^{\infty} \binom{2j+1}{K-1} \frac{1}{2^{2j+1}}$$

o que é igual a

$$\sum_{j=2K}^{\infty} \binom{j-1}{K-1} \frac{1}{2^j} + \sum_{j=K-1}^{\infty} \binom{2j+1}{K-1} \left(1 - \frac{2j+2}{2(2j-K+3)}\right) \frac{1}{2^{2(j+2)}}.$$

Mas

$$\binom{2j+1}{K-1} \frac{1}{2^{2(j+2)}}$$

são probabilidades oriundas de uma distribuição binomial negativa, logo o segundo elemento na soma acima, é limitado por

$$\left(1 - \frac{2K}{2K+2}\right).$$

Por conseguinte, o limite de (5.13) quando  $K \rightarrow \infty$  é o mesmo que o limite da expressão

$$\sum_{j=2K}^{\infty} \binom{j-1}{K-1} \frac{1}{2^j}$$

que é a probabilidade de uma distribuição binomial negativa com parâmetros  $K$  e  $1/2$ , para valores maiores ou iguais a  $2K$ . Esta probabilidade converge para  $1/2$  pelo Teorema Central do Limite. □

## 5.4 Algoritmo para Simulação

Igualmente como visto na seção 3.2 o método para simular valores oriundos de uma certa cópula multivariada é baseado na técnica de amostragem condicional ver Cherubini, U., Luciano, E., Vecchiato, W. (2004). Vamos focar agora para o caso em que desejamos gerar valores oriundos de uma cópula  $K$ -dimensional. Seja  $U_1, U_2, \dots, U_K$  uniformemente distribuídos e tendo  $C$  como sendo sua função de distribuição, então a distribuição condicional de  $U_i$  dado os valores de  $U_1, U_2, \dots, U_{i-1}$ , é dado por

$$\begin{aligned} C_i(u_i|u_1, \dots, u_{i-1}) &= Pr(U_i \leq u_i | U_1 = u_1, \dots, U_{i-1} = u_{i-1}) \\ &= \left( \frac{[\partial^{i-1} C_i(u_1, \dots, u_i)] / [\partial u_1, \dots, \partial u_{i-1}]}{[\partial^{i-1} C_{i-1}(u_1, \dots, u_{i-1})] / [\partial u_1, \dots, \partial u_{i-1}]} \right) \end{aligned} \quad (5.14)$$

com  $i = 2, \dots, K$ . Obviamente assumimos que ambos, numerador e denominador, existem e este não é zero, desta forma o algoritmo para a simulação pode ser reescrito da seguinte forma

- (i) Defina  $C_i(u_1, u_2, \dots, u_i) = C(u_1, u_2, \dots, u_i, 1, \dots, 1)$  para  $i = 1, \dots, K$ ;
- (ii) Simule um valor aleatório  $u_1$  oriundo de uma  $U(0, 1)$ ;
- (iii) E generalizando, gere um valor aleatório  $u_i$  oriundo de  $C_i(\cdot | u_1, \dots, u_{i-1})$  para  $i = 1, \dots, K$ .

Logo, para se obter o valor de  $u_i$  através de  $C_i(\cdot | u_1, \dots, u_{i-1})$ , basta gerar um valor  $q$  oriundo de uma  $U(0, 1)$ , e definir  $u_i = C_i^{-1}(q | u_1, \dots, u_{i-1})$  que é obtido da equação  $q = C_i(u_i | u_1, \dots, u_{i-1})$ . Para que possa gerar valores aleatórios oriundos da cópula  $K$ -extremal, torna-se necessário conhecermos qual é a distribuição de  $(U_i | U_1, U_2, \dots, U_{i-1})$ .

**Teorema 5.6** *A distribuição condicional  $(U_i | U_1, U_2, \dots, U_{i-1})$  para a cópula  $K$ -extremal é dada pela seguinte expressão.*

$$C_i(u_i | u_1, \dots, u_{i-1}) = \frac{\psi_i(u_i)}{\psi_{i-1}(u_{i-1})}. \quad (5.15)$$

**Prova:** Calculando inicialmente o numerador de (5.14), tem-se,

$$\begin{aligned} & \frac{\partial^{i-1} C_i(u_1, \dots, u_i)}{\partial u_1, \dots, \partial u_{i-1}} \\ = & \frac{\partial^{i-1} \left[ -\psi_i(u_i) \sum_{j=1}^{i-1} \frac{-\log(\psi_j(u_j))^j}{j!} J_{i-j}(-\log \psi_{j+1}(u_{j+1}), \dots, -\log \psi_i(u_i)) \right]}{\partial u_1 \dots \partial u_{i-1}}. \end{aligned} \quad (5.16)$$

Eliminando os termos que não dependem de todas as variáveis  $u_1, \dots, u_{i-1}$ , conclui-se

$$\frac{\partial^{i-1} C_i(u_1, \dots, u_i)}{\partial u_1, \dots, \partial u_{i-1}} = \frac{\partial^{i-1} \left[ -\psi_i(u_i) \prod_{j=1}^{i-1} (-\log(\psi(u_j))) \right]}{\partial u_1 \dots \partial u_{i-1}} =$$

mas como

$$\frac{d \log \psi_m}{du} = (-1)^{m-1} \left( \psi_m \frac{(\log \psi_m)^{m-1}}{(m-1)!} \right)^{-1}$$

conclui-se

$$= (-1)^i \psi_i(u_i) (-1)^{i-1} \prod_{j=1}^{i-1} (-1)^{j-1} \left( \psi_j(u_j) \frac{\log(\psi_j(u_j))^{j-1}}{(j-1)!} \right)^{-1}. \quad (5.17)$$

Já o denominador em (5.14), nada mais é do que a função de densidade da cópula K-extremal para  $K = i - 1$ , logo:

$$\begin{aligned} & \frac{\partial^{i-1} C_{i-1}(u_1, \dots, u_{i-1})}{\partial u_1, \dots, \partial u_{i-1}} \\ & \left( \prod_{j=1}^{i-2} (-1)^{j-1} \psi_j(u_j) \frac{(\log \psi_j(u_j))^{j-1}}{(j-1)!} \right)^{-1} \left( -\frac{(\log \psi_{i-1}(u_{i-1}))^{i-2}}{(i-2)!} \right)^{-1} \end{aligned} \quad (5.18)$$

finalmente, substituindo os valores de (5.17) e (5.18) em (5.14), obtemos:

$$C_i(u_i | u_1, \dots, u_{i-1}) = \frac{\psi_i(u_i)}{\psi_{i-1}(u_{i-1})}.$$

□

Fazendo  $q = C_i(u_i | u_1, \dots, u_{i-1})$ , temos:

$$u_i = C_i^{-1}(q | u_1, \dots, u_{i-1}) = \psi_i^{-1}(q \cdot \psi_{i-1}(u_{i-1}))$$

mas pela definição 5.2, conclui-se:

$$u_i = \psi_i(q, \psi_{i-1}(u_{i-1})) \sum_{j=0}^{i-1} (-1)^j \frac{(\log \psi_i(q, \psi_{i-1}(u_{i-1})))^j}{j!}.$$

Este último resultado implica que inúmeras simulações podem ser feitas, tanto de forma a possibilitar um maior conhecimento sobre a estrutura de dependência da cópula K-extremal, como também para auxiliar em futuras modelagens de dados reais. A figura (5.2) ilustra gráficos de cruzamento entre as variáveis uniformizadas, a partir de dados simulados da cópula K-extremal ( $K = 4$ ).

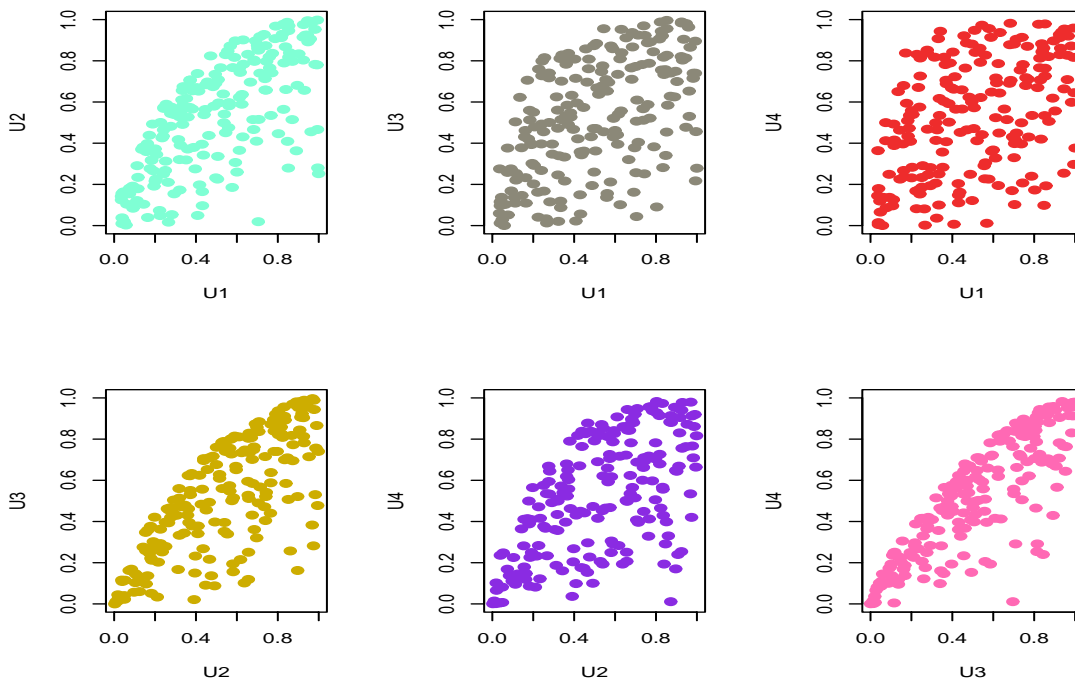


Figura 5.2: Gráficos com cruzamentos entre dados simulados da cópula K-extremal, com  $K = 4$

## 5.5 Simulações de processos (i.i.d.)

Como foi feito na seção (3.4), queremos verificar os resultados assintóticos para a cópula K-extremal. As simulações foram divididas em 5 etapas. Como forma simplificadora,



optamos por estudar a aderência dos dados simulados à cópula K-extremal com  $K = 4$ . Também aqui, assumimos  $m = 10$ , ou seja, para cada dimensão no intervalo  $(0, 1)$  foram feitas 10 partições de igual tamanho, o que resultou em 10.000 células no  $\mathbb{R}^4$ . Além disso, devido ao intenso cálculo computacional que seria necessário para se obter a probabilidade de cada célula, optamos por utilizar uma aproximação, que consistiu em multiplicar o volume de cada célula pelo valor da densidade da cópula 4-extremal em seu ponto médio no quadrado 4-dimensional. Foram construídos 12 experimentos de simulação. Estes tiveram como primeiro objetivo verificar a convergência dos resultados para várias distribuições de probabilidade contínua, sendo estas: (i) Normal(0,1), (ii) Normal(0,4), (iii) t-st(3), (iv) t-st(4), (v) Gama(1,3), (vi) Uniforme(0,1), para dois tamanhos de amostra  $N = 100$  e  $N = 300$ . Para cada processo de simulação foi calculada a estatística de teste  $D$  e o  $p$ -valor na respectiva distribuição qui-quadrado. A partir daí, como forma de verificar a aderência dos dados à cópula 4-extremal, foram calculados a média, o erro padrão da distribuição amostral dos  $p$ -valores obtidos e o percentual dos  $p$ -valores que foram maiores do que 0.05 para as 200 repetições de cada experimento realizado.

Para todos os processos em que os dados são *i.i.d.*, conforme ilustra a tabela (5.1) e nas figuras (A.1) a (A.3), constatamos que a cópula K-extremal com  $K = 4$  é ideal para representar a estrutura de dependência multivariada das 4 maiores estatísticas de ordem. Somente no caso uniforme, foi que obtemos um maior percentual de rejeição para o tamanho de amostra  $N = 100$ , mas, quando a amostra passou para  $N = 300$ , imediatamente houve uma expressiva diminuição deste percentual. Neste processo de simulação nosso intuito foi verificar e demonstrar a aderência dos resultados teóricos a dados simulados, sujeitos às hipóteses feitas previamente. A partir deste ponto, surge também o interesse de verificar qual seria a capacidade de modelagem da cópula K-extremal a conjuntos de dados oriundos de processos estacionários, que serão detalhados na seção 5.7.

Tabela 5.1: Tabela com dados simulados para a cópula 4-extremal - processos i.i.d.

	Média	Erro padrão	% > 0.05
Processo i.i.d. (N=100)			
$N(0, 1)$	0.8345	0.2770	97%
$N(0, 4)$	0.8380	0.2728	95%
$t - st(3)$	0.8496	0.2636	96%
$t - st(4)$	0.8585	0.2362	98%
$Gama(1, 3)$	0.8740	0.4208	99%
$Uniforme(0, 1)$	0.6370	0.4208	76%
Processo i.i.d. (N=300)			
$N(0, 1)$	0.7049	0.2863	98%
$N(0, 4)$	0.7290	0.2860	97%
$t - st(3)$	0.6470	0.3266	94%
$t - st(4)$	0.7196	0.2938	98%
$Gama(1, 3)$	0.7603	0.2705	98%
$Uniforme(0, 1)$	0.6800	0.3238	93%

## 5.6 Verificação empírica para a K-extremal

Nosso intuito, nesta seção, é verificar, com a utilização de dados reais, a aderência da estrutura de dependência da cópula K-extremal. Os dados utilizados nesta aplicação são os mesmos da seção 3.5, sendo que, para este caso, foram coletadas as 3 maiores observações para cada bloco. Após o ajuste da estrutura de dependência da cópula 3-extremal aos dados, obtemos um p-valor igual a 0.6573. Este valor aceita a hipótese nula, ou seja, que a cópula 3-extremal é apropriada para exprimir a estrutura de dependência, entre as três maiores estatísticas de ordem padronizadas.

## 5.7 Processos Estacionários

A generalização da teoria de valores extremos para processos estacionários está bem detalhada no trabalho de Leadbetter (1983). Esta foi fundamental para unificar os resultados, como também para desenvolver uma caracterização amplamente aplicável, adotando apenas fracas suposições de regularidade.

Os resultados ilustrados no capítulo 2 foram obtidos a partir de processos oriundos de uma seqüência de variáveis aleatórias independentes (para um melhor detalhamento veja o teorema 2.5). No entanto, para os tipos de dados aos quais os modelos de valores extremos são comumente aplicados, a independência temporal é uma suposição nada realista. Condições extremas freqüentemente persistem durante algumas observações consecutivas, trazendo à tona o questionamento quanto à adequação desses modelos. Uma investigação detalhada dessas questões requer um tratamento matemático num nível mais elevado. Contudo, a idéia básica não é difícil e o principal resultado obtido tem uma interpretação intuitiva bastante simples. Um detalhamento mais preciso é dado em Leadbetter *et al* (1983).

A generalização mais natural de uma seqüência de variáveis aleatórias independentes é para séries estacionárias. Estacionariedade é uma suposição bem mais realista para vários processos, correspondendo a séries cujas variáveis podem ser mutuamente dependentes, mas cujas propriedades estocásticas são homogêneas através do tempo.

A dependência em séries estacionárias pode assumir diversas formas, sendo, portanto, impossível desenvolver uma caracterização geral do comportamento de extremos, a menos que algumas restrições sejam impostas. É usual assumir a condição que restringe a extensão da dependência de longo prazo em níveis extremos, de tal forma que eventos  $X_i > u$  e  $X_j > u$  sejam aproximadamente independentes, para valores de  $u$  suficientemente grandes, e os tempos  $i$  e  $j$  bastante distanciados no tempo. Em outras palavras, eventos extremos ocorridos em tempos bastantes distanciados entre si são aproximadamente independentes. Muitas séries estacionárias satisfazem esta propriedade. Mas o mais importante é que se trata de uma propriedade freqüentemente plausível para processos físicos (Rubem, A.P.S. (2006)).

Segundo Morettin e Tolo (1981), processos estacionários são uma classe de processos estocásticos que satisfazem suposições simplificadoras. Intuitivamente, um processo  $\{X_t\}$  é estacionário se ele se desenvolve no tempo de modo que a escolha de uma origem dos tempos não é importante. Em outras palavras, as características de  $X_{t+h}, \forall h$  são as mesmas de  $X_t$ . Podemos dar como exemplo de processos estacionários as medidas das vibrações de um avião em regime estável de vôo horizontal durante seu cruzeiro, como também as várias formas de "ruídos" podem ser considerados processos estacionários.

Tecnicamente, há duas formas de estacionariedade: fraca (ou ampla de segunda ordem) e estrita (ou forte). Para o nosso estudo, iremos detalhar somente a estacionariedade estrita ou forte. Como forma de simplificação iremos usar a notação "estacionário" para "estritamente estacionário".

**Definição 5.3** (Morettin e Tolo (1981)) *Um processo estocástico  $\{X_t\}, t \in T$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações do tempo, ou seja:*

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$$

para quaisquer  $t_1, \dots, t_n, h \in T$ . Isto significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo, e portanto tem média e variância constantes para todo  $t \in T$ , e  $cov(X_n, X_m) = cov(X_0, X_{|n-m|}) \forall n, m$ .

Em Embrechts, P., Klüppelberg, C., Mikosch, T. (1997) é mencionado ser impossível generalizar a teoria de valores extremos para todos os processos estritamente estacionários, entretanto para uma classe destes processos que sejam assintoticamente independentes isso será possível. Para tanto, definiremos formalmente para uma seqüência qualquer  $(u_n)_{n=1}^{+\infty}$  de números reais, as condições  $D(u_n)$  e  $D'(u_n)$  que serão mencionadas a seguir.

**Condição 5.1**  $D(u_n)$ : Para qualquer inteiro  $p, q$  e  $n$

$$1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$$

desde que  $j_1 - i_p \geq l$ , nos temos

$$\left| P \left( \max_{i \in A_1 \cup A_2} X_i \leq u_n \right) - P \left( \max_{i \in A_1} X_i \leq u_n \right) P \left( \max_{i \in A_2} X_i \leq u_n \right) \right| \leq \alpha_{n,l},$$

onde  $A_1 = \{i_1, \dots, i_p\}$ ,  $A_2 = \{j_1, \dots, j_q\}$  e  $\alpha_{n,l} \rightarrow 0$  quando  $n \rightarrow \infty$  para alguma seqüência  $l = l_n = o(n)$ .

**Condição 5.2**  $D'(u_n)$ : A seguinte relação deve ser satisfeita:

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = 0.$$

De posse destas condições podemos obter generalizações do teorema 2.5 para processos estacionários.

**Teorema 5.7** (Embrechts, P., Klüppelberg, C., Mikosch, T. (1997)) Seja  $\{X_n\}$  um processo estacionário e defina  $M_n = \max\{X_1, \dots, X_n\}$ . Então, se existem seqüências de constantes  $\{a_n > 0\}$  e  $\{b_n\}$  tais que

$$Pr \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \rightarrow G_1(x)$$

onde  $G_1$  é uma função de distribuição não-degenerada, e as condições  $D(a_n x + b_n)$  e  $D'(a_n x + b_n)$  são satisfeitas para todo  $x$  real, então  $G_1$  é membro da família GEV.

**Teorema 5.8** (Embrechts, P., Klüppelberg, C., Mikosch, T. (1997)) Seja  $\{X_n\}$  um processo estacionário e  $\{X_n^*\}$  uma seqüência de variáveis aleatórias independentes com a

mesma distribuição marginal. Defina  $M_n = \max\{X_1, \dots, X_n\}$  e  $M_n^* = \max\{X_1^*, \dots, X_n^*\}$ .

Assumindo  $\forall x \in \mathbb{R}$ ,  $D(a_n x + b_n)$  e  $D'(a_n x + b_n)$  são satisfeitas,

$$Pr \left\{ \frac{M_n^* - b_n}{a_n} \leq z \right\} \rightarrow G_1^*(z)$$

quando  $n \rightarrow \infty$  para seqüências normalizadoras  $\{a_n > 0\}$  e  $\{b_n\}$ , onde  $G_1$  é uma função de distribuição não-degenerada, se e somente se

$$Pr \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \rightarrow G_1(z),$$

onde

$$G_1(z) = [G_1^*(z)]^\theta \tag{5.19}$$

para uma constante  $\theta$  tal que  $0 < \theta \leq 1$ .

O teorema (5.8) implica que, se a seqüência de máximos de uma série estacionária converge, o que é assegurado pelo teorema (5.7), desde que as condições  $D(a_n x + b_n)$  e  $D'(a_n x + b_n)$  sejam satisfeitas, a distribuição limite está relacionada com a distribuição limite de uma seqüência independente, conforme a equação (5.19). Este resultado é consistente com o do teorema 5.7, pois se  $G_1^*$  pertence à família GEV, o mesmo ocorre com  $[G_1^*]^\theta$ . Mais precisamente, se  $G_1^*$  tem distribuição GEV com parâmetros  $(\mu^*, \sigma^*, \xi^*)$ , e  $\xi^* \neq 0$ , então  $G_1$  tem distribuição GEV com parâmetros  $(\mu, \sigma, \xi)$ , onde

$$\mu = \mu^* - \frac{\sigma^*}{\xi^*} (1 - \theta^{-\xi^*})$$

e

$$\sigma = \sigma^* \theta^{\xi^*}.$$

Também pode-se estender os resultados ilustrados na seção (2.2.2) com relação à distribuição limite para a  $K$ -ésima estatística de ordem.

**Teorema 5.9** (Embrechts, P., Klüppelberg, C., Mikosch, T. (1997)) *Seja  $\{X_n\}$  um processo estritamente estacionário, e denotando  $M_{k,n}$  como a  $k$ -ésima estatística de ordem. Supondo existirem seqüências de constantes  $a_n > 0$  e  $b_n$  e assumindo  $\forall x \in \mathbb{R}$ ,  $D(a_n x + b_n)$  e  $D'(a_n x + b_n)$  são satisfeitas,*

$$Pr \left\{ \frac{M_{1,n} - b_n}{a_n} \leq z \right\} \rightarrow G_1(z) \text{ quando } n \rightarrow \infty,$$

onde  $G_1$  é uma função de distribuição não-degenerada, tal que  $G_1$  é uma GEV dada por (2.9), então:

$$Pr \left\{ \frac{M_{k,n} - b_n}{a_n} \leq z \right\} \rightarrow G_k(z),$$

onde

$$G_k(z) = G_1(z) \sum_{s=0}^{k-1} \frac{[-\log(G_1(z))]^s}{s!}. \quad (5.20)$$

para todo  $k \geq 1$ .

O teorema 5.9 mostra a similaridade entre os resultados assintóticos obtidos para sequências estacionárias com sequências *i.i.d.* Este ganho só foi possível em virtude de as condições  $D(a_n x + b_n)$  e  $D'(a_n x + b_n)$  terem sido assumidas. Logo, naturalmente surge o interesse de verificar se a cópula K-extremal é adequada para representar também a estrutura de dependência destes processos estacionários. Entre os processos estacionários que satisfazem as condições anteriores, podemos destacar os processos ARMA gaussianos.

## 5.8 Simulação de processos ARMA

Para verificar se a cópula K-extremal também pode ser utilizada para estes processos, foram construídos novos experimentos, inicialmente foram gerados dados oriundos de processos auto-regressivos AR(1),  $Z_t = \phi_1 Z_{t-1} + \epsilon_t$ , onde  $\epsilon_t$  são erros Gaussianos independentes, aqui neste caso tendo distribuição  $N(0, 1)$ . Os dados provenientes desta etapa estão na tabela 5.2 e nas figuras A.4 a A.5. As conclusões a que podemos chegar são que somente para o processo onde a dependência temporal é extremamente fraca, ou seja,  $\phi = 0.1$ , os valores das estatísticas observados não sofreram grandes mudanças quando o tamanho da amostra aumentou. Além disso, fica evidente que quanto maior for a dependência temporal e quanto maior for nossa amostra, mais rapidamente o índice de *p-valores*  $> 0.05$  diminui, chegando ao extremo para  $\phi = 0.8$ , em que obtemos um percentual de *p-valores*  $> 0.05$  igual 0%.

Outro experimento consistiu em gerar dados oriundos de processos médias móveis MA(1),  $Z_t = \theta_1 \epsilon_{t-1} + \epsilon_t$ , onde  $\epsilon_t$  são erros gaussianos, aqui neste caso tendo distribuição

Tabela 5.2: Tabela resumo com dados simulados para a cópula 4-extremal - AR(1)

Processo fracamente estacionários - AR(1)	Média	Erro padrão	% > 0.05
Erros Gaussianos( N=100)			
$\phi_1 = 0.1$	0.8278	0.2789	98%
$\phi_1 = 0.4$	0.6487	0.3330	95%
$\phi_1 = 0.6$	0.2836	0.3341	59%
$\phi_1 = 0.8$	0.0074	0.0307	0.5%
Erros Gaussianos (N=300)			
$\phi_1 = 0.1$	0.6981	0.2763	97%
$\phi_1 = 0.4$	0.2553	0.2717	68%
$\phi_1 = 0.6$	0.0020	0.0087	1%
$\phi_1 = 0.8$	$4.35 \times e^{-17}$	$4.46 \times e^{-16}$	00%

$N(0, 1)$ . Os dados oriundos desta etapa estão na tabela 5.3 e nas figuras A.6 a A.7. A maioria das simulações neste caso ocasionou que a cópula 4-extremal torna-se uma opção para ajuste. Mas, da mesma forma que nos processos AR(1), também aqui, conforme o parâmetro  $\theta_1$  aumenta, também o índice de *p-valores* maiores do que 0.05 diminui, sendo que com menor intensidade do que no caso anterior. Também, quando o tamanho da amostra aumenta, o índice de *p-valores* maiores do que 0.05 diminui.

Para as simulações oriundas de processos ARMA(1,1), foram feitas algumas combinações entre os parâmetros  $\phi$  e  $\theta$ . Além disso, somente foram geradas simulações com tamanho de amostra igual  $N = 300$ . Os resultados encontram-se na tabela 5.4, bem como nas figuras A.8 a A.10. Aqui, igualmente como nos dois experimentos anteriores de simulação, observa-se que, dependendo da combinação feita entre os parâmetros  $\phi$  e  $\theta$ , a cópula K-extremal não pode ser empregada.

O último experimento consistiu em verificar o comportamento do ajuste dos dados pela cópula K-extremal para erros não gaussianos, em particular, para erros com distribuição  $U(-1, 1)$ . As simulações foram oriundas de um processo AR(1). Os resultados encontram-se na tabela 5.5, bem como nas figuras A.11 a A.12. Neste caso, não poderíamos aceitar a hipótese de que a cópula K-extremal é apropriada para ajustar os



Tabela 5.3: Tabela resumo com dados simulados para a cópula 4-extremal - MA(1)

Processo fracamente estacionários - MA(1)	Média	Erro padrão	% > 0.05
Erros Gaussianos( N=100)			
$\theta_1 = 0.1$	0.8166	0.2753	99%
$\theta_1 = 0.4$	0.6859	0.3399	94%
$\theta_1 = 0.6$	0.6539	0.3482	91%
$\theta_1 = 0.8$	0.6188	0.3650	89%
Erros Gaussianos (N=300)			
$\theta_1 = 0.1$	0.7155	0.2894	98%
$\theta_1 = 0.4$	0.4692	0.3120	86%
$\theta_1 = 0.6$	0.3464	0.3133	77%
$\theta_1 = 0.8$	0.2649	0.2854	65%

dados.

Diante destes resultados, chega-se à conclusão que a cópula K-extremal não pode ser utilizada para representar a estrutura de dependência entre as K-maiores estatísticas para todas as classes de seqüências estacionárias. Este fato ocorre porque, apesar das condições  $D(u_n)$  e  $D'(u_n)$  serem satisfeitas para o caso dos processos ARMA gaussianos, estas não garantem que a distribuição conjunta também é assintoticamente igual à obtida para o caso de seqüências *i.d.d.*. Entretanto, segundo Embrechts, P., Klüppelberg, C., Mikosch, T. (1997), para que os resultados assintóticos sejam os mesmos, é necessário introduzir uma analogia para o caso k-dimensional da condição  $D(u_n)$ .

**Condição 5.3**  $D_k(\mathbf{u}_n)$ : Para qualquer inteiro fixo  $p, q$

$$1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$$

desde que  $j_1 - i_p \geq l$ , temos

$$\left| P \left( X_{i_m} \leq u_n^{(s_m)}, m = 1, \dots, p, X_{j_r} \leq u_n^{(s'_r)}, m = 1, \dots, q \right) - P \left( X_{i_m} \leq u_n^{(s_m)}, m = 1, \dots, p, \right) P \left( X_{j_r} \leq u_n^{(s'_r)}, m = 1, \dots, q \right) \right| \leq \alpha_{n,l},$$

para qualquer  $1 \leq s_l, s'_r \leq k$ , e  $\alpha_{n,l} \rightarrow 0$  quando  $n \rightarrow \infty$  para alguma seqüência  $l = l_n = o(n)$ .

Tabela 5.4: Tabela resumo com dados simulados para a cópula 4-extremal - ARMA(1)

Processo fracamente estacionários - ARMA(1)	Média	Erro padrão	% > 0.05
Erros Gaussianos( N=300)			
$\phi_1 = .1, \theta_1 = 0.1$	0.7721	0.3157	96%
$\phi_1 = .4, \theta_1 = 0.1$	0.5534	0.3590	86%
$\phi_1 = .6, \theta_1 = 0.1$	0.1817	0.2601	50%
$\phi_1 = .8, \theta_1 = 0.1$	0.2218	0.3075	47%
$\phi_1 = .1, \theta_1 = 0.8$	0.0031	0.0190	2%
$\phi_1 = .4, \theta_1 = 0.8$	0.2670	0.3327	58%
$\phi_1 = .6, \theta_1 = 0.8$	0.0848	0.1885	23%
$\phi_1 = .8, \theta_1 = 0.8$	0.0022	0.0145	2%
$\phi_1 = .4, \theta_1 = 0.4$	0.3912	0.3686	72%
$\phi_1 = .5, \theta_1 = 0.5$	0.2111	0.2944	47%

Além disso, não é necessário definir uma extensão para a condição  $D'(u_n)$ , se simplesmente assumir que  $D'(u_n^{(i)})$  seja satisfeita separadamente para cada  $i = 1, \dots, k$ . A pergunta que pode ser feita, neste ponto, é:

**Existe alguma classe de processos estacionários em que a distribuição conjunta satisfaça as condições  $D_k(\mathbf{u}_n)$  e  $D'(u_n^{(i)})$ ,  $i = 1, \dots, k$ ?**

A princípio, não foi possível identificar tais classes, principalmente pela dificuldade de verificar a condição  $D_k(\mathbf{u}_n)$ . Estudos posteriores podem ter como objetivo verificar quais classes atenderiam esta condição. Além disso, também algum cálculo empírico poderia ser proposto, tendo como objetivo principal evidenciar a aplicabilidade da cópula K-extremal.

Tabela 5.5: Tabela resumo com dados simulados para a cópula 4-extremal - AR(1)

Processo fracamente estacionários - AR(1)	Média	Erro padrão	% > 0.05
Erros Uniformes ( N=100)			
$\phi_1 = 0.1$	0.5799	0.4260	72%
$\phi_1 = 0.4$	0.6369	0.3617	90%
$\phi_1 = 0.6$	0.2368	0.2946	57%
$\phi_1 = 0.8$	0.0116	0.0601	0.4%
Erros Uniformes (N=300)			
$\phi_1 = 0.1$	0.3192	0.3695	54%
$\phi_1 = 0.4$	0.1202	0.1991	42%
$\phi_1 = 0.6$	0.0004	0.0028	00%
$\phi_1 = 0.8$	$3.21 \times e^{-18}$	$3.46 \times e^{-17}$	00%

# Capítulo 6

## Conclusão

Vários resultados de extrema relevância foram obtidos nos capítulos anteriores. O primeiro resultado, que foi a cópula Bi-extremal, definida completamente por suas funções de densidade e distribuição, possibilitou um extenso estudo sobre as propriedades contidas nessa nova estrutura assintótica de dependência não-linear (capítulo 4). Essa também poderá ser fundamental para novos estudos de simulações. É interessante ressaltar que a estrutura de dependência desta nova cópula poderá ser empregada como uma função de ligação entre marginais univariadas que geraram distribuições bi-variadas.

Já a cópula K-extremal, em sua essência, teve como objetivo principal generalizar os resultados obtidos da cópula Bi-extremal. No decorrer desse trabalho, vários resultados teóricos foram obtidos. Entre esses, o de maior destaque foi constatar que, independentemente da distribuição contínua primitiva  $F$  dos dados, a cópula das K-maiores estatísticas de ordem padronizadas é a cópula K-extremal. Além disso, essa possibilita, através de suas expressões exatas para a densidade e a função de distribuição, a propagação de estudos sobre a dependência assintótica não-linear entre quaisquer pares de estatísticas de ordem.

Como desafios futuros temos: Inicialmente obter uma família de cópulas para a Bi-extremal. Esse estudo poderá ter como referência os resultados obtidos em Nelsen (2006). Além disso, contruir uma biblioteca para o *R-project* com as funções utilizadas em todo este estudo, que possibilitarão a propagação das pesquisas relacionadas com a cópula K-extremal. Um grande desafio seria também identificar quais processos estacionários

se enquadrariam na estrutura de dependência da cópula K-extremal. O pesquisador Thomas Mikosch, um dos autores do livro *Modelling extremal events for insurance and finance*, em contato por *email*, relata que a condição  $D_K$  necessária para que a distribuição conjunta das K-maiores estatísticas de ordem seja a mesma à do caso *i.i.d.* é implicitamente satisfeita nos processos de misturas forte, ver Doukhan (1994). Esta última proposta de trabalho tende a ser de extrema importância para futuras aplicações dos resultados já obtidos, visto que, na prática, muitas séries temporais não são processos totalmente *i.i.d.*.

# Referências Bibliográficas

- [1] Anjos, U., Kolev, N., Tanaka, N. (2005). "Copula associated to order statistics". BJPS 19, 111-124.
- [2] Anjos, U., Kolev, N. (2005). Representation of bivariate copulas via local measure of dependence. Technical Report RT-MAE 2005-03, University of São Paulo. Submitted.
- [3] Avérous, J., Genest, C., Kochar, S. (2005). On the dependence structure of order statistics. *Journal of Multivariate Analysis* 94, 159-171.
- [4] Belitsky, V., Moreira, F.M. (2007). *A thorough yet simple exposition of the peaks-over-threshold and its employment for risk estimation*. Third Brazilian Conference on Statistical Modelling in Insurance and Finance, 2007.
- [5] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.
- [6] Charpentier, A. (2004). Extremes and dependence: a copula approach. Submitted. Available at <http://www.crest.fr/pageperso/lfa/charpent/charpent.htm>
- [7] Chen, Y. (2007). *A note on the relationship between Spearman's  $\rho$  and Kendall's  $\tau$  for extreme order statistics*, *Journal of Stat. Plain. and Inf.* 137,2165-2171.
- [8] Cherubini, U., Luciano, E., Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley Finance: Chichester.
- [9] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer: Berlin.

- [10] DeGroot, M. H. (2002). Probability and Statistics, Addison-Wesley, 3rd ed.
- [11] Doukhan, P. (1994). *Mixing: Properties and Examples* Lecture Notes in Statistics. 1994
- [12] Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag: Berlin.
- [13] Fermanian, JD. (2005). Goodness of fit tests for copulas. *Journal of Multivariate Analysis*, 119-152.
- [14] Fisher, R.A., and Tippett, L.H.C.(1928) *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*, Proceedings of the Cambridge Philosophical Society, n.24, p.180-190.
- [15] Fredricks, G.A., Nelsen, R.B. (2007). *On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variable*. Journal of Statistical Planning and Inference, in press.
- [16] Frees, E.W, Valdez, E. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, **2**, 1-25.
- [17] Gnedenko, B.V. (1943). *Sur la distribution limite du terme maximum d'une serie aléatoire*, Annals of Mathematics n.44,p.423-453.
- [18] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.
- [19] James B. R. (1981). *Probabilidade: um curso em nível intermediário*, CNPq-IMPA Projeto Euclides, Rio de Janeiro.
- [20] Jenkinson, A.F. (1955). *The frequency distribution of the annual maximum (or minimum) values of meteorological events*, Quarterly Journal of the Royal Meteorological Society 81, 158-172, 1955.
- [21] Leadbetter, M. (1983). *Extremes and local dependence in stationary sequences*, Zeit, Wahrscheinl-theorie, 16, 291-396.

- [22] Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, Berlin.
- [23] Magalhães, M.N. (2006). *Probabilidade e Variáveis Aleatórias*. EDUSP - Editora da Universidade de São Paulo, São Paulo-Brasil.
- [24] Morettin, P.A., Toloi, C.M.C. (1981). *Modelos para Previsão de Séries Temporais*, 13º Colóquio Brasileiro de Matemática. CNPQ-IMPA
- [25] Nelsen, R.B. (1992). *On measures of association as measures of positive dependence*. *Statistics Probab. Letters*, 14, 269-274.
- [26] Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer: New York.
- [27] Patton, A., Andrew, J. (2006). *Asymmetric Exchange Rate Dependence*. *International Economic Review*, Vol. 47, No. 2, pp. 527-556, May 2006.
- [28] Rubem, A.P.S. (2006). *Modelagem de extremos baseada nas r-maiores estatísticas de ordem*. Dissertação de Mestrado, IM-UFRJ, 2006.
- [29] Schweizer, B., Wolff, E.F. (1981) *On nonparametric measures of dependence for random variables*. *The Annals of Statistics*, 9, 879-885.
- [30] Smith, R.L. (1986). *"Extreme value theory based on the r largest annual events"*, *J. Hydrology*, 86, 27-43.
- [31] Sklar, A. (1959). *Fonctions de répartition á n dimensions et leurs marges*. *Publ. Inst. Statist. Univ. Paris* **8**, 229-231.
- [32] Schmitz, V. (2004). *"Revealing the dependence structure between  $X_{(1)}$  and  $X_{(n)}$ "*. *J. Statist. Plann. Inference*, 123, 1, 41-47.
- [33] Tawn, J. (1988). *Bivariate extreme value theory: models and estimation*. *Biometrika* **75**, 397-415.
- [34] Weissman, I. (1978) *Estimation of parameters and large quantiles based on the k-largest observations*, *Journal of American Statistical Society*, v.73, 812-815.



# Apêndice A

## Gráficos das simulações

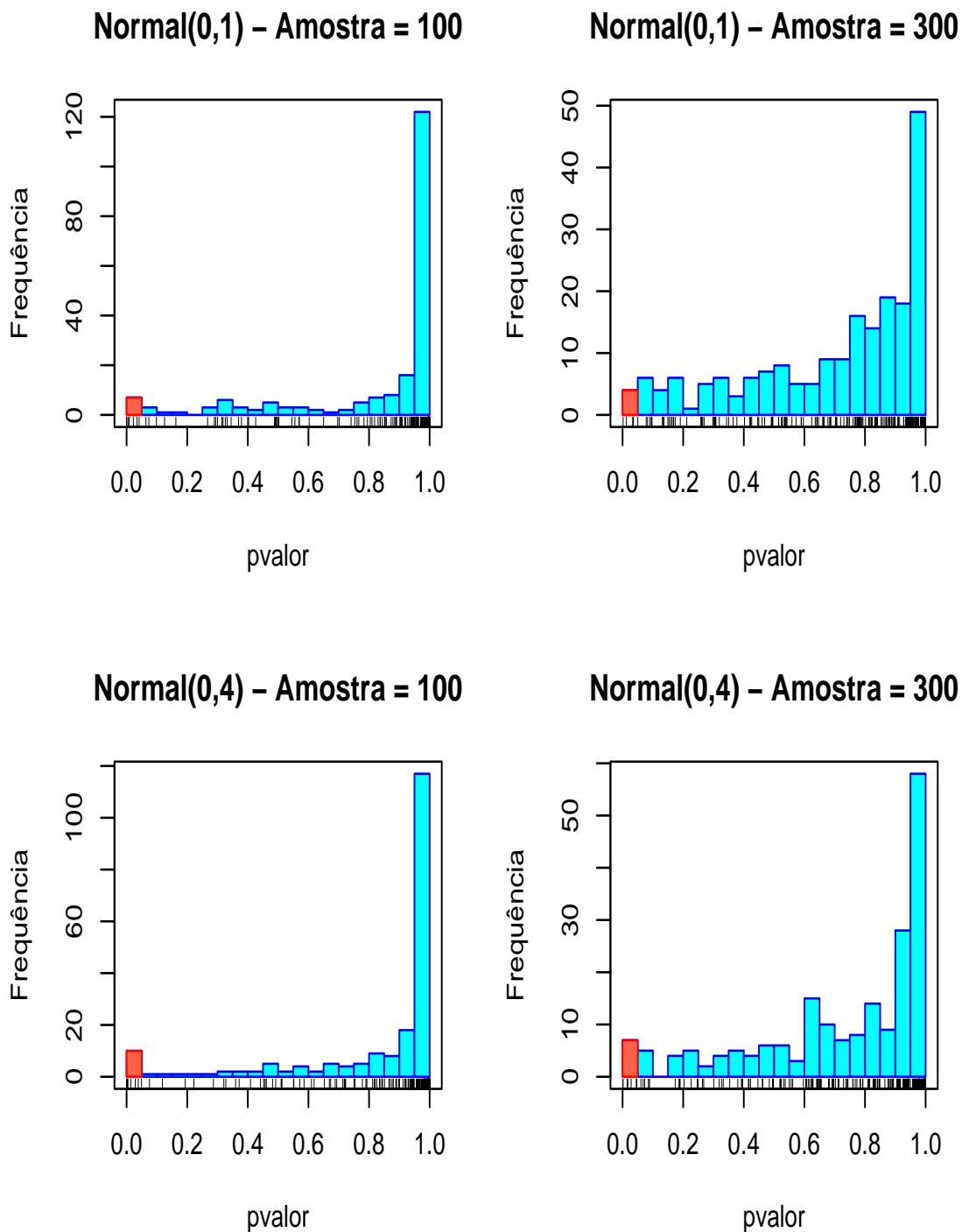


Figura A.1: Histograma das simulações de processos (*i.i.d*) da cópula *K*-extremal com  $K = 4$ , para  $Normal(0,1)$  e  $Normal(0,4)$

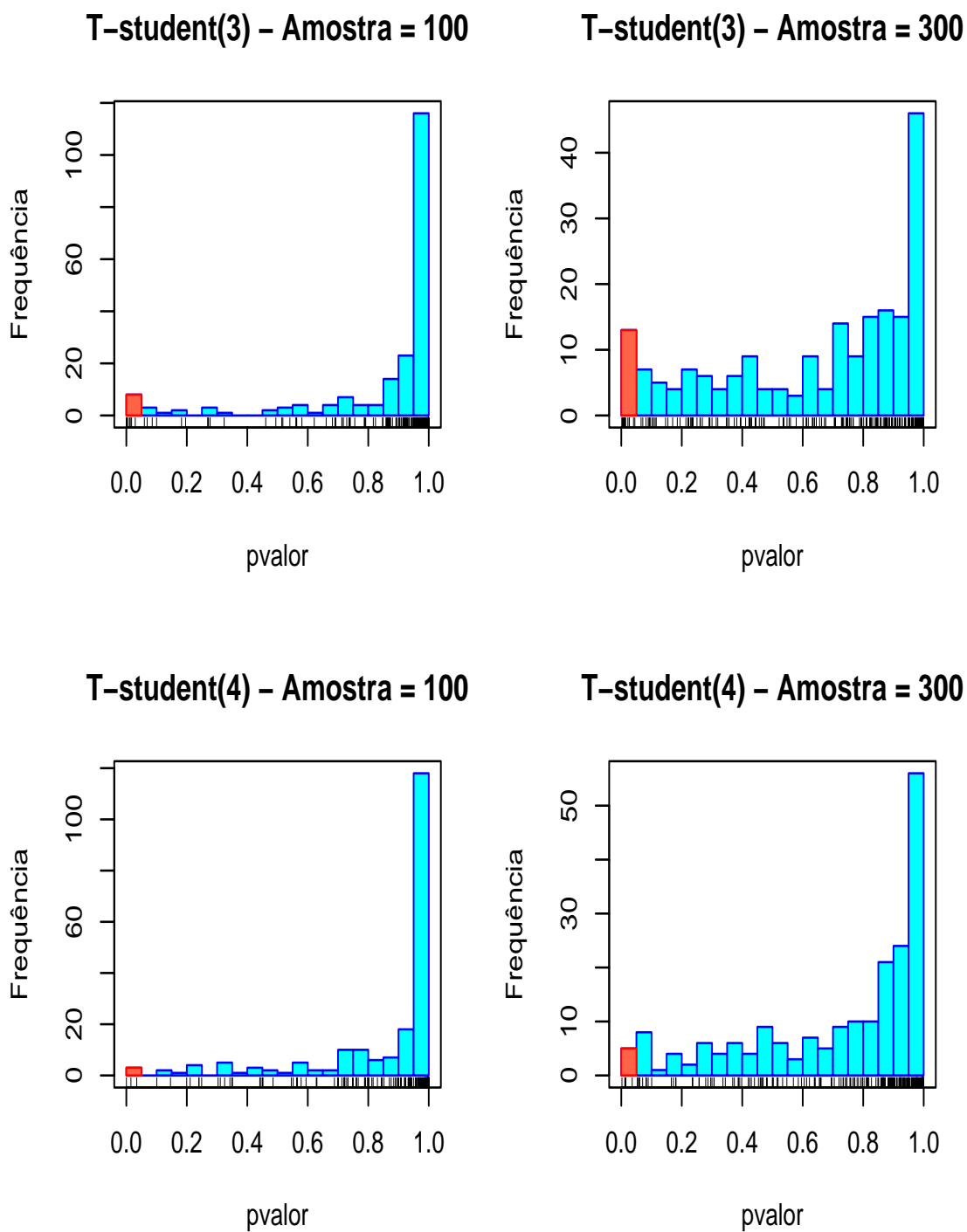
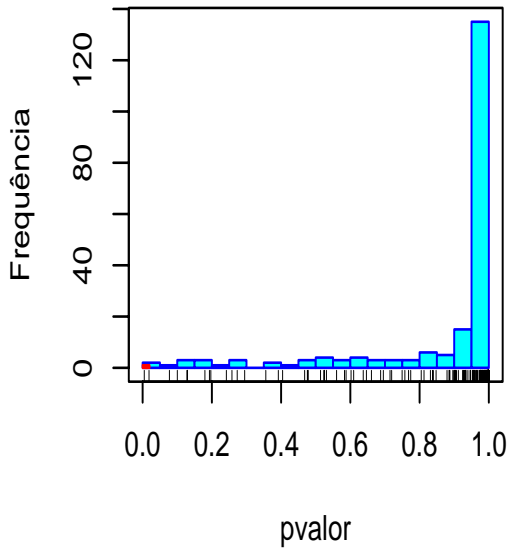
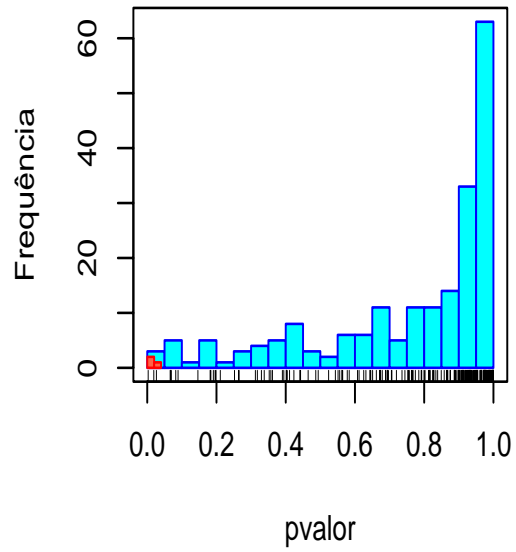


Figura A.2: Histograma das simulações de processos (*i.i.d*) da cópula *K*-extremal, com  $K = 4$ , para a *T-student*(3) e *T-student*(4)

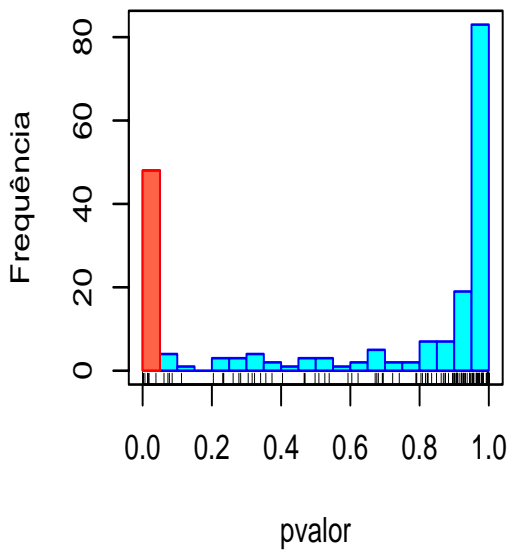
**Gama(1,3) - Amostra = 100**



**Gama(1,3) - Amostra = 300**



**Uniforme(0,1) - Amostra = 100**



**Uniforme(0,1) - Amostra = 300**

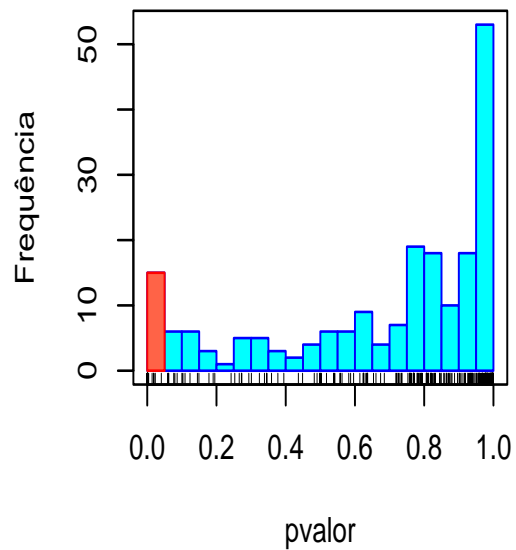
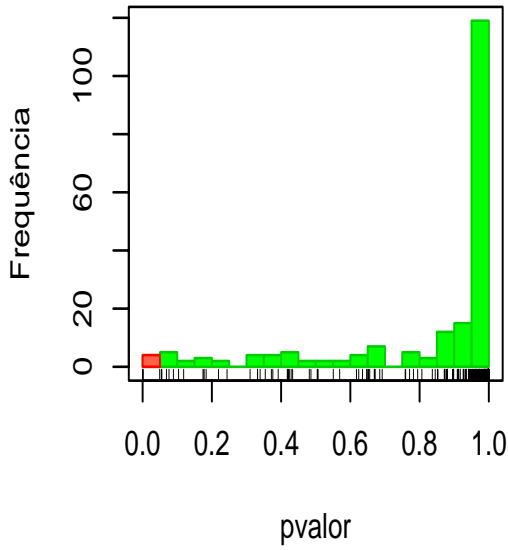
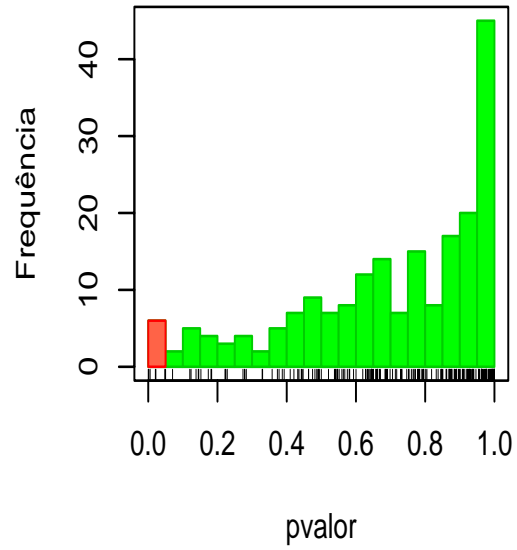


Figura A.3: Histograma das simulações de processos (*i.i.d*) da cópula *K*-extremal com  $K = 4$ , para *Gama*(1,3) e *Uniforme*(0,1)

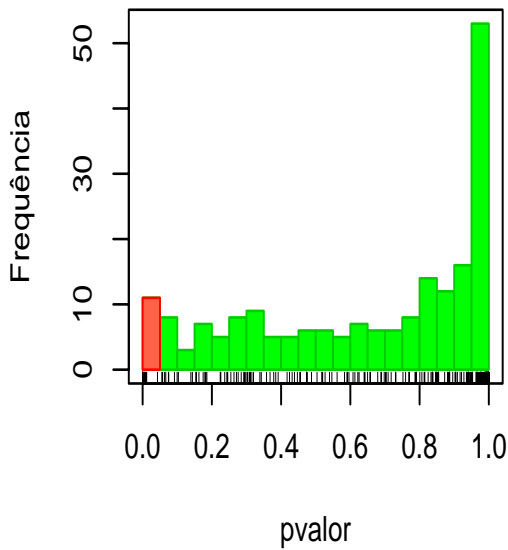
**AR(1) -  $f_i=1$  - Amostra = 100**



**AR(1) -  $f_i=1$  - Amostra = 300**



**AR(1) -  $f_i=4$  - Amostra = 100**



**AR(1) -  $f_i=4$  - Amostra = 300**

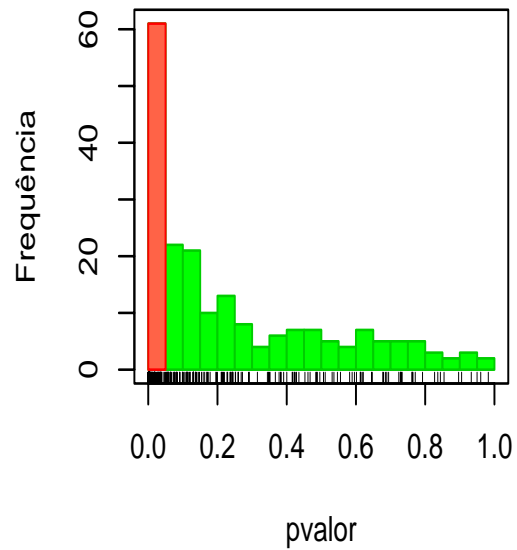
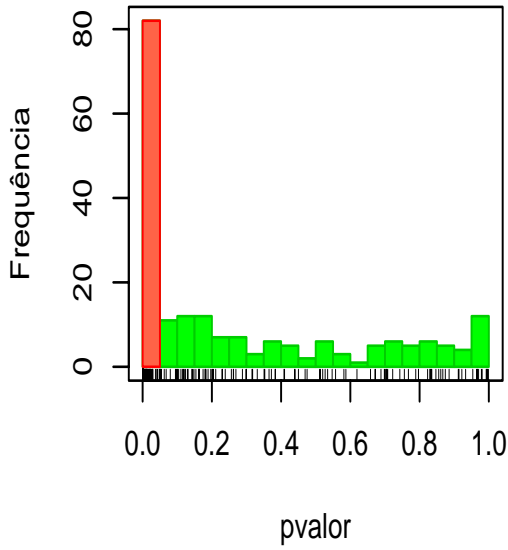
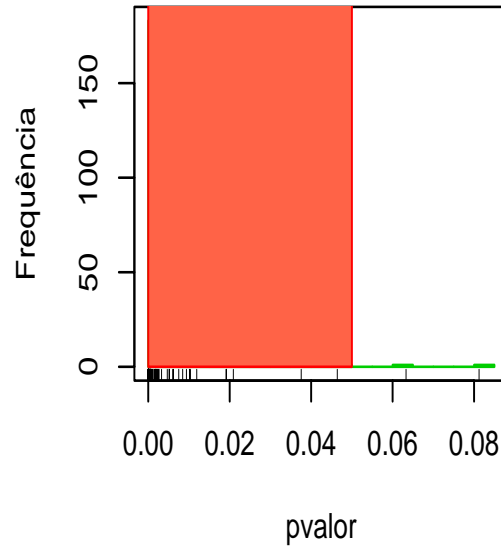


Figura A.4: *Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para  $K = 4$ .*

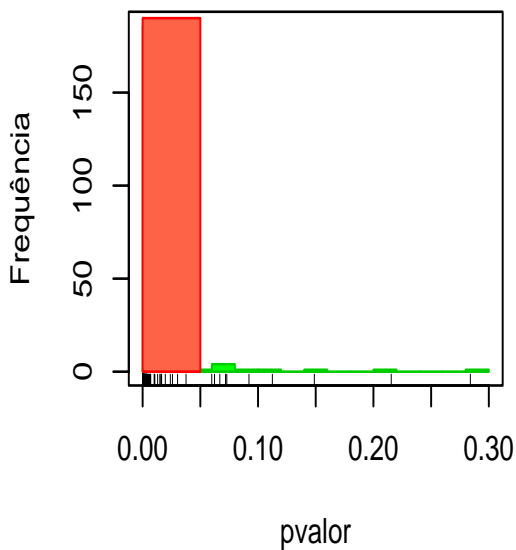
**AR(1) -  $f_i=0.6$  - Amostra = 100**



**AR(1) -  $f_i=0.6$  - Amostra = 300**



**AR(1) -  $f_i=0.8$  - Amostra = 100**



**AR(1) -  $f_i=0.8$  - Amostra = 300**

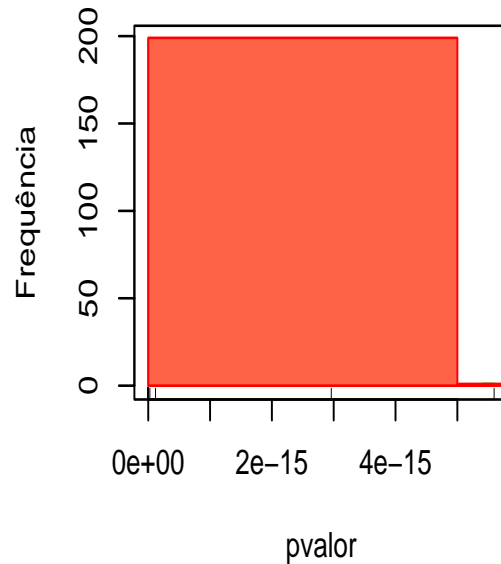
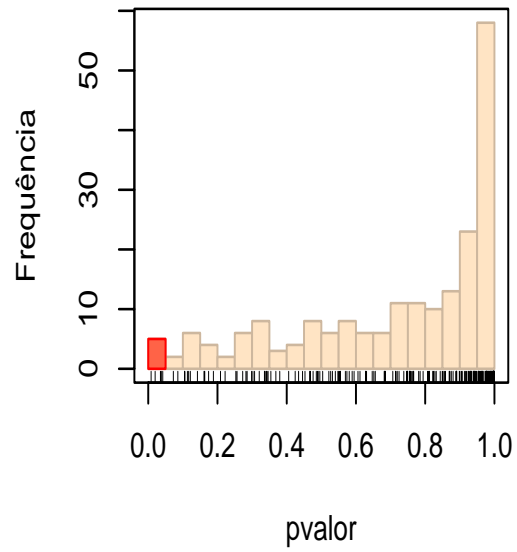
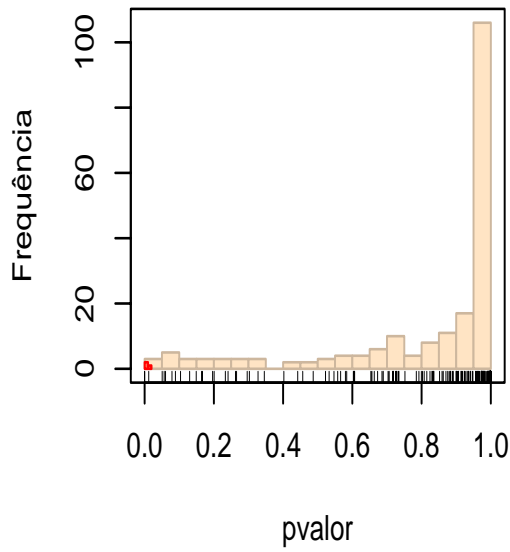


Figura A.5: *Histograma das simulações de processos fracamente estacionários AR(1), da cópula K-extremal, para  $K = 4$ .*

**MA(1) - Theta=.1 - Amostra = 10**      **MA(1) - Theta=.1 - Amostra = 30**



**MA(1) - Theta=.4 - Amostra = 10**      **MA(1) - Theta=.4 - Amostra = 30**

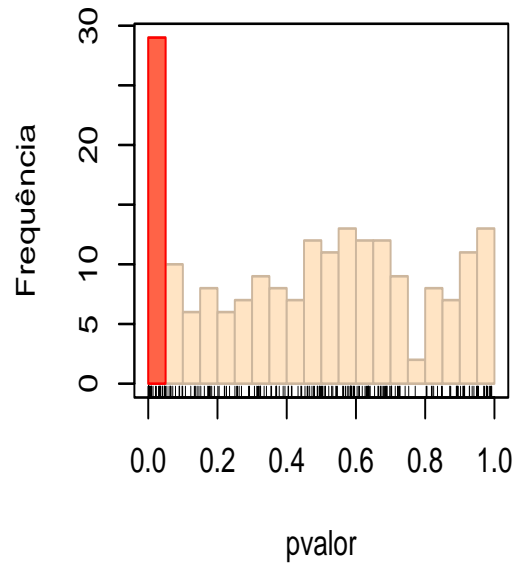
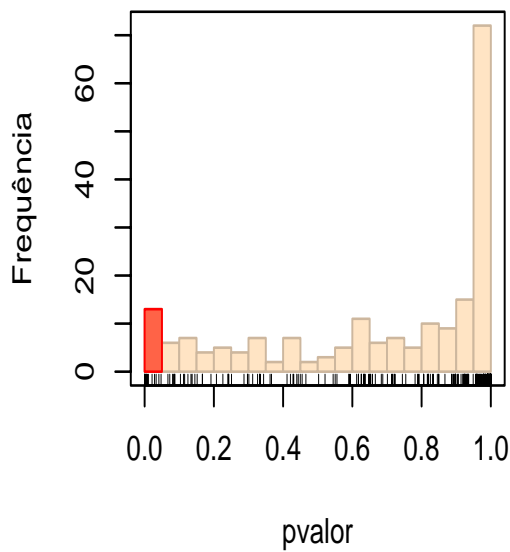
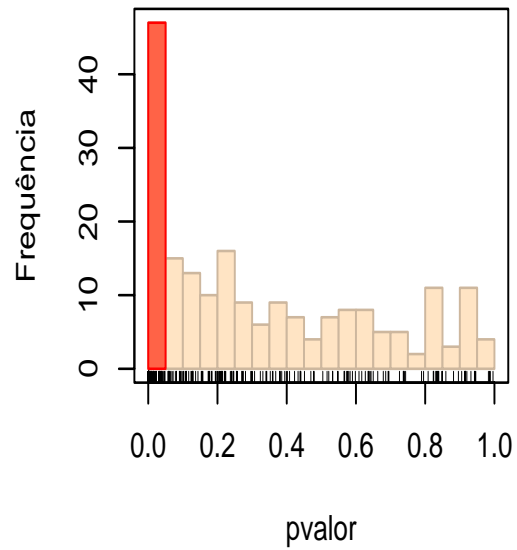
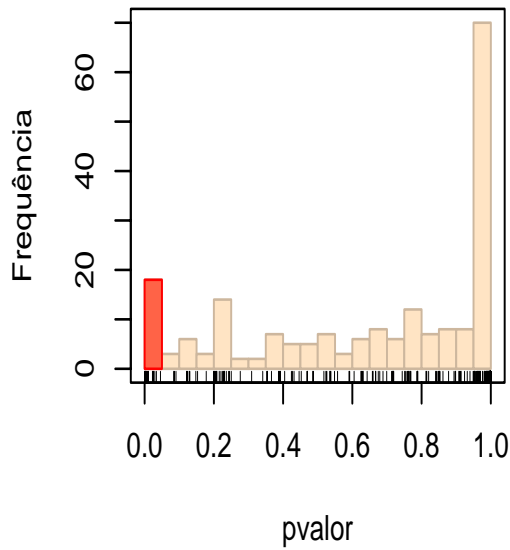


Figura A.6: *Histograma das simulações de processos fracamente estacionários MA(1), da cópula K-extremal, para K = 4.*

**MA(1) - Theta=.6 - Amostra = 10**    **MA(1) - Theta=.6 - Amostra = 30**



**MA(1) - Theta=.8 - Amostra = 10**    **MA(1) - Theta=.8 - Amostra = 30**

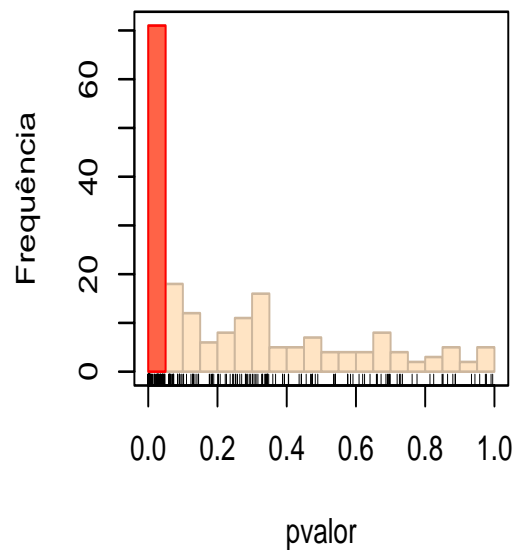
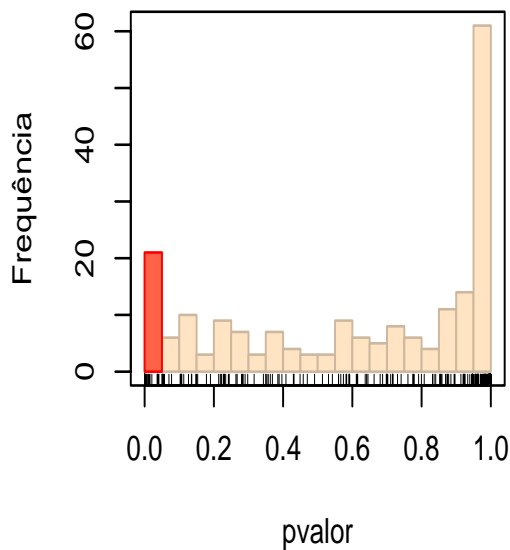
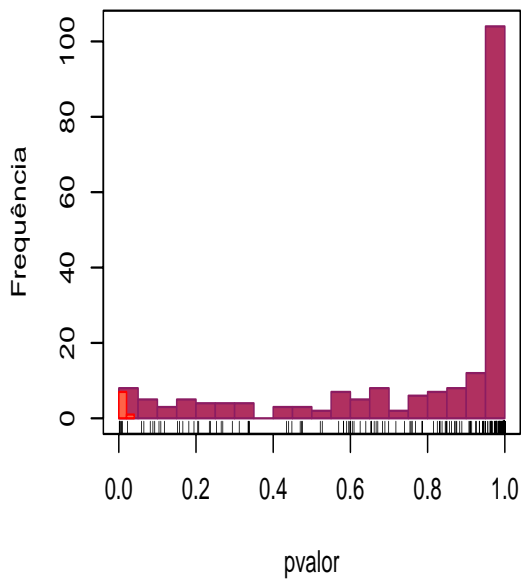


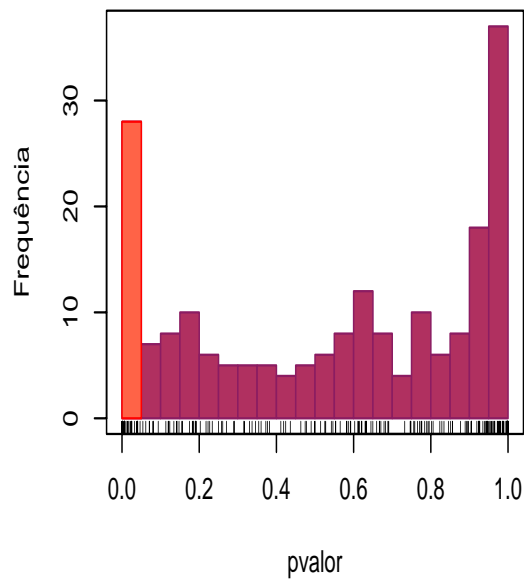
Figura A.7: *Histograma das simulações de processos fracamente estacionários MA(1), da cópula K-extremal, para K = 4.*



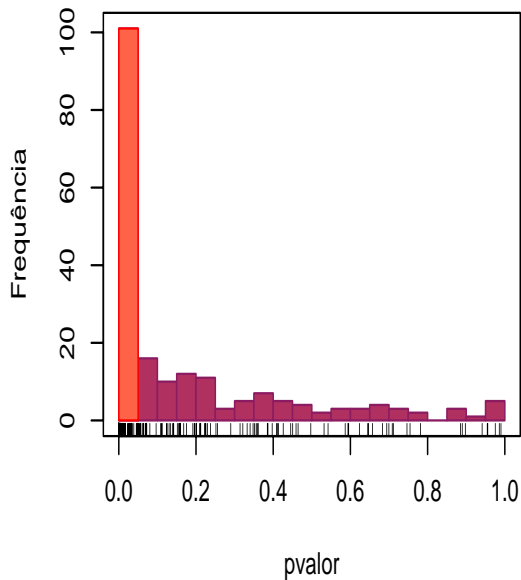
**ARMA(1) ,  $\phi_1=1$  e  $\theta=1$  - Amostra = 300**



**ARMA(1) ,  $\phi_1=4$  e  $\theta=1$  - Amostra = 300**



**ARMA(1) ,  $\phi_1=6$  e  $\theta=1$  - Amostra = 300**



**ARMA(1) ,  $\phi_1=8$  e  $\theta=1$  - Amostra = 300**

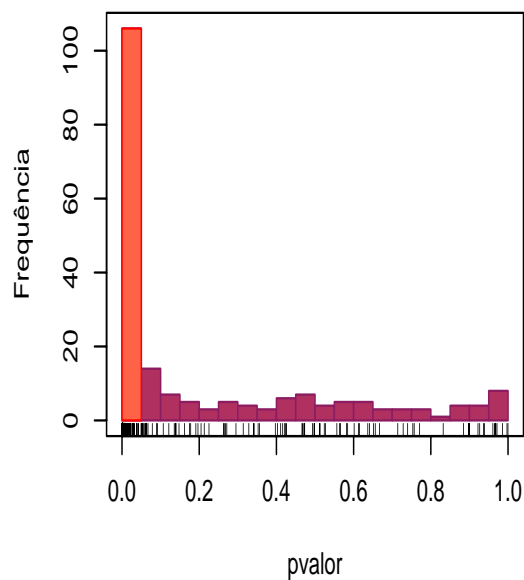
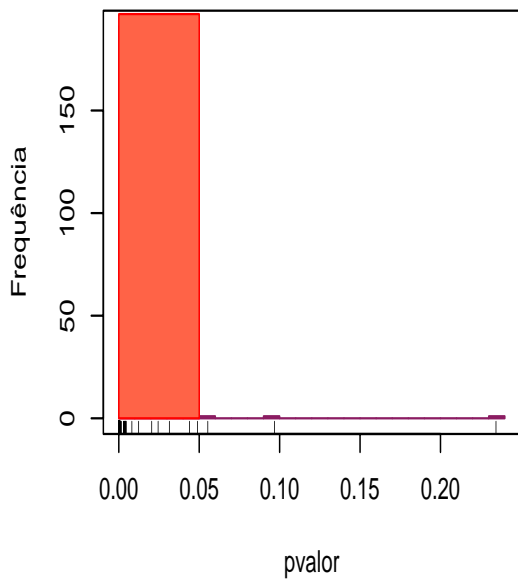
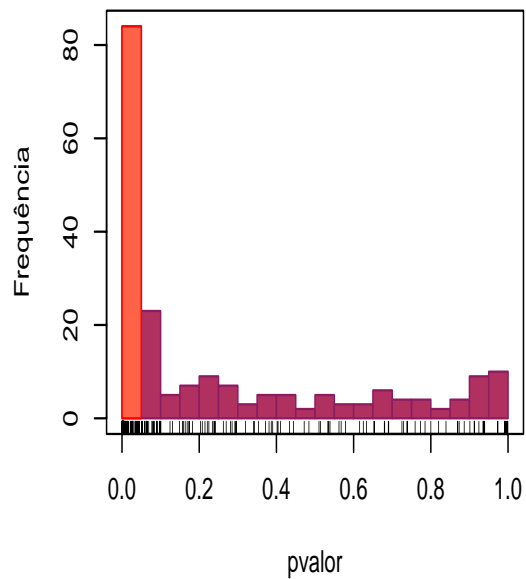


Figura A.8: *Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para  $K = 4$ .*

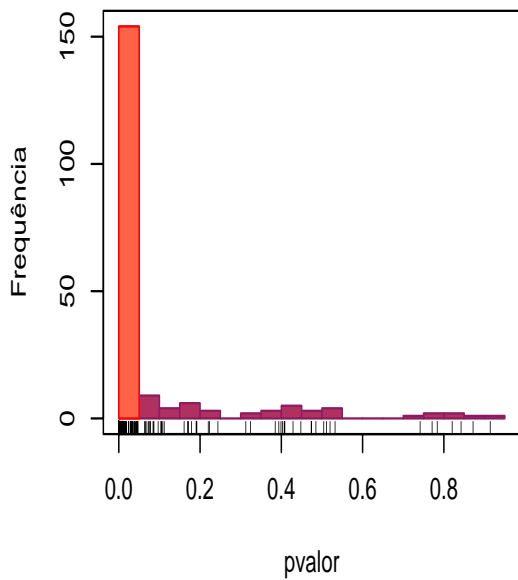
**ARMA(1) ,  $\phi_i=1$  e  $\Theta=0.8$  - Amostra = 300**



**ARMA(1) ,  $\phi_i=0.4$  e  $\Theta=0.8$  - Amostra = 300**



**ARMA(1) ,  $\phi_i=0.6$  e  $\Theta=0.8$  - Amostra = 300**



**ARMA(1) ,  $\phi_i=0.8$  e  $\Theta=0.8$  - Amostra = 300**

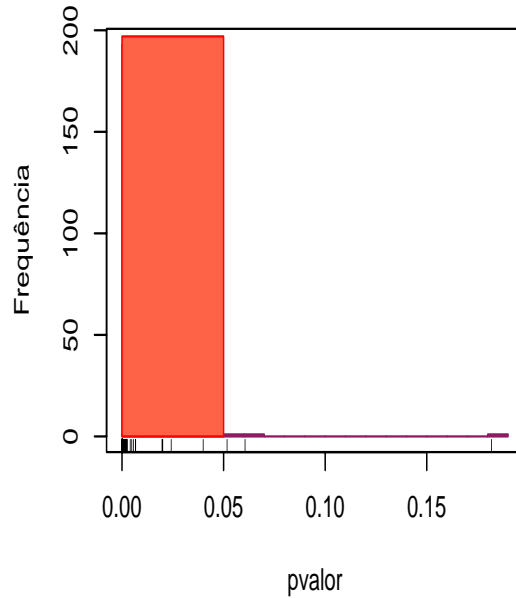
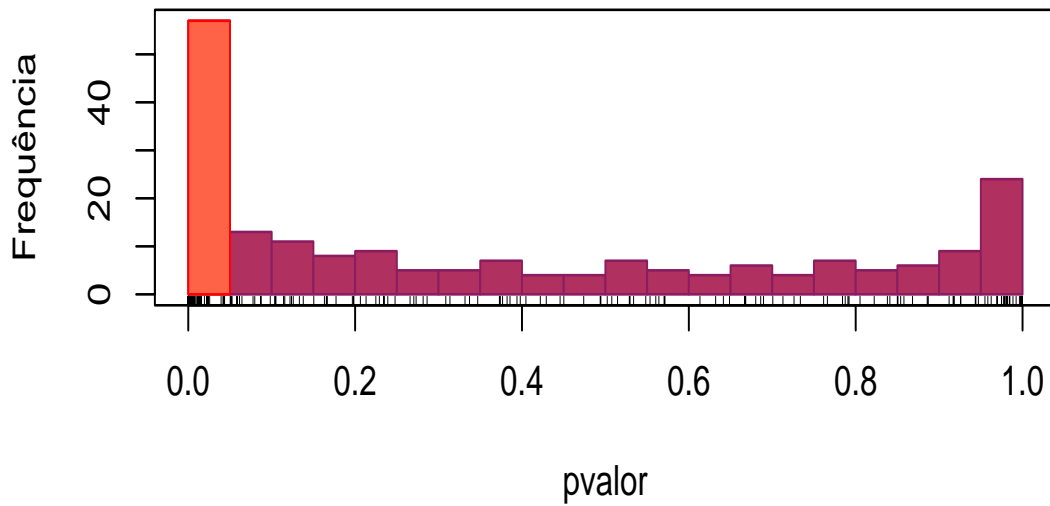


Figura A.9: *Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para  $K = 4$ .*

### ARMA(1) , $\phi_i=.4$ e $\Theta=.4$ - Amostra = 300



### ARMA(1) , $\phi_i=.5$ e $\Theta=.5$ - Amostra = 300

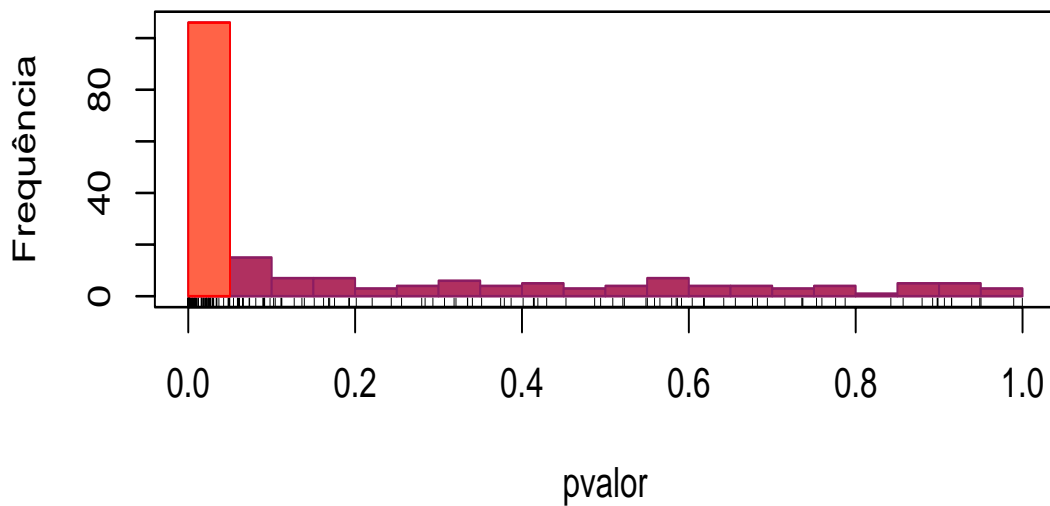
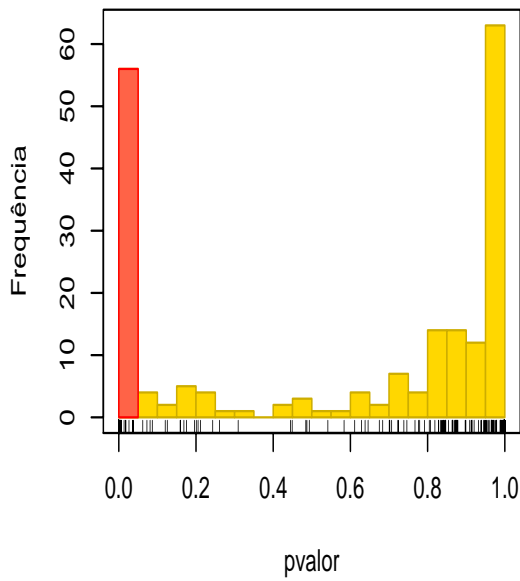
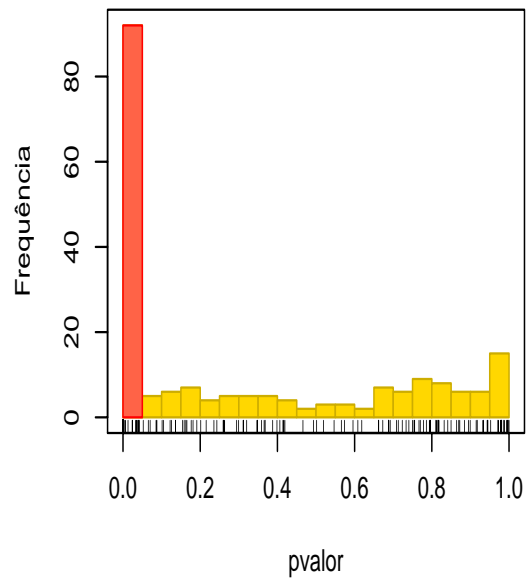


Figura A.10: *Histograma das simulações de processos fracamente estacionários ARMA(1), da cópula K-extremal, para  $K = 4$ .*

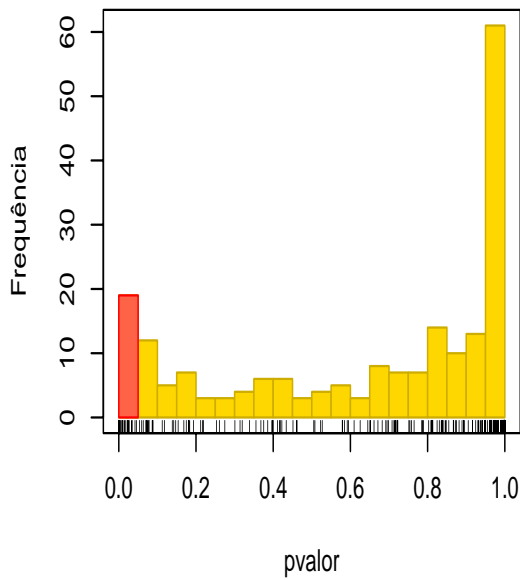
**AR(1),  $f_i=1$  e erro  $U(-1,1)$  - Amostra = 100**



**AR(1),  $f_i=1$  e erro  $U(-1,1)$  - Amostra = 300**



**AR(1),  $f_i=4$  e erro  $U(-1,1)$  - Amostra = 100**



**AR(1),  $f_i=4$  e erro  $U(-1,1)$  - Amostra = 300**

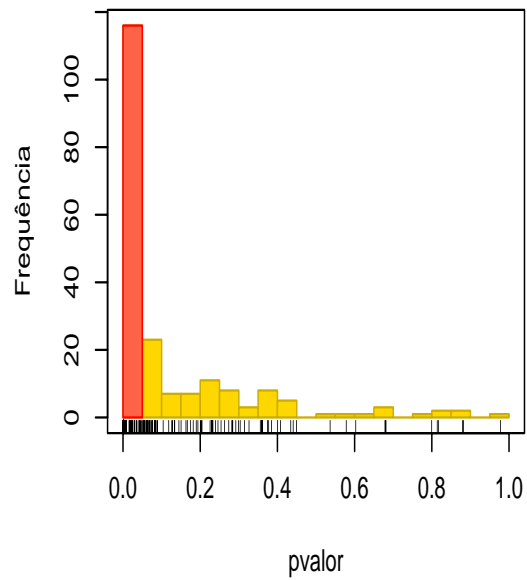
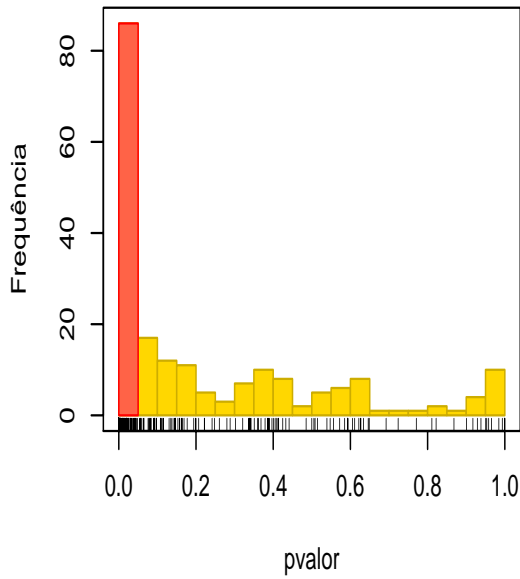
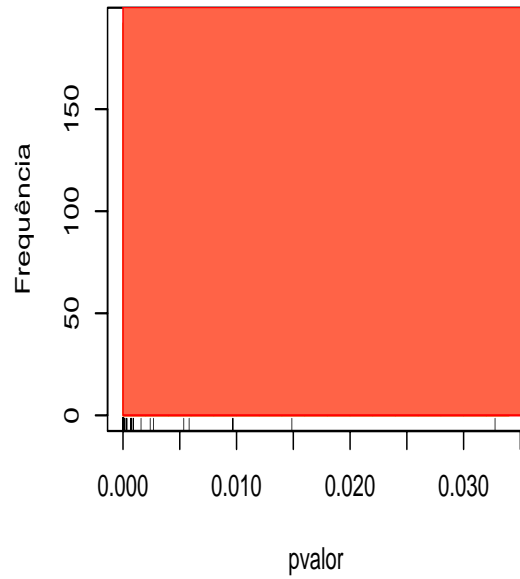


Figura A.11: *Histograma das simulações de processos fracamente estacionários  $AR(1)$ , da cópula  $K$ -extremal, para  $K = 4$  com erros  $U(-1,1)$ .*

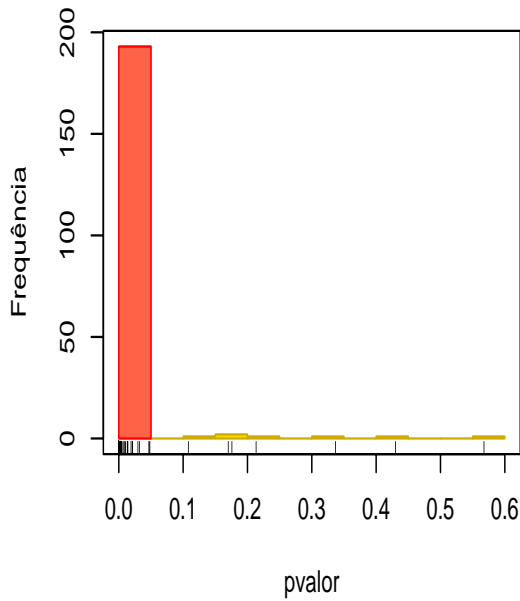
AR(1),  $\phi=0.6$  e erro  $U(-1,1)$  - Amostra = 100



AR(1),  $\phi=0.6$  e erro  $U(-1,1)$  - Amostra = 300



AR(1),  $\phi=0.8$  e erro  $U(-1,1)$  - Amostra = 100



AR(1),  $\phi=0.8$  e erro  $U(-1,1)$  - Amostra = 300

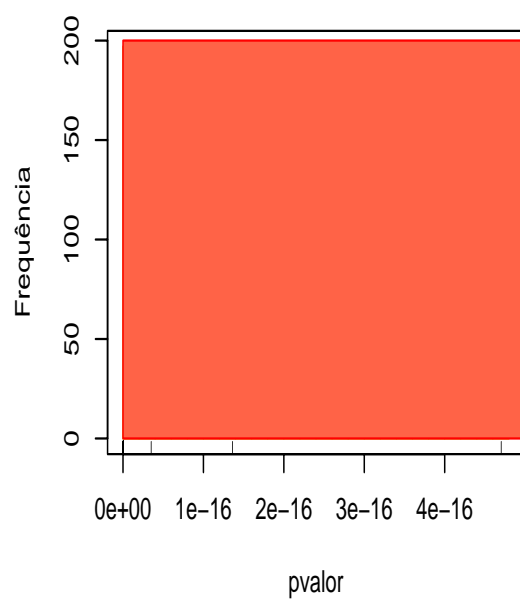


Figura A.12: Histograma das simulações de processos fracamente estacionários  $AR(1)$ , da cópula  $K$ -extremal, para  $K = 4$  com erros  $U(-1,1)$ .

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)