Teoria de Valores Extremos e Cópulas: Distribuição Valor Extremo e Cópulas Arquimedianas Generalizadas Trivariadas

Márcio Luis Lanfredi Viola

IMECC-UNICAMP

2006

Teoria de Valores Extremos e Cópulas: Distribuição Valor Extremo e Cópulas Arquimedianas Generalizadas Trivariadas

Márcio Luis Lanfredi Viola

Agradecimentos

Agradeço a todos que contribuiram, direta ou indiretamente para a realização deste trabalho.

À minha orientadora Verónica Andrea González-López pela sua paciência e atenção.

À minha co-orientadora Laura Letícia Ramos Rifo pela leitura da dissertação.

À todos os meus amigos, especialmente, Marcelo Cristino Gama e Marcos Eduardo Valle pelas discussões matemáticas e estatísticas e pelas dicas de latex.

À CAPES pelo suporte financeiro.

Resumo

Sob a ótica da Teoria de Cópulas, a modelagem multidimensional pode ser considerada decorrente de dois processos: estimação das distribuições de probabilidade marginais e modelagem de uma estrutura de dependência multidimensional que age sobre tais distribuições marginais, sendo esta última, denominada cópula. Neste trabalho, as estruturas marginais de interesse correspondem ao máximo de uma variável aleatória e, consequentemente, a Teoria de Valores Extremos apresenta-se como uma alternativa natural para a modelagem. Nesta dissertação, serão estudados os tipos de dependência entre variáveis aleatórias, a construção e implementação de modelos de Teoria de Cópulas assim como, os resultados básicos de convergência utilizados na Teoria de Valores Extremos. Sob o escopo da Teoria de Valores Extremos, os métodos de estimação pontual de Máxima Verossimilhança e L-momentos serão comparados através de algumas simulações e, adicionalmente, serão abordadas as condições que asseguram a validade das propriedades assintóticas do Método de Máxima Verossimilhança bem como as principais propriedades de ambos os métodos citados. As teorias citadas serão aplicadas no contexto de Lingüística na modelagem multidimensional de características do sinal acústico observadas em regiões de baixa, média e alta freqüência de frases da língua inglesa.

Abstract

Introdução

Esta dissertação aborda tópicos sobre a Teoria de Valores Extremos, Métodos de Estimação e Teoria de Cópulas e a posterior aplicação destas teorias em dados reais provenientes da área de Linguística.

Como um dos propósitos será a modelagem dos máximos de uma variavel aleatória, será utilizada a Teoria de Valores Extremos pois esta fornece a distribuição assintótica do máximo de uma variável aleatória. Neste caso, tal distribuição assintótica é a Distribuição Valor Extremo Generalizada (GEV). A fundamentação desta teoria será dada no Capítulo 1 e está baseada no conceito de variável aleatória max-estável.

O próximo passo requer a utilização de métodos de estimação para obter-se os parâmetros da Distribuição GEV. Para isto, pode-se utilizar o Método de Máxima Verossimilhança sendo que a Distribuição GEV possui suporte que depende de seus parâmetros, ou seja, a estimação deles é um caso não-regular pois tal distribuição não satisfaz as condições de regularidade referentes à estimação pelo Método de Máxima Verossimilhança. Com isso, as propriedades assintóticas, como consistência, normalidade assintótica dos estimadores dos parâmetros da Distribuição GEV, obtidos pelo Método de Máxima Verossimilhança, podem não valer, embora elas permanecem válidas quando $\alpha < 1/2$ [**?**]. Isto será discutido no Capítulo 2.

Como uma alternativa ao Método de Máxima Verossimilhança, será considerado o Método dos L-momentos [**?**]. Este método é baseado em quantidades denominadas L-momentos que são funções da esperança de estatísticas de ordem. Os L-momentos possuem propriedades como a caracterização de distribuições [**?**, **?**], são não-viciados e possuem normalidade assintótica, independentemente do suporte da distribuição depender dos seus parâmetros [**?**]. Este método será discutido no Capítulo 2.

Hosking [**?**] comparou os métodos de L-momentos e Máxima Verossimilhança utilizando a eficiência assintótica dada por $eff(\hat{\theta}_i) := lim_{n\to\infty}\frac{Var(\tilde{\theta}_i)}{Var(\hat{\theta}_i)}$ para cada elemento θ_i do vetor de parâmetros θ onde $\tilde{\theta}_i$ e $\hat{\theta}_i$ são, respectivamente, os estimadores de máxima verossimilhança e L-momentos de θ_i. Hosking [**?**] ressalta várias vantagens do Método dos L-momentos em relação ao Método de Máxima Verossimilhança: Os estimadores obtidos pelo Método dos L-momentos apresentaram, em geral, vício pequeno sendo que este decresce rapidamente quando aumenta-se o tamanho amostral.

Adicionalmente, os desvios padrões dos estimadores obtidos pelo Método dos L-momentos foram comparavéis com aqueles obtidos pelo Método de Máxima Verossimilhança para tamanhos amostrais moderados (por exemplo, tamanho amostral 100) e foram, frequentemente, substancialmente menores que os desvios padrões dos estimadores obtidos pelo Método de Máxima Verossimilhança para pequenas amostras (por exemplo, tamanho amostral 15, 25).

Nesta dissertação, a comparação entre os métodos será feita através dos valores das estimativas pontuais dos parâmetros da Distribuição GEV obtidas a partir de amostras geradas de tamanhos 15, 50, 100, 500 e 1000. As comparações entre os métodos serão apresentadas no Capítulo 2. Este tipo de comparação será feita para conhecer-se o comportamento pontual das estimativas dos métodos dos L-momentos e Máxima Verossimilhança para pequenas e grandes amostras e em relação aos valores de inicialização dos métodos numéricos.

Para a utilização dos métodos de estimação, citados anteriormente, necessitou-se de procedimentos numéricos. Para a obtenção das estimativas pelos métodos de L-momentos e Máxima Verossimilhança utilizou-se, respectivamente, os Métodos da Bissecção e Nelder-Mead. O primeiro destes obtém as raízes de uma equação e o outro maximiza uma função. Como os métodos numéricos requerem valores iniciais para a inicialização, tomou-se o cuidado de que os valores iniciais estivessem próximos dos verdadeiros valores dos parâmetros através do uso de uma perturbação dos verdadeiros valores. Para o método da Bissecção utilizou-se o intervalo inicial $[\alpha - \varepsilon, \alpha + \varepsilon]$ e, para o Método de Máxima Verossimilhança, utilizou-se os valores iniciais $\mu + \varepsilon$, $\sigma + \varepsilon$ e $\alpha + \varepsilon$ onde μ, σ e α são os parâmetros da Distribuição GEV e ε é a perturbação. Para a comparação dos métodos de estimação, considerou-se o mesmo valor de ε para ambos os métodos numéricos. Este cuidado é necessário pois não é conveniente compararmos os métodos com valores iniciais muito distantes já que, para valores iniciais muito diferentes, os resultados fornecidos pelos métodos numéricos podem ser totalmente discrepantes e, consequentemente, a comparação ficará comprometida. De forma alguma, isto representa uma restrição dos métodos de estimação pois é apenas um cuidado tomado para a comparação de tais métodos. Outros métodos numéricos poderiam ser utilizados sendo que os métodos da Bissecção e Nelder-Mead foram escolhidos por não utilizarem derivadas.

O próximo passo será a modelagem da função de distribuição de probabilidade conjunta dos máximos. Para isto será utilizada a Teoria de Cópulas.

No Capítulo 3 serão construídos modelos de cópulas trivariadas com dependência positiva. Para

isto, será mostrada a construção dos modelos de cópulas arquimedianas bivariadas e a extensão de tais modelos para o caso trivariado. Isto requer o uso da Transformada de Laplace que permitirá a construção de cópulas arquimedianas e as generalizações destas.

As estruturas de cópulas estão associadas com tipos específicos de dependência. Por exemplo, pode-se notar que as cópulas arquimedianas bidimensionais possuem densidade TP_2 que é um dos tipos de dependência positiva. Sendo assim, serão explorados diversos tipos de dependências multidimensionais. No Capítulo 3 será feita uma coletânea dos tipos de dependências abordando propriedades e exemplos.

Na aplicação, será considerada apenas a extensão trivariada do modelo de Kimeldorf e Sampson e do modelo de Gumbel. Como uma extensão da dissertação, posteriormente, outros modelos de cópula poderão ser testados.

No Capítulo 4, será mostrada uma aplicação da Teoria de Valores Extremos, Métodos de Estimação e Teoria de Cópulas na modelagem de energias construídas a partir de um sinal acústico, em três regiões de frequências, referentes às línguas inglesa e francesa. Foi proposta a obtenção da função de distribuição acumulada conjunta dos log-retornos máximos nas três regiões de frequências associadas à cada frase através dos modelos de cópula citados. Primeiramente, para cada região de frequências, serão obtidos os log-retornos das energias com o propósito de modelar a variação da energia no tempo t em relação à energia no tempo $t-1$, isto é, será modelada a taxa de mudança entre energias consecutivas. Mas, os log-retornos de cada frase podem apresentar auto-correlações. Então, serão utilizados modelos auto-regressivos a fim de eliminar tais auto-correlações tornando, assim, as observações independentes de uma dada amostra de forma que a hipótese de independência do Teorema de Fisher-Tippett seja satisfeita.

Após a aplicação de modelos auto-regressivos, será obtido, para cada frase, o máximo dos seus resíduos sendo que tais máximos serão modelados através das teorias de Valores Extremos e Cópulas. A modelagem dos log-retornos máximos das energias é apenas uma proposta sendo que poderiam ser modeladas, diretamente, as energias.

Os modelos auto-regressivos poderão apresentar, para algumas frases, ordem alta. Uma alternativa a estes modelos, seria a utilização de outras técnicas que tornasse os dados independentes. Outras técnicas serão investigadas futuramente.

Os modelos generalizados trivariados e as cópulas empíricas foram comparados.

Para cada generalização do modelo de Kimeldorf e Sampson e do modelo de Gumbel, a cópula $C(u_1, u_2, u_3)$ e as cópulas marginais $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ foram comparadas, respectivamente, com as distribuições empíricas através do Teste de Wilcoxon a fim de verificar-se a proximidade de tais distribuições. Também, comparou-se tais modelos, considerando-se a maior distância, entre as cópulas $C(u_1, u_2, u_3)$, $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ e as respectivas distribuições empíricas. Este procedimento será repetido comparando-se as cópulas $C(u_1, u_2, u_3)$, $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ resultantes da generalização do modelo de Kimeldorf e Sampson com as cópulas resultantes da generalização do modelo de Gumbel para cada idioma.

Se as energias estressadas podem ser modeladas através das generalizações do modelo de Kimeldorf e Sampson e do modelo de Gumbel, estes modelos propostos podem ser usados para caracterizar a dependência das energias estressadas nas três regiões de frequências, podendo, futuramente, serem utilizados para a discriminação (em termos dos parâmetros dos modelos citados) de diversos idiomas.

Portanto, os Capítulos 1, 2 e 3 fornecem o suporte para a realização do objetivo desta dissertação que é o estudo do ajuste dos modelos de cópula, citados anteriormente, aplicados às línguas inglesa e francesa em relação à modelagem multidimensional de características de estresse do sinal acústico quantificadas pelas energias observadas nas três regiões de frequências sendo que o próximo passo seria o estudo de técnicas de agrupamento para que seja criada uma região de confiança em torno do ponto (θ_1, θ_2) correspondente aos modelos de cópula e aos idiomas considerados, de forma que uma dada frase possa ser discriminada automaticamente e alocada em algum idioma. Para isso, poderia ser estudada a distribuição amostral dos estimadores de (θ_1, θ_2) onde θ_1 e θ_2 são os parâmetros dos modelos de cópula.

Conteúdo

1 Caracterização da Convergência do Máximo de Variáveis Aleatórias **1**

1.1 **Introdução** 1

1.2 **Preliminares** 2

1.3 **Teorema de Fisher-Tippett e sua Implicação** 11

2 Estimação **15**

2.1 **Introdução** 15

2.2 **L-Momentos de uma Distribuição de Probabilidade** 16

2.2.1 **Estimação por L-momentos dos parâmetros da Distribuição GEV** . 24

2.3 **Método de Máxima Verossimilhança** 30

2.3.1 **Estimação por Máxima Verossimilhança dos parâmetros da Distribuição GEV** 45

2.4 **Comparação entre os Métodos dos L-momentos e Máxima Verossimilhança** 46

3 Cópula **65**

3.1 **Introdução** 65

3.2 **Tipos de Dependência entre Variáveis Aleatórias** 67

3.2.1 **Funções Totalmente Positivas e Totalmente Negativas** 67

3.2.2 **Dependência do Quadrante Positivo (e Negativo) e do Octante** . . . 75

3.2.3 **Variáveis Aleatórias Estocasticamente Crescentes e Decrescentes, Dependência Crescente na Cauda à Direita e Dependência Decrescente na Cauda à Esquerda** 78
3.2.4 **Variáveis Aleatórias Negativamente Associadas** 84
3.2.5 **Implicações e Contra-exemplos envolvendo os Conceitos de Dependência** 86
3.3 **Construção de Cópulas com Dependência Positiva** 94
3.4 **Implementação de um Modelo de Cópula** 105

4 Aplicação **107**
4.1 **Introdução** 107
4.2 **Metodologia** 110
4.3 **Comparação entre os Modelos** 114
4.4 **Conclusão** 115

Lista de Figuras

1.1 Distribuições Weibull, Fréchet e Gumbel geradas considerando-se os parâmetros $\mu = 0$, $\sigma = 1$ e, respectivamente, $\alpha = 0,5$, α= - 0,5 e α=0 13

1.2 Distribuições Weibull, Fréchet e Gumbel geradas considerando-se os parâmetros $\mu = 0$, $\sigma = 1$ e, respectivamente, α=2, α= -2 e α=0 13

3.1 À esquerda: Gráfico de Dispersão dos 1000 valores gerados das v.a. independentes $X \sim Exp(2)$ e $Y \sim Exp(10)$. À direita: Gráfico de Dispersão dos 1000 valores gerados das v.a. $X \sim Exp(2)$ e $Y \sim Exp(10)$ avaliados em suas respectivas f.d.a. 66

3.2 Gráficos da família B_2 correspondentes aos parâmetros $\delta = 0,5$ e $\delta = 2$ 101

3.3 Gráficos da família B_4 correspondentes aos parâmetros $\delta = 2$, $\delta = 5$ e $\delta = 9$ 101

3.4 Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1 103

3.5 Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1 103

3.6 Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1 103

3.7 Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1 104

3.8 Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1 104

3.9 Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1 104

3.10 Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1 104

3.11 Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1 104

3.12 Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1 104

3.13 Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2 104

3.14 Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2 104

3.15 Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2 104

3.16 Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2 104

3.17 Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2 104
3.18 Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2 104
3.19 Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2 104
3.20 Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2 104
3.21 Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2 104

4.1 Acima: Onda acústica resultante da pronúncia da frase, na língua italiana, "I genitori lasciano Marco senza risorse". Embaixo: Espectograma correspondente à onda acima indicando um tempo t e sua frequência f associada sendo que este espectograma foi obtido considerando-se uma discretização temporal de 2 ms 109
4.2 Acima: Indicação de regiões de alta e baixa enegia no espectograma. Embaixo: As regiões de frequência 80 Hz à 800 Hz, 800 Hz à 1500 Hz e 1500 Hz à 3000 Hz são mostradas no espectograma 109
4.3 Gráficos de Dispersão dos máximos uniformizados referentes às três regiões de frequência da língua inglesa 113
4.4 Gráficos de Dispersão dos máximos uniformizados referentes às três regiões de frequência da língua francesa 114
4.5 Gráfico de θ_2 versus θ_1 considerando-se os modelos de cópula M_1 e M_2 e as línguas inglesa e francesa 116
4.6 Figuras correspondentes aos "simplex" de duas e três variáveis 136

Lista de Tabelas

2.1 L-momentos de algumas distribuições 20

2.2 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.16) 28

2.3 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.17) 29

2.4 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.17) 30

4.5 Estatísticas provenientes de amostras de tamanho 15 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 51

4.6 Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 52

4.7 Estatísticas provenientes de amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 53

4.8 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 54

4.9 Estatísticas provenientes de amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 55

4.10 Estatísticas provenientes de amostras de tamanho 15 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$ 56

4.11 Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$ 57

4.12 Estatísticas provenientes de amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$ 58
4.13 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$ 59
4.14 Estatísticas provenientes de amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$ 60
4.15 Estimativas provenientes de 100 amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,7$ 61
4.16 Estimativas provenientes de 100 amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,7$ 62
4.17 Estimativas provenientes de 100 amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0$ e perturbação $\varepsilon = 0,15$ 63

2.1 Distribuição conjunta de probabilidade de (X_1, X_2, X_3, X_4) 92
2.2 Distribuição conjunta de probabilidade de X e Y 93
2.3 Distribuição conjunta de probabilidade de X e Y 93
2.4 Distribuição conjunta de probabilidade de X e Y 93
2.5 Distribuição conjunta de probabilidade de (Y_1, Y_2) 93

2.1 Testes de Kolmogorov-Smirnov e Wilcoxon para verificar o ajuste da Distribuição GEV aos dados relativos à língua inglesa 113
2.2 Testes de Kolmogorov-smirnov e Wilcoxon para verificar o ajuste da Distribuição GEV aos dados relativos à língua francesa 114
2.3 Estimativas dos parâmetros da Distribuição GEV correspondentes aos log-retornos extremos das energias nas três regiões de frequência da língua inglesa obtidas através dos Métodos dos L-momentos e Máxima Verossimilhança considerando-se um modelo autoregressivo de ordem 7 117
2.4 Estimativas dos parâmetros da Distribuição GEV correspondentes aos log-retornos extremos das energias nas três regiões de frequência da língua francesa obtidas através dos Métodos dos L-momentos e Máxima Verossimilhança considerando-se um modelo autoregressivo de ordem 7 118

2.5 Estimativas dos parâmetros dos modelos de cópula M_1 e M_2 obtidas através do Método de Máxima Verossimilhança considerando a uniformização dos máximos da língua inglesa pela aplicação da Distribuição GEV com os parâmetros estimados pelos métodos dos L-momentos e Máxima Verossimilhança 119
2.6 Estimativas dos parâmetros dos modelos de cópula M_1 e M_2 obtidas através do Método de Máxima Verossimilhança considerando a uniformização dos máximos da língua francesa pela aplicação da Distribuição GEV com os parâmetros estimados pelos métodos dos L-momentos e Máxima Verossimilhança 120
3.7 Teste de Wilcoxon entre os os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua inglesa 121
3.8 Distância máxima entre os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua inglesa 121
3.9 Teste de Wilcoxon e distância máxima entre os modelos de cópula M_1 e M_2 considerando-se a língua inglesa 122
3.10 Teste de Wilcoxon entre os os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua francesa 122
3.11 Distância máxima entre os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua francesa 123
3.12 Teste de Wilcoxon e distância máxima entre os modelos de cópula M_1 e M_2 considerando-se a língua francesa 123
4.13 Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 126
4.14 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$ 127
4.15 Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = -0,25$ 128
4.16 Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,55$ 129

Notação

v.a.: Variável aleatória

f.d.a.: Função de distribuição acumulada

i.i.d.: v.a. independentes e identicamente distribuídas

$\xrightarrow{d}$: Convergência em distribuição

$\stackrel{d}{=}$: Igualdade entre distribuições

$lim_{x\downarrow 0}h(x)$: Limite de $h(x)$ quando x tende ao zero pela direita

$f(x_1, ..., x_n) \uparrow x_i$: Função f crescente em x_i, $\forall x_j$, $i \neq j$

$f(x_1, ..., x_n) \downarrow x_i$: Função f decrescente em x_i, $\forall x_j$, $i \neq j$

x_{F_X}: Suporte da distribuição associada à variável aletaória X

$D(G)$: Domínio de atração de G

sup: Supremo

inf: Ínfimo

max: Máximo

Distribuição GEV: Distribuição Valor Extremo

LM: Método dos L-momentos

MV: Método de Máxima Verossimilhança

$x(G)$: Quantil da função de distribuição acumulada $G(x)$

$X_{i,n}$: Para uma amostra $\mathbf{X} = (X_1, ..., X_n)$, $X_{i,n}$ representa a i-ésima estatística de ordem da amostra $\mathbf{X}$ de tamanho n

Q1: Primeiro quartil

Q2: Mediana

Q3: Terceiro quartil

$p(\mathbf{x}|\theta)$: Probabilidade de $\mathbf{x}$ condicionado à um dado θ

$L(\theta|\mathbf{x})$: Função de verossimilhança, que depende de θ, condicionada à uma dada amostra $\mathbf{x}$

$Ln(x)$: Logaritmo Natural de x

$||$: Determinante

$B = A^T$: B é a matriz transposta de A

Capítulo 1

Caracterização da Convergência do Máximo de Variáveis Aleatórias

1.1 Introdução

Seja X_1, X_2, ... uma sequência de variáveis aleatórias (v.a.) independentes e identicamente distribuídas (i.i.d.) tendo função de distribuição F_X e consideremos M_n como sendo o máximo das primeiras n variáveis aleatórias, isto é, $M_n = max(X_1, ..., X_n)$.

A função de distribuição da variável aleatória M_n, $F_{M_n}(x)$, é dada por:

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = (F_X(x))^n$$

Mas, para n muito grande, a função de distribuição de M_n pode ser degenerada. Com isso, seria de grande utilidade algum resultado assintótico para o máximo.

A Teoria de Valores Extremos assegura a existência de uma distribuição assintótica não-degenerada, G, para uma transformação linear de M_n, isto é, para constantes apropriadas $a_n > 0$ e $b_n \in \mathbb{R}$ tem-se que:

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \tag{1.1}$$

ou, equivalentemente,

$$\lim_{n\to\infty} P(M_n \leq a_n x + b_n) = \lim_{n\to\infty} F_X^n(a_n x + b_n) = G(x)$$

para $x \in \mathbb{R}$ já que as possíveis funções de distribuição G (estas serão apresentadas na no Teorema) são funções contínuas em $\mathbb{R}$ [**?**, **?**].

Se (1.1) for verificado, F_X pertencerá ao domínio de atração do máximo associado à distribuição de valores extremos G, denotado por $F_X \in D(G)$), ou seja, a coleção das F_X tais que os máximos associados possuem uma distribuição limite não-degenerada é denominada domínio de atração da G.

Na próxima seção será obtido que a distribuição assintótica G é do tipo valor extremo, ou seja, ela pode ser a distribuição Weibull, Fréchet ou Gumbel. Para isto serão introduzidos alguns conceitos importantes para a formulação e demonstração do Teorema de Fisher-Tippett, que será feita na Seção 1.3 [**?**].

1.2 Preliminares

Para a identificação da distribuição assintótica não-degenerada G, que são possíveis leis limites para o máximo de uma sequência i.i.d. de variáveis aleatórias, define-se uma classe de distribuições denominadas max-estáveis. Mas, primeiramente, será apresentado o conceito e alguns resultados sobre a inversa de funções monótonas.

Definição 1.1. *Se $\psi(x)$ for uma função não-decrescente e contínua à direita, define-se a função inversa generalizada, ψ^{-1}, de $\psi(x)$, no intervalo (inf$\{\psi(x)\}$,sup$\{\psi(x)\}$), como sendo $\psi^{-1}(y)$=inf$\{x|\psi(x) \geq y\}$.*

Observa-se que a função de distribuição de uma v.a. é não-decrescente e contínua à direita. Assim, a inversa generalizada de uma função de distribuição G, $G^{-1}(y)$=inf$\{x|G(x) \geq y\}$, $0 < x < 1$, é chamada função quantil da função de distribuição G.

Lema 1.1.

1. *Para ψ definido acima, se a>0, $b \in \mathbb{R}$ e $c \in \mathbb{R}$ constantes e $H(x) = \psi(ax+b) - c$ tem-se que $H^{-1}(y) = a^{-1}(\psi^{-1}(y+c) - b)$;*
2. *Para ψ definido no item 1., se ψ^{-1} for contínua então $\psi^{-1}(\psi(x)) = x$;*
3. *Se G for uma função distribuição não-degenerada, existem $y_1 < y_2$ tais que $G^{-1}(y_1) < G^{-1}(y_2)$ são bem definidas e finita.*

Demostração:

1. Tem-se que $H^{-1}(y) = inf\{x|\psi(ax+b)-c \geq y\} = a^{-1}(inf\{(ax+b)|\psi(ax+b) \geq y+c\}-b) = a^{-1}(\psi^{-1}(y+c)-b)$ como requerido.

2. A partir da definição de ψ^{-1} é claro que $\psi^{-1}(\psi(x)) \leq x$. Se a desigualdade estrita for válida para algum x, a definição de ψ^{-1} mostra a existência de z<x com $\psi(z) \geq \psi(x)$ e, portanto, $\psi(z) = \psi(x)$ pois ψ é não-decrescente. Para y=$\psi(z)$=$\psi(x)$ tem-se que $\psi^{-1}(y) \leq z$, considerando para y>$\psi(z) = \psi(x)$ temos $\psi^{-1}(y) \geq x$, contrariando a continuidade de ψ^{-1}. Então, $\psi^{-1}(\psi(x)) = x$.

3. Se G for não-degenerada existem $x_1^{'} < x_2^{'}$ tais que $0 < G(x_1^{'}) = y_1 < G(x_2^{'}) = y_2 \leq 1$. Claramente, $x_1 = G^{-1}(y_1)$ e $x_2 = G^{-1}(y_2)$ são ambos bem definidos. Também, $G^{-1}(y_2) \geq x_1^{'}$ e a igualdade implicaria $G(z) \geq y_2$ para todo $z > x_1$ de modo que $G(x_1^{'}) = lim_{\varepsilon \downarrow 0} G(x_1^{'}+\varepsilon) = G(x_1^{'}+) \geq y_2$, contrariando $G(x_1^{'}) = y_1$. Então, $G^{-1}(y_2) > x_1^{'} \geq x_1 = G^{-1}(y_1)$, como requerido. □

Corolário 1.1. *Se G for uma função de distribuição não-degenerada, $a > 0$, α>0, $b \in \mathbb{R}$ e $\beta \in \mathbb{R}$ forem constantes tais que $G(ax+b) = G(\alpha x+\beta)$ para todo x, então $a = \alpha$ e $b = \beta$.*

Demonstração: Tome $y_1 < y_2$ e $-\infty < x_1 < x_2 < \infty$, pelo item 3. do Lema 1.1, de modo que $x_1 = G^{-1}(y_1)$ e $x_2 = G^{-1}(y_2)$. Considerando a inversa de $G(ax+b)$ e $G(\alpha x+\beta)$ tem-se que $a^{-1}(G^{-1}(y)-b) = \alpha^{-1}(G^{-1}(y)-\beta)$ para todo y, pelo item 1. do Lema 1.1. Aplicando isto a y_1 e y_2, dados acima, obtem-se $a^{-1}(x_1-b) = \alpha^{-1}(x_1-\beta)$ e $a^{-1}(x_2-b) = \alpha^{-1}(x_2-\beta)$, a partir das quais, segue-se que $a = \alpha$ e $b = \beta$. □

Teorema 1.1 (Khintchine)**.** *Seja $\{F_n\}$ uma sequência de funções de distribuição e G uma função de distribuição não-degenerada. Sejam $a_n > 0$ e $b_n \in \mathbb{R}$ constantes tais que*

$$F_n(a_n x + b_n) \xrightarrow{d} G(x) \tag{2.2}$$

Seja G_* uma função de distribuição não-degenerada tal que

$$F_n(\alpha_n x + \beta_n) \xrightarrow{d} G_*(x) \tag{2.3}$$

se e somente se

$$\begin{aligned} a_n^{-1}\alpha_n &\longrightarrow a \\ a_n^{-1}(\beta_n - b_n) &\longrightarrow b \end{aligned} \tag{2.4}$$

para algum $a > 0$ *e* $b \in \mathbb{R}$ *onde* $\alpha_n > 0$ *e* $\beta_n \in \mathbb{R}$*, então*

$$G_*(x) = G(ax + b) \tag{2.5}$$

Demonstração: Dados $a > 0$, $b \in \mathbb{R}$, $\alpha>0$, $\beta \in \mathbb{R}$ e considerando $\alpha_n^{'} = a_n^{-1}\alpha_n$, $\beta_n^{'} = a_n^{-1}(\beta_n - b_n)$ e $F_n^{'}(x) = F_n(a_n x + b_n)$ pode-se reescrever (2.2), (2.3) e (2.4), respectivamente, da seguinte maneira:

$$F_n^{'}(x) \xrightarrow{d} G(x), \tag{2.6}$$

$$\begin{aligned} F_n(\alpha_n x + \beta_n) = F_n(a_n\alpha_n^{'}x + (a_n\beta_n^{'} + b_n)) &= F_n(a_n(\alpha_n^{'}x + \beta_n^{'}) + b_n) = \\ &= F_n^{'}(\alpha_n^{'}x + \beta_n^{'}) \xrightarrow{d} G_*(x) \end{aligned} \tag{2.7}$$

e

$$\begin{aligned} \alpha_n^{'} &\longrightarrow a \\ \beta_n^{'} &\longrightarrow b \end{aligned} \tag{2.8}$$

para algum $a > 0$ e $b \in \mathbb{R}$.

Se (2.6) e (2.8) valem então (2.7) vale com $G_*(x) = G(ax + b)$ pois, por (2.7) e (2.8), $F_n^{'}(a_n x + b_n) \xrightarrow{d} G_*(x)$, e, por (2.6), $F_n^{'}(a_n x + b_n) \xrightarrow{d} G(a_n x + b_n)$. Então, (2.2) e (2.4) implicam em (2.3) e (2.5).

A prova estará completa se mostrarmos que (2.6) e (2.7) implicam em (2.8) para, então, (2.5) ser válida.

Como, por hipótese, G_* é não-degenerada, existem dois pontos $x^{'}$ e $x^{''}$ (que podem ser pontos de continuidade de G_*) tais que $0 < G_*(x^{'}) < 1$ e $0 < G_*(x^{''}) < 1$.

A sequência $\{\alpha_n^{'}x^{'} + \beta_n^{'}\}$ precisa ser limitada pois, se não fosse, uma subsequência $\{n_k\}$ poderia ser escolhida de modo que $\{\alpha_{n_k}^{'}x^{'} + \beta_{n_k}^{'}\} \longrightarrow \pm\infty$, o que, por (2.6), (já que G é uma função de distribuição) implicaria que o limite de $F_{n_k}^{'}(\alpha_{n_k}^{'}x^{'} + \beta_{n_k}^{'})$ seria zero ou um, contradizendo (2.7) para $x = x^{'}$. Então, $\{\alpha_n^{'}x^{'} + \beta_n^{'}\}$ é limitada e, similarmente, $\{\alpha_n^{'}x^{''} + \beta_n^{'}\}$ também é, o que mostra que as sequências $\{\alpha_n^{'}\}$ e $\{\beta_n^{'}\}$ são limitadas.

Dado que as sequências $\{\alpha_n^{'}\}$ e $\{\beta_n^{'}\}$ são limitadas, existem constantes $a > 0$ e $b \in \mathbb{R}$ e uma subsequência $\{n_k\}$ de inteiros tais que $\alpha_{n_k}^{'} \longrightarrow a$ e $\beta_{n_k}^{'} \longrightarrow b$ e, segue-se que, $F_{n_k}^{'}(\alpha_{n_k}^{'}x + \beta_{n_k}^{'}) \xrightarrow{d} G(ax + b)$ donde, por (2.7), $G(ax + b) = G_*(x)$.

Por outro lado, se houvesse uma outra subsequência $\{m_k\}$ de inteiros tais que $\alpha_{m_k}^{'} \longrightarrow a^{'} > 0$ e $\beta_{m_k}^{'} \longrightarrow b^{'} \in \mathbb{R}$, ter-se-ia, pelo Corolário 1.1, $G(a^{'}x + b^{'}) = G_*(x) = G(ax + b)$ e, portanto, $a^{'} = a$ e $b^{'} = b$. Então, $\alpha_n^{'} \longrightarrow a$ e $\beta_n^{'} \longrightarrow b$ como requerido para completar a prova. □

Definição 1.2. *Seja uma sequência de variáveis aleatórias X_1, .., X_n i.i.d. e seja X uma variável aleatória não-degenerada e i.i.d. com a seqüência. A variável aleatória X é chamada max-estável se existirem constantes apropriadas $a_n > 0$, $b_n \in \mathbb{R}$ para cada $n \geq 2$ que satisfaça max(X_1, ..., X_n)$\stackrel{d}{=} a_nX + b_n$.*

Definição 1.3. *Duas funções de distribuição G_1 e G_2 são do mesmo tipo se $G_2(x) = G_1(ax + b)$ para algumas constantes $a>0$ e $b \in \mathbb{R}$.*

Pelas definições 1.2 e 1.3, pode-se dizer que uma função de distribuição G não-degenerada é max-estável se, para cada $n \geq 2$ existirem constantes $a_n > 0$ e $b_n \in \mathbb{R}$ tais que $G^n(a_nx + b_n) = G(x)$, ou seja, a função de distribuição G^n é do mesmo tipo que G.

Definição 1.4. *Se $F_X^n(a_nx + b_n) \xrightarrow{d} G(x)$ valer para alguma sequência $\{a_n > 0\}$ e $\{b_n \in \mathbb{R}\}$, F_X pertencerá ao domínio de atração de G, $F_X \in D(G)$.*

Teorema 1.2. *1. Uma função de distribuição não-degenerada G é max-estável se e somente se existir uma sequência $\{F_n\}$ de funções de distribuição e constantes $a_n > 0$ e $b_n \in \mathbb{R}$ tais que*

$$F_n(a_{nk}x + b_{nk}) \xrightarrow{d} G^{\frac{1}{k}}(x) \tag{2.9}$$

quando $n \longrightarrow \infty$ para cada k = 1, 2, ...

2. *Em particular, se G for não-degenerada, $D(G)$ será não-vazio se e somente se G for max-estável, $G \in D(G)$. Desta maneira, a classe das funções de distribuição não-degeneradas G que aparecem como limite em lei, em (1.1) (para X_1, X_2, ... i.i.d), coincide com a classe das funções de distribuição max-estáveis.*

Demonstração:

1. Se G for não-degenerada, $G^{\frac{1}{k}}$ também será, para cada $k = 1, 2, \ldots$. Se (2.9) valer para cada k, o Teorema 1.1 implicará que $G^{\frac{1}{k}}(x) = G(\alpha_k x + \beta_k)$ para algum $\alpha_k > 0$ e $\beta_k \in \mathbb{R}$ e, assim, G será max-estável. Reciprocamente, se G for max-estável, $G^n(a_n x + b_n) = G(x)$ para algum $a_n > 0$ e $b_n \in \mathbb{R}$ e, considerando $F_n = G^n$,

$$F_n(a_{nk}x + b_{nk}) = (G^{nk}(a_{nk}x + b_{nk}))^{\frac{1}{k}} = (G(x))^{\frac{1}{k}}$$

de forma que (2.9) segue trivialmente.

2. Se G for max-estável, $G^n(a_n x + b_n) = G(x)$ para algum $a_n > 0$ e $b_n \in \mathbb{R}$ e, quando $n \longrightarrow \infty$, $G \in D(G)$. Reciprocamente, se $D(G)$ for não-vazio, $F \in D(G)$, ou seja, $F^n(a_n x + b_n) \xrightarrow{d} G(x)$. Então, $F^{nk}(a_{nk}x + b_{nk}) \xrightarrow{d} G(x)$ ou $F^n(a_{nk}x + b_{nk}) \xrightarrow{d} G^{\frac{1}{k}}(x)$. Assim (2.9) vale com $F_n = F^n$ e, assim, G é max-estável. □

Corolário 1.2. *Se G for max-estável, existem funções reais $g_1(s) > 0$ e $g_2(s) \in \mathbb{R}$ definidas para $s > 0$ tais que*

$$G^s(g_1(s)x + g_2(s)) = G(x) \tag{2.10}$$

para todos reais $x, s > 0$.

Demonstração: Dado que G é max-estável, existem $a_n > 0$ e b_n tais que

$$G^n(a_n x + b_n) = G(x) \tag{2.11}$$

e, desta forma, $G^{[ns]}(a_{[ns]}x + b_{[ns]}) = G(x)$ onde [.] denota a parte inteira. Disto é facilmente visto (por exemplo, tomando logaritmo) que:

$$G^n(a_{[ns]}x + b_{[ns]}) \xrightarrow{d} G^{\frac{1}{s}}(x) \tag{2.12}$$

Pelos limites (2.12) e (2.11) e desde que $G^{\frac{1}{s}}$ seja não-degenerada, o Teorema 1.1 pode ser aplicado, com $\alpha_n = a_{[ns]}$ e $\beta_n = b_{[ns]}$, para mostrar que $G(g_1(s)x+g_2(s)) = G^{\frac{1}{s}}(x)$ para algum $g_1(s) > 0$ e $g_2(s) \in \mathbb{R}$, como desejado. □

Definição 1.5. *Seja F_X a função de distribuição de uma v.a. . O limite superior x_{F_X} do suporte da distribuição F_X é definido com $x_{F_X} = sup\{x \in \mathbb{R} | F_X(x) < 1\}$.*

Teorema 1.3. *Toda distribuição max-estável possui uma das seguintes estruturas funcionais citadas a seguir* [1]*:*

Tipo I (Distribuição Gumbel):

$$G(x) = exp(-e^{-x}) \quad para \quad -\infty < x < \infty$$

Tipo II (Distribuição Fréchet):

$$G(x) = exp(-x^{-\alpha}) \quad para\ algum \quad \alpha > 0 \quad e\ para \quad x > 0$$

Tipo III (Distribuição Weibull):

$$G(x) = exp(-(-x)^{\alpha}) \quad para\ algum \quad \alpha > 0 \quad e\ para \quad x \leq 0$$

[1] As distribuições do Tipo I, II e III são denominadas Distribuição Valor Extremo Generalizada (GEV).

Reciprocamente, as distribuições do Tipo I, II e III são max-estável.

Demonstração: A recíproca segue facilmente. Para o Tipo I temos:

Dados $a_n > 0$ e $b_n \in \mathbb{R}$, deve ser verificada a relação $G^n(a_n x + b_n) = G(x)$. De fato,

$$(exp(-e^{-(a_n x + b_n)}))^n = exp(-e^{-(a_n x + b_n - Ln(n))})$$

Então, tomando $a_n = 1$ e $b_n = Ln(n)$ verifica-se a definição de distribuição max-estável. Similarmente, temos a recíproca para os Tipos II e III.

Agora, provaremos a afirmação direta:

Se G for max-estável então (2.10) será valida para todo $s > 0$ e para todo x. Se $0 < G(x) < 1$, (2.10) fornece $-sLn(G(a(s)x + b(s))) = -Ln(G(x))$ donde $-Ln(-Ln(G(a(s)x + b(s)))) - Ln(s) = -Ln(-Ln(G(x)))$.

Observa-se, a partir da propriedade max-estável com $n = 2$, que G não pode ter saltos em algum limite superior (ou inferior) finito do suporte da distribuição G. Então a função não-decrescente $\psi(x) = -Ln(-Ln(G(x)))$ é tal que $inf\{\psi(x)\}$= -∞, $sup\{\psi(x)\}$=∞ e, desta forma, existe uma função inversa $U(y)$ definida para todo real y. Além do mais, $\psi(a(s)x + b(s)) - Ln(s) = \psi(x)$. Então, pelo item (i) do Lema 1.1, temos:

$$(\psi(a(s)x + b(s)) - Ln(s))^{-1} = \psi^{-1}(x) \Longrightarrow \frac{\psi^{-1}(y + Ln(s)) - b(s)}{a(s)} = \psi^{-1}(x)$$

isto é,

$$\frac{U(y + Ln(s)) - b(s)}{a(s)} = U(y)$$

Efetuando $U(y) - U(0)$ temos:

$$\begin{aligned} U(y)-U(0) &= \frac{U(y+Ln(s))-b(s)}{a(s)} - \frac{U(Ln(s))-b(s)}{a(s)} = \\ &= \frac{U(y+Ln(s))-U(Ln(s))}{a(s)} = \frac{U(y+Ln(s))-U(Ln(s))+U(0)-U(0)}{a(s)} \end{aligned}$$

Considerando $z = Ln(s)$, $\tilde{a} = a(e^z)$ e $\tilde{U}(y) = U(y) - U(0)$ temos:

$$\tilde{U}(y+z) - \tilde{U}(z) = \tilde{U}(y)\tilde{a}(z) \tag{2.13}$$

para todo y, z reais.

Agora, analisaremos dois casos:

Caso 1: $\tilde{a}(z) = 1$ para todo z.

Neste caso, de (2.13) obtemos $\tilde{U}(y+z) = \tilde{U}(y) + \tilde{U}(z)$.

A única solução monótona crescente é dada por $\tilde{U}(y) = \rho y$ para algum $\rho>0$, ou seja, $U(y) - U(0) = \rho y$ ou $\psi^{-1}(y) = U(y) = \rho y + \nu$, $\nu = U(0)$. Dado que a expressão ψ^{-1} é contínua, pelo item (ii) do Lema 1.1, temos x$=\psi^{-1}(\psi(x)) = \rho\psi(x) + \nu$ ou $\psi(x) = \frac{x-\nu}{\rho}$ que fornece:

$$G(x) = exp\left(-e^{-\frac{(x-\nu)}{\rho}}\right)$$

quando $0 < G(x) < 1$ pois $\psi(x) = -Ln(-Ln(G(x))$.

Como notado anteriormente, G não pode ter saltos em algum limite superior (ou inferior) finito do suporte da distribuição G. Desta forma, para todo x, a forma supra-citada (Tipo I) existe.

Caso 2: $\tilde{a}(z) \neq 1$ para algum z.

Neste caso, intercalando y e z em (2.13) e subtraindo (2.13), obtemos:

$$\tilde{U}(y+z) - \tilde{U}(y+z) + \tilde{U}(z) - \tilde{U}(y) = -\tilde{U}(y)\tilde{a}(z) + \tilde{U}(z)\tilde{a}(y)$$

isto é,

$$\tilde{U}(y)(1-\tilde{a}(z)) = \tilde{U}(z)(1-\tilde{a}(y)) \tag{2.14}$$

Assim, (2.14) fornece:

$$\tilde{U}(y) = \frac{\tilde{U}(z)}{1-\tilde{a}(z)}(1-\tilde{a}(y)) = c(1-\tilde{a}(y)) \tag{2.15}$$

onde c=$\frac{\tilde{U}(z)}{1-\tilde{a}(z)} \neq 0$ (dado que $\tilde{U}(z) = 0$ implicaria $\tilde{U}(y) = 0$ para todo y e, então, $U(y) = U(0)$, constante).

Avaliando a expressão (2.13) de acordo com (2.15) obtemos:

$$\tilde{U}(y+z) - \tilde{U}(z) = \tilde{U}(y)\tilde{a}(z)$$

isto é,

$$c(1-\tilde{a}(y+z)) - c(1-\tilde{a}(z)) = c(1-\tilde{a}(y))\tilde{a}(z)$$

A expressão acima fornece:

$$\tilde{a}(y+z) = \tilde{a}(y)\tilde{a}(z)$$

Mas $\tilde{a}$ é monótona (por (2.15)) e, então, a única solução monótona não-constante da equação acima possui a forma $\tilde{a}(y) = e^{\rho y}$ para $\rho \neq 0$. Então, (2.13) fornece $\psi^{-1}(y) = U(y) = \nu + c(1-e^{\rho y})$ com ν=U(0).

Dado que $-Ln(-Ln(G(x)))$ é crescente então U também é. Além disso, temos que $c < 0$ se $\rho > 0$ e $c > 0$ se ρ<0.

Pelo item (ii) do Lema 1.1 obtemos:

$$x = \psi^{-1}(\psi(x)) = \nu + c(1-e^{\rho\psi(x)}) = \nu + c(1-(-Ln(G(x)))^{-\rho})$$

fornecendo

$$G(x) = exp\left(-\left(1-\frac{x-\nu}{c}\right)^{-\frac{1}{\rho}}\right)$$

com $0 < G(x) < 1$.

Novamente, a partir da continuidade da G em algum limite finito do seu suporte da distribuição, tem-se que G é do Tipo II ou do Tipo III, com $\alpha = \frac{1}{\rho}$ ou $\alpha = -\frac{1}{\rho}$, de acordo, respectivamente, com

$\rho > 0$(c<0) ou $\rho < 0$(c>0). □

1.3 Teorema de Fisher-Tippett e sua Implicação

Teorema 1.4. *(Teorema de Fisher-Tippett) Seja $M_n = max(X_1, ..., X_n)$ onde X_i, $i = 1, ..., n$, são variáveis aleatórias i.i.d. . Se para algumas constantes (em n) $a_n > 0$ e $b_n \in \mathbb{R}$, temos*

$$P(a_n^{-1}(M_n - b_n) \leq x) \xrightarrow{d} G(x) \tag{3.16}$$

*para alguma G não-degenerada, então G é um dos três tipos citados no Teorema 1.3. Reciprocamente, cada função de distribuição G do tipo valor extremo aparece como um dos limites em (3.16) e, de fato, aparece quando G é, ela mesma, a função de distribuição de cada X_i [**?**, **?**].*

Demonstração: Se (3.16) vale, o Teorema 1.2 mostra que G é max-estável e, então, pelo Teorema 1.3, é do tipo valor extremo. Reciprocamente, se G é do tipo valor extremo, ela é max-estável, pelo Teorema 1.3, e, assim, o item (ii) do Teorema 1.2 mostra que $G \in$ D(G). □

Pelo Teorema 1.4, pode-se, então, estimar a distribuição assintótica de $\frac{M_n - b_n}{a_n}$ diretamente da família G sem fazer-se alguma referência à distribuição de X pois temos que a distribuição G (que corresponde à distribuição dos máximos $M_n = max(X_1, ..., X_n)$ onde X_i, i=1, ..., n, são variáveis aleatórias i.i.d.) é uma das três dadas no Teorema 1.3.

A expressão seguinte incorpora as três distribuições de valor extremo:

$$G(x) = \begin{cases} exp\left(-(1+\alpha x)^{-\frac{1}{\alpha}}\right) & \alpha \neq 0 \\ \\ exp(-exp(-x)) & \alpha = 0 \end{cases}$$

onde $1 + \alpha x > 0$.

Pode-se, também, utilizar a parametrização $\alpha^{'} = -\alpha$ e, assim, obter:

$$G(x) = \begin{cases} exp\left(-(1-\alpha^{'}x)^{\frac{1}{\alpha^{'}}}\right) & \alpha^{'} \neq 0 \\ \\ exp(-exp(-x)) & \alpha^{'} = 0 \end{cases}$$

onde $1 - \alpha^{'}x > 0$.

Pode-se, ainda, incorporar parâmetros de locação e escala na distribuição de valor extremo substituindo-se x por $\frac{x-\mu}{\sigma}$ onde $\mu \in \mathbb{R}$ e $\sigma > 0$. Assim, a família de locação-escala correspondente é dada por

$$G(x) = \begin{cases} exp\left(-(1-\alpha^{'}(\frac{x-\mu}{\sigma}))^{\frac{1}{\alpha^{'}}}\right) & \alpha^{'} \neq 0 \\ \\ exp(-exp(-\frac{x-\mu}{\sigma})) & \alpha^{'} = 0 \end{cases}$$

onde $1 - \alpha^{'}(\frac{x-\mu}{\sigma}) > 0$, $\mu \in \mathbb{R}$ e $\sigma > 0$.

A condição $1-\alpha^{'}(\frac{x-\mu}{\sigma}) > 0$ fornece que x é limitado superiormente e inferiormente por $\mu+\sigma/\alpha^{'}$, respectivamente, se $\alpha^{'} > 0$ e $\alpha^{'} < 0$.

No decorrer do texto será utilizada a família locação-escala da GEV supra-citada.

O Teorema de Fisher-Tippett fornece a distribuição limite para o máximo coletado em blocos de tamanho n. Sendo $x_1, ..., x_m$ as realizações da variável X, considera-se blocos de tamanho k desta amostra, isto é, os dados amostrais são particionados em b blocos tais que $bk \leq m$. Definimos, então:

$$\begin{aligned} M_1 &= max\{x_1, ..., x_k\} \\ M_2 &= max\{x_{k+1}, ..., x_{2k}\} \\ &\vdots \\ M_b &= max\{x_{bk-k+1}, ..., x_{bk}\} \end{aligned}$$

Através dessa nova amostra, $M_1, ..., M_b$, pode-se obter os estimadores de $\hat{\alpha}$, $\hat{\mu}$ e $\hat{\sigma}$ e, assim, determinar a distribuição extremal G.

Figura 1.1: Distribuições Weibull, Fréchet e Gumbel geradas considerando-se os parâmetros $\mu = 0$, $\sigma = 1$ e, respectivamente, $\alpha = 0,5$, α= - 0,5 e α=0

Figura 1.2: Distribuições Weibull, Fréchet e Gumbel geradas considerando-se os parâmetros $\mu = 0$, $\sigma = 1$ e, respectivamente, α=2, α= -2 e α=0

Capítulo 2

Estimação

2.1 Introdução

O Teorema de Fisher-Tippett fornece a distribuição assintótica G para o máximo de variáveis aleatórias independentes e com a mesma distribuição. A distribuição G depende dos parâmetros de locação μ, escala σ e forma α. Portanto, estes parâmetros devem ser estimados a fim de obter-se $G_{\hat{\mu},\hat{\sigma},\hat{\alpha}}(x) = G(x|\hat{\mu}, \hat{\sigma}, \hat{\alpha})$.

Diversos tipos de inferência podem ser realizados através de $G_{\hat{\alpha},\hat{\mu},\hat{\sigma}}(x)$ como, por exemplo, a predição e cálculo de quantis que estabeleçam medidas de risco na ocorrência de eventos extremos.

Este capítulo aborda dois métodos de estimação de parâmetros: Método de Máxima Verossimilhança [**?**, **?**, **?**, **?**] e o Método dos L-Momentos [**?**, **?**].

Com a finalidade de compará-los, serão geradas amostras de uma Distribuição GEV e serão obtidas as estimativas fornecidas por ambos os métodos. Também, serão mencionadas as condições que garantem as propriedades assintóticas dos estimadores de máxima verossimilhança e casos em que tais propriedades assintóticas ficam comprometidas, em particular, o caso da Distribuição GEV sendo que este comprometimento ocorre para outras distribuições. Com isso, o Método dos L-Momentos é um método alternativo de estimação que possui normalidade assintótica.

2.2 L-Momentos de uma Distribuição de Probabilidade

Um problema comum em Estatística é a estimação pontual na Estatística Clássica, a partir de uma amostra de tamanho n, dos parâmetros de uma dada distribuição de probabilidade cuja especificação envolve um número finito p de parâmetros desconhecidos. Analogamente ao método dos momentos usuais, o método dos L-momentos obtém estimadores pela igualdade dos primeiros p L-momentos amostrais às correspondentes quantidades populacionais.

Definição 2.1. *Seja $(X_1, ...X_n)$ uma amostra aleatória da variável aletória real X com função de distribuição acumulada $G(x)$ e função quantil $x(G)$, e sejam $X_{1,n}, ..., X_{n,n}$ as estatísticas de ordem obtidas a partir da distribuição de X. Define-se o L-momento de ordem r de X, λ_r, como:*

$$\lambda_r := \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k C_{r-1,k}\mathbb{E}(X_{r-k,r}) \tag{2.1}$$

para $r = 1, 2, ...$ onde $C_{r,k} = \begin{pmatrix} r \\ k \end{pmatrix}$.

Observe que λ_r é uma função linear das esperanças das estatísticas de ordem. A esperança das estatísticas de ordem pode ser escrita como:

$$\mathbb{E}(X_{j,r}) = \frac{r!}{(j-1)!(r-j)!}\int_0^1 x[G(x)]^{j-1}[1-G(x)]^{r-j}dG \tag{2.2}$$

Substituindo a expressão (2.2) em (2.1), tem-se:

$$\lambda_r = \int_0^1 x(G)P^*_{r-1}(G)dG \tag{2.3}$$

$$P^*_r(G) = \sum_{k=0}^{r} p^*_{r,k}G^k \tag{2.4}$$

$$p^*_{r,k} = (-1)^{r-k}C_{r,k}C_{r+k,k} \tag{2.5}$$

com r = 1, 2, ...

Por exemplo, utilizando as equações (2.3), (2.4) e (2.5) tem-se que os quatro primeiros L-momentos são dados, respectivamente, por:

$$
\begin{aligned}
\lambda_1 &= \mathbb{E}(X) = \int_0^1 x(G)dG \qquad (2.6)\\
\lambda_2 &= \frac{1}{2}\mathbb{E}(X_{2,2} - X_{1,2}) = \int_0^1 x(G)(2G-1)dG\\
\lambda_3 &= \frac{1}{3}\mathbb{E}(X_{3,3} - 2X_{2,3} + X_{1,3}) = \int_0^1 x(G)(6G^2 - 6G + 1)dG\\
\lambda_4 &= \frac{1}{4}\mathbb{E}(X_{4,4} - 3X_{3,4} + 3X_{2,4} - X_{1,4}) = \int_0^1 x(G)(20G^3 - 30G^2 + 12G - 1)dG
\end{aligned}
$$

Os momentos λ_1 e λ_2 podem ser considerados como medidas, respectivamente, de locação e escala.

O uso dos L-momentos para descrever distribuições de probabilidade é justificado pelo teorema a seguir:

Teorema 2.1.

1. *Os L-momentos, λ_r, $r = 1, 2, ...$, da variável aletória X existem se e somente se X tiver esperança finita.*

2. *Uma distribuição cuja esperança exista é caracterizada pelos L-momentos {λ_r:r=1, 2, ...}.*

Demonstração: [?]

1. A esperança finita de X implica na esperança finita de todas as estatísticas de ordem e, assim, o resultado segue imediatamente.

2. Primeiramente, será mostrado que a distribuição é caracterizada pelo conjunto {E($X_{r,r}$), r=1, 2, ...}.

Sejam X e Y variáveis aleatórias com distribuições acumuladas F e G e funções quantis $x(u)$ e $y(u)$, respectivamente.

Considere $\xi_r^X := E(X_{r,r}) = r\int x[F(x)]^{r-1}dF$ e $\xi_r^Y := E(Y_{r,r}) = r\int x[G(x)]^{r-1}dG$.

Então:

$$\xi_{r+2}^X - \xi_{r+1}^X = \int_0^1 [(r+2)u^{r+1} - (r+1)u^r]x(u)du \tag{2.7}$$

Considerando $g = x(u)$ e $df = (r+2)u^{r+1} - (r+1)u^r]du$ sendo, respectivamente, $dg = dx(u)$ e $f = u^{r+2} - u^{r+1}$ tem-se, utilizando integração por partes em (2.7),

$$\int_0^1 [(r+2)u^{r+1} - (r+1)u^r]x(u)du = (u^{r+2} - u^{r+1})x(u)|_0^1 - \int_0^1 (u-1)u^{r+1}dx(u) \tag{2.8}$$

Como $(u^{r+2} - u^{r+1})x(u)|_0^1 = 0$, a expressão (2.8) torna-se

$$\int_0^1 [(r+2)u^{r+1} - (r+1)u^r]x(u)du = -\int_0^1 (u-1)u^{r+1}dx(u) = \int_0^1 u^r u(1-u)dx(u) \tag{2.9}$$

Da expressão (2.9) tem-se

$$\xi_{r+2}^X - \xi_{r+1}^X = \int_0^1 u^r u(1-u)dx(u) = \int_0^1 u^r dz_X(u)$$

onde $z_X(u)$, definido por $dz_X(u) = u(1-u)dx(u)$, é uma função crescente em (0,1).

Se $\xi_r^X = \xi_r^Y$, $r = 1, 2, ...$, então $\int_0^1 u^r dz_X(u) = \int_0^1 u^r dz_Y(u)$, $r = 0, 1, ...$. Então, z_X e z_Y são distribuições que possuem os mesmos momentos no intervalo finito (0,1). Consequentemente, $z_X = z_Y$ implicando que $x(u) = y(u)$. Com isso a distribuição com esperança finita é caracterizada pela sequência $\{\xi_r : r = 1, 2, ...\}$.

Utilizando (2.3), tem-se:

$$\lambda_r = \sum_{k=1}^{r} p_{r-1,k-1}^* k^{-1}\xi_k \tag{2.10}$$

e

$$\xi_r = \sum_{k=1}^{r} \frac{(2k-1)r!(r-1)!}{(r-k)!(r-1+k)!}\lambda_k \tag{2.11}$$

Portanto, pelas equações (2.10) e (2.11), uma dada sequência de λ_r determina uma única sequência de ξ_r. □

Com o resultado anterior, uma distribuição pode ser especificada pelos L-momentos mesmo se alguns dos momentos convencionais não existir, sendo que tal especificação é sempre única.

Para uma v. a. contínua com função densidade $f(x)$, os momentos usuais de ordem r centrados em uma constante a e os momentos centrais usuais de ordem r são dados, respectivamente, por $\mu_r^{'} = \int_{-\infty}^{\infty}(x-a)^r f(x)dx$ e $\mu_r = \int_{-\infty}^{\infty}(x-\mu_r^{'})^r f(x)dx$. Para $a = 0$ e $r = 1$ temos $\mu_1^{'} = \int_{-\infty}^{\infty} xf(x)dx$ que corresponde à esperança de uma v.a. e, para $r = 2$ temos $\mu_2 = \int_{-\infty}^{\infty}(x-\mu_1^{'})^2 f(x)dx$ que corresponde à variância de uma v.a. . Para v.a. discretas, basta trocar, nas definições acima de $\mu_r^{'}$ e μ_r, a integral pelo somatório.

Através dos momentos $\mu_r^{'}$ e μ_r, define-se os coeficientes de assimetria (medida da simetria de uma distribuição) e curtose (medida do grau de achatamento de uma distribuição) respectivamente, por $\beta_1^2 = \frac{\mu_3^2}{\mu_2^3}$ e $\beta_2 = \frac{\mu_4}{\mu_2^2}$. Os coeficientes de assimetria e curtose amostrais são dados, considerando-se uma amostra aleatória $x_1, ..., x_n$ da v.a. X, respectivamente, por $b_1^2 = \frac{m_3^2}{m_2^3}$ e $b_2 = \frac{m_4}{s^4}$ onde $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$ e $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ [**?**].

Também, define-se o coeficiente de variação, como sendo a razão entre o desvio padrão e a esperança, que é uma medida de dispersão utilizada na comparação da dispersão entre diferentes populações pois é uma medida independente de escala [**?**].

Pode-se definir razão de L-momentos como sendo $\tau_r = \frac{\lambda_r}{\lambda_2}$, $r = 3, 4, ...$. Também, é possível definir a razão $\tau = \frac{\lambda_2}{\lambda_1}$ sendo, tal quantidade, análoga ao coeficiente de variação. As quantidades τ_r e τ são limitadas como mostra o teorema seguinte.

Teorema 2.2. *Seja X uma variável aleatória não degenerada com esperança finita. Então, as razões de L-momentos de X satisfazem* $|\tau_r| < 1$ *para* $r \geq 3$. *E, se* $X \geq 0$ *então* τ *satisfaz* $0<\tau<1$.

Demostração: Para a demonstração consulte [**?**].

Tabela 2.1: L-momentos de algumas distribuições

Distribuição	G(x) ou x(G)	λ_1 e λ_2	τ_3 e τ_4
Uniforme	$x = \alpha + (\beta - \alpha)G$	$\lambda_1 = \frac{1}{2}(\alpha + \beta)$, $\lambda_2 = \frac{1}{6}(\beta - \alpha)$	$\tau_3 = 0, \tau_4 = 0$
Exponencial	$x = \xi - \alpha Ln(1 - G)$	$\lambda_1 = \xi + \alpha$, $\lambda_2 = \frac{\alpha}{2}$	$\tau_3 = \frac{1}{3}, \tau_4 = \frac{1}{6}$
Logística	$x = \xi + \alpha Ln(\frac{G}{1-G})$	$\lambda_1 = \xi$, $\lambda_2 = \alpha$	$\tau_3 = 0, \tau_4 = \frac{1}{6}$
Normal	$G = \Phi(\frac{x-\mu}{\sigma})$	$\lambda_1 = \mu$, $\lambda_2 = \pi^{-1}\sigma$	$\tau_3 = 0$, $\tau_4 = 30\pi^{-1}\tan^{-1}\sqrt{2} - 9 = 0{,}1226$

Os L-momentos $\lambda_1, ..., \lambda_r$ e as razões de L-momentos $\tau_1, ..., \tau_r$ são quantidades usadas para resumir uma distribuição. Os L-momentos são análogos aos momentos centrais e a razão de L-momentos é análoga à razão de momentos. Em particular, λ_1, λ_2, τ_3 e τ_4 podem ser tomados como medidas de locação, escala, assimetria e curtose, respectivamente. Já a quantidade τ_5 pode ser interpretada como uma medida de tendência à bimodalidade de uma distribuição [**?**].

Na prática, os L-momentos precisam ser estimados a partir de uma amostra aleatória de uma distribuição desconhecida. Pelo fato de λ_r ser uma função da esperança das estatísticas de ordem de uma amostra de tamanho r, é natural estimá-los pela U-estatística [**?**, **?**], isto é, a correspondente função da média das estatísticas de ordem amostrais sobre todas as sub-amostras de tamanho r, que podem ser construídas a partir da amostra observada de tamanho n. A U-estatística é definida

por

$$U = \frac{1}{\begin{pmatrix} n \\ r \end{pmatrix}} \sum_{1 \leq i_1 <} \sum_{i_2 <} ... \sum_{< i_r \leq n} \phi(X_{i_1}, ..., X_{i_r})$$

onde $\phi(X_{i_1}, ..., X_{i_r})$ representa o núcleo da U-estatística. Então, considerando-se a amostra $x_1, x_2, ..., x_n$ e a amostra ordenada $x_{1,n} \leq x_{2,n} \leq ... \leq x_{n,n}$ e considerando-se o núcleo da U-estatística como sendo a estatística de ordem $x_{i_{r-k},n}$, define-se o r-ésimo L-momento amostral como sendo

$$l_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k C_{r-1,k} C_{n,r}^{-1} \sum_{1 \leq i_1 <} \sum_{i_2 <} ... \sum_{< i_r \leq n} x_{i_{r-k},n}$$

com $r = 1, 2, ..., n$.

Em particular,

$$\begin{aligned}
l_1 &= n^{-1} \sum_i x_i \\
l_2 &= \frac{1}{2} C_{n,2}^{-1} \sum_{i>} \sum_j (x_{i,n} - x_{j,n}) \\
l_3 &= \frac{1}{3} C_{n,3}^{-1} \sum_{i>} \sum_{j>} \sum_k (x_{i,n} - 2x_{j,n} + x_{k,n}) \\
l_4 &= \frac{1}{4} C_{n,4}^{-1} \sum_{i>} \sum_{j>} \sum_{k>} \sum_l (x_{i,n} - 3x_{j,n} + 3x_{k,n} - x_{l,n})
\end{aligned}$$

As U-estatísticas são não viciadas e possuem propriedades como normalidade assintótica (veja o anexo sobre U-estatísticas) e moderada resistência à influência de " outliers " [**?**].

Para calcular l_r, não é necessário iterar sobre todas as sub-amostras de tamanho r. A estatística pode ser expressa, explicitamente, como uma combinação linear das estatísticas de ordem da amostra de tamanho n [**?**]. Considerando os estimadores da U-estatística de $\mathbb{E}(X_{r,r})$, mostra-se que o seu estimador pode ser escrito como rb_{r-1} onde:

$$b_r = n^{-1} \sum_{i=1}^{n} \frac{(i-1)(i-2)...(i-r)}{(n-1)(n-2)...(n-r)} x_{i,n}$$

Então, por (**??**), temos:

$$l_r = \sum_{k=0}^{r-1} p^*_{r-1,k} b_k$$

A distribuição amostral exata dos L-momentos é difícil de obter. É comum aplicar-se a teoria assintótica à distribuição dos L-momentos amostrais e à razão de L-momentos como mostra o seguinte teorema.

Teorema 2.3. *Considere X uma variável aleatória real com distribuição acumulada G e λ_r os L-momentos com variância finita. Sejam l_r, $r = 1, 2, ..., m$, os L-momentos amostrais calculados através de uma amostra aleatória de tamanho n e $\tau_r = \frac{\lambda_r}{\lambda_2}$ e $t_r = \frac{l_r}{l_2}$, $r = 3, 4, ..., m$. Então, quando $n \to \infty$:*

1. *$n^{1/2}(l_r - \lambda_r)$, $r = 1, 2, ..., m$, converge em distribuição para a distribuição normal multivariada $N(0, \Lambda)$ cujos elementos Λ_{rs}, $r, s = 1, 2, ..., m$, de Λ são dados por:*

$$\Lambda_{rs} = \iint_{x<y} [P^*_{r-1}(G(x))P^*_{s-1}(G(y)) + P^*_{s-1}(G(x))P^*_{r-1}(G(y))]G(x)(1 - G(y))dxdy$$

   onde P^*_r é dado por (2.4).

2. *O vetor $n^{1/2}[(l_1 - \lambda_1)(l_2 - \lambda_2)(t_3 - \tau_3)(t_4 - \tau_4)...(t_m - \tau_m)]^t$ converge em distribuição para a distribuição normal multivariada $N(0, T)$ onde os elementos T_{rs}, $r, s = 1, 2, .., m$, de T são dados por:*

$$T_{rs} = \begin{cases} \Lambda_{rs} & r \leq 2, s \leq 2 \\ \frac{\Lambda_{rs} - \tau_r \Lambda_{2s}}{\lambda_2} & r \geq 3, s \leq 2 \\ \frac{\Lambda_{rs} - \tau_r \Lambda_{2s} - \tau_s \Lambda_{2r} + \tau_r \tau_s \Lambda_{22}}{\lambda_2^2} & r \geq 3, s \geq 3 \end{cases}$$

Demonstração: Para a demonstração consulte [**?**].

As correspondentes versões amostrais dos coeficientes de assimetria e curtose calculados por L-momentos são dadas, respectivamente, por $t_3 = \frac{l_3}{l_2}$ e $t_4 = \frac{l_4}{l_2}$.

Hosking [**?**] verificou a relação entre os coeficiente de curtose β_2 e $\tau_4 = \frac{\lambda_4}{\lambda_2}$ e o poder do Teste de Shapiro-Wilk [**?**]. Nesse artigo, Hosking considera vários tipos de distribuições, como N(0,1), t_{10}, Laplace(0,1), Logística(0,1) entre outras, com diferentes valores do coeficiente de curtose β_2 e τ_4, como distribuições alternativas à normal no teste de normalidade de Shapiro-Wilk com amostras de tamanho 20 e nível de significância 5%. Obtendo-se os gráficos do poder do teste versus β_2 e poder do teste versus τ_4, Hosking verificou que o gráfico do poder do teste versus β_2 apresentou comportamento linear crescente apenas para as distribuições com curtose próxima de 3 sendo que os demais pontos ficaram dispersos sem um comportamento funcional explícito. Já o gráfico do poder do teste versus τ_4 apresentou comportamento linear crescente. A mesma verificação do comportamento obteve-se considerando os coeficientes de assimetria usual, β_1, o coeficiente calculado por L-momentos, $\tau_3 = \frac{\lambda_3}{\lambda_2}$, e o poder do teste de normalidade de Shapiro-Wilk, porém, o comportamento linear apresentou-se mais evidente para o caso em que foi considerado o coeficiente de curtose. Com isso conclui-se que, se o objetivo for utilizar o coeficiente de curtose, é melhor usar o coeficiente calculado por L-momentos. Como os momentos são, frequentemente, usados como medidas sumárias da forma de uma distribuição, a verificação que Hosking obteve, sugere que os L-momentos provê boas medidas sumárias da forma da distribuição.

Hosking [**?**], também, gerou 50 amostras aleatórias de tamanho 100 a partir de três distribuições diferentes: Distribuição GEV com coeficiente de assimetria $\beta_1 = 3$, Distribuição Weibull com coeficiente de assimetria $\beta_1 = 1$ e Distribuição Weibull com coeficiente de assimetria, calculado por L-momentos, $\tau_3 = 1$, para analisar a discriminação das distribuições através dos coeficientes de assimetria amostrais de β_1 e τ_3 e dos coeficientes de curtose amostrais de β_2 e τ_4. Com as distribuições geradas, foram obtidos os gráficos do coeficiente de curtose amostral b_2 versus o coeficiente de assimetria amostral b_1 e do coeficiente de curtose amostral t_4 versus o coeficiente de assimetria t_1. No gráfico de b_1 versus b_2, os pontos gerados das distribuições GEV e Weibull citadas formaram uma nuvem de pontos sobrepostos impedindo, assim, a discriminação das distribuições. Já no gráfico de t_3 versus t_4, os pontos gerados das distribuições formaram nuvens separadas de pontos permitindo, assim, a discriminação das distribuições.

Com isso, os coeficientes de assimetria e curtose calculados por L-momentos podem ser utilizados

em uma análise exploratória de dados seja para o reconhecimento da forma (assimetria, curtose) da distribuição ou, ainda, para reconhecer misturas de distribuições.

2.2.1 Estimação por L-momentos dos parâmetros da Distribuição GEV

Nesta seção serão obtidos os estimadores dos parâmetros μ, σ e α da Distribuição GEV utilizando-se o Método dos L-momentos. Também, serão feitas simulações a fim de comparar-se os estimadores obtidos com uma aproximação proposta por Hosking.

A distribuição GEV é dada por:

$$G(y) = \begin{cases} exp\left(-(1-\alpha y)^{\frac{1}{\alpha}}\right) & \alpha \neq 0 \\ \\ exp(-exp(-y)) & \alpha = 0 \end{cases}$$

onde $1 - \alpha y > 0$ e $y = \frac{x-\mu}{\sigma}$.

A seguir, serão calculados os 3 primeiros L-momentos da distribuição GEV para $\alpha \neq 0$.

Da expressão de $G(x)$ temos que:

$$x(G) = \begin{cases} \mu + \frac{\sigma}{\alpha}\left(1 - (-ln(G))^{\alpha}\right) & \alpha \neq 0 \\ \\ \mu - \sigma ln(-ln(G)) & \alpha = 0 \end{cases}$$

Para o cálculo do r-ésimo L-momento, será definida a seguinte integral: $\int_0^1 x(G)G^n dG$ onde G é a distribuição GEV.

$$\begin{aligned} I_n &= \int_0^1 x(G)G^n dG = \int_0^1 \left(\mu + \frac{\sigma}{\alpha}\left[1 - (-ln(G))^{\alpha}\right]\right) G^n dG = \\ &= \int_0^1 \left(\mu + \frac{\sigma}{\alpha}\right) G^n dG - \frac{\sigma}{\alpha}\int_0^1 (-ln(G))^{\alpha} G^n dG = \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\int_0^1 (-ln(G))^{\alpha} G^n dG \end{aligned} \tag{2.12}$$

Utilizando a mudança de variável $u = -ln(G)$ em (2.12) temos:

$$I_n = \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\int_0^1 (-ln(G))^\alpha G^n dG \;=\; \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) + \frac{\sigma}{\alpha}\int_\infty^0 u^\alpha e^{-(n+1)u} du =$$

$$= \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\int_0^\infty u^\alpha e^{-(n+1)u} du \tag{2.13}$$

Através da mudança de variável $v = (n+1)u$ e da utilização da função gama, $\Gamma(n) = \int_0^\infty u^{n-1}e^{-u}du$, em (2.13) temos:

$$I_n = \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\int_0^\infty u^\alpha e^{-(n+1)u} du = \frac{1}{n+1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\int_0^\infty \frac{v^\alpha}{(n+1)^{\alpha+1}} e^{-v} du =$$

$$= (n+1)^{-1}\left(\mu + \frac{\sigma}{\alpha}\right) - \frac{\sigma}{\alpha}\left((n+1)^{-\alpha-1}\right)\Gamma(1+\alpha) = (n+1)^{-1}\left[\mu + \frac{\sigma}{\alpha}\left(1 - (n+1)^{-\alpha}\Gamma(1+\alpha)\right)\right]$$

Assim,

$$I_n = (n+1)^{-1}\left[\mu + \frac{\sigma}{\alpha}\left(1 - (n+1)^{-\alpha}\Gamma(1+\alpha)\right)\right] \tag{2.14}$$

Através das expressões (2.14) e (2.6) obtem-se os L-momentos de ordem r=1, 2 e 3, dados a seguir:

$$\lambda_1 = I_0 = \mu + \frac{\sigma}{\alpha}\left(1 - \Gamma(1+\alpha)\right)$$

$$\lambda_2 = 2I_1 - I_0 = \frac{\sigma}{\alpha}\left(1 - 2^{-\alpha}\right)\Gamma(1+\alpha)$$

$$\lambda_3 = 6I_2 - 6I_1 + I_0 = \frac{\sigma}{\alpha}\Gamma(1+\alpha)\left(\frac{3}{2^\alpha} - \frac{2}{3^\alpha} - 1\right) = \frac{\sigma}{\alpha}\Gamma(1+\alpha)\left(\frac{3}{2^\alpha} - \frac{2}{3^\alpha} - 2 + 3\right) =$$

$$= \frac{\sigma}{\alpha}\Gamma(1+\alpha)[-3(1-2^{-\alpha}) + 2(1-3^{-\alpha})]$$

Da expressão de λ_3 e λ_2 obtém-se $\tau_3 = \frac{\lambda_3}{\lambda_2} = 2\left(\frac{1-3^{-\alpha}}{1-2^{-\alpha}}\right) - 3$.

Igualando-se as quantidades λ_1, λ_2 e $\frac{\lambda_3}{\lambda_2}$ aos respectivos L-momentos amostrais obtemos os estimadores de μ, σ e α por L-momentos.

O estimador dos parâmetros μ e σ podem ser obtidos de forma explícita, em função do estimador de α:

$$\hat{\sigma} = \frac{l_2\hat{\alpha}}{(1-2^{-\hat{\alpha}})\Gamma(1+\hat{\alpha})}$$

$$\hat{\mu} = l_1 - \frac{\hat{\sigma}}{\hat{\alpha}}(1-\Gamma(1+\hat{\alpha})) \tag{2.15}$$

Para estimar-se α, não existe solução explícita. O estimador de α é obtido resolvendo-se numericamente a equação:

$$t_3 = \frac{l_3}{l_2} = 2\left(\frac{1-3^{-\alpha}}{1-2^{-\alpha}}\right) - 3 \tag{2.16}$$

Será utilizado o Método da Bissecção [**?**](veja o anexo sobre este método) para a obtenção de $\hat{\alpha}$.

Hosking, Wallis e Wood [**?**] forneceram a seguinte aproximação para a equação (2.16):

$$\alpha^* = 7,859c + 2,9554c^2 \tag{2.17}$$

onde $c = \frac{2}{\left(3+\frac{l_3}{l_2}\right)} - \frac{ln(2)}{ln(3)}$.

Porém a aproximação proposta vale somente se $-1/2 \leq \alpha \leq 1/2$. Pela Tabela 2.2, observa-se o comportamento das estimativas dos parâmetros dentro e fora deste intervalo. Hosking, Wallis e Wood [**?**] utilizaram a aproximação anterior para modelar fenômenos naturais extremos em hidrologia como, por exemplo, na modelagem da distribuição dos máximos anuais de correnteza de rios.

Para comparar as estimativas $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\alpha}$ utilizando-se as equações (2.15),(2.16) e (2.15),(2.17) foram geradas 500 amostras de tamanho 500 provenientes de uma Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$. Para o parâmetro α foi considerado os valores -0,5; -0,2; 0; 0,2; 0,5 e 4. As estatísticas sumárias obtidas encontam-se nas Tabelas 2.2 e 2.3.

Pela Tabela 2.2 observa-se que as estimativas $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\alpha}$ dadas por (2.15),(2.16) e (2.15),(2.17), no intervalo $-1/2 \leq \alpha \leq 1/2$, apresentaram-se muito próximas. Para α=4, as estimativas obtidas utilizando a aproximação proposta por Hosking e Wallis [**?**] apresentaram uma pequena discrepância em relação às estimativas dadas por (2.15) e (2.16).

Também, foram geradas 500 amostras de tamanho 500 provenientes de uma Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e $\alpha = 8$. Utilizando-se a aproximação (2.17), observa-se, pela Tabela 2.4, uma grande discrepância das estimativas e dos parâmetros populacionais $\mu = 0$, $\sigma = 1$ e $\alpha = 8$.

Tabela 2.2: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.16)

Parâmetro	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP
	$\hat{\mu}$	-1,061	-0,153	-0,028	0,0681	0,183	-0,067	0,187
α=4	$\hat{\sigma}$	0,275	0,714	1,061	1,46	4,874	1,19	0,663
	$\hat{\alpha}$	3,051	3,39	3,701	4,134	4,917	3,774	0,467
	$\hat{\mu}$	-0,365	-0,076	-0,008	0,073	0,266	-0,005	0,107
α=0,5	$\hat{\sigma}$	0,748	0,946	0,998	1,047	1,247	0,996	0,078
	$\hat{\alpha}$	0,275	0,435	0,493	0,549	0,765	0,494	0,082
	$\hat{\mu}$	-0,300	-0,079	0,003	0,081	0,371	0,0027	0,112
α=0,2	$\hat{\sigma}$	0,733	0,955	1,004	1,054	1,248	1,006	0,078
	$\hat{\alpha}$	0,013	0,149	0,199	0,248	0,418	0,202	0,071
	$\hat{\mu}$	-0,288	-0,081	0,003	0,082	0,393	0,0034	0,114
α=0	$\hat{\sigma}$	0,702	0,950	1,005	1,053	1,313	1,006	0,085
	$\hat{\alpha}$	-0,199	-0,045	0,0019	0,055	0,245	0,0036	0,076
	$\hat{\mu}$	-0,341	-0,082	-0,0047	0,075	0,315	-0,0024	0,112
α= - 0,2	$\hat{\sigma}$	0,728	0,930	0,994	1,063	1,249	0,996	0,095
	$\hat{\alpha}$	-0,439	-0,255	-0,199	-0,122	0,088	-0,192	0,089
	$\hat{\mu}$	-0,281	-0,075	0,0241	0,101	0,461	0,021	0,123
α= - 0,5	$\hat{\sigma}$	0,631	0,931	1,029	1,110	1,666	1,031	0,140
	$\hat{\alpha}$	-0,882	-0,528	-0,441	-0,376	-0,135	-0,454	0,121

Tabela 2.3: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.17)

Parâmetro	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP
	$\hat{\mu}$	-2,331	-0,834	-0,530	-0,343	0,0020	-0,622	0,383
α=4	$\hat{\sigma}$	0,853	2,455	3,434	5,024	35,16	4,02	2,515
	$\hat{\alpha}$	1,923	2,541	2,697	2,863	3,244	2,699	0,225
	$\hat{\mu}$	-0,377	-0,078	-0,0051	0,071	0,400	-0,0057	0,115
α=0,5	$\hat{\sigma}$	0,759	0,939	0,988	1,058	1,275	0,995	0,087
	$\hat{\alpha}$	0,254	0,443	0,499	0,549	0,765	0,497	0,082
	$\hat{\mu}$	-0,355	-0,079	-0,0092	0,071	0,283	-0,0052	0,108
α=0,2	$\hat{\sigma}$	0,780	0,943	0,999	1,045	1,210	0,996	0,076
	$\hat{\alpha}$	-0,021	0,147	0,194	0,243	0,409	0,197	0,072
	$\hat{\mu}$	-0,349	-0,084	-0,0087	0,074	0,298	-0,0063	0,112
α=0	$\hat{\sigma}$	0,758	0,938	0,998	1,051	1,239	0,997	0,082
	$\hat{\alpha}$	-0,226	-0,054	-0,0079	0,054	0,209	-0,0019	0,074
	$\hat{\mu}$	-0,274	-0,092	-0,0067	0,076	0,426	-0,0030	0,117
α= - 0,2	$\hat{\sigma}$	0,679	0,920	0,996	1,058	1,359	0,998	0,1065
	$\hat{\alpha}$	-0,565	-0,255	-0,186	-0,130	0,088	-0,196	0,097
	$\hat{\mu}$	-0,282	-0,075	-0,024	0,101	0,460	0,021	0,123
α= - 0,5	$\hat{\sigma}$	0,632	0,936	1,033	1,109	1,664	1,035	0,138
	$\hat{\alpha}$	-0,868	-0,527	-0,441	-0,376	-0,136	-0,453	0,119

Tabela 2.4: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e α utilizando-se as equações (2.15) e (2.17)

Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP
$\hat{\mu}$	-46000	-72,77	-33,82	-15,55	-1,086	-177,4	2070,661
$\hat{\sigma}$	9,613	140,9	388,7	1080	1532000	4687	68881,34
$\hat{\alpha}$	2,628	3,058	3,156	3,231	3,303	3,13	0,130

2.3 Método de Máxima Verossimilhança

Dada uma distribuição de probabilidade cujos parâmetros são desconhecidos, tem-se como objetivo, a partir de uma amostra de uma determinada população, estimar o valor do parâmetro da distribuição de probabilidade que representa toda a população. Assim, considerando $\mathbf{X} = (X_1, ..., X_n)$ uma amostra aleatória i.i.d. com distribuição de probabilidade $p(\mathbf{x} \mid \theta)$ onde θ é o vetor de parâmetros a serem estimados, deseja-se dar uma estimativa do valor de θ ou de uma função de $\tau(\theta)$ a partir de valores observados $x_1, ..., x_n$. O Método de Máxima Verossimilhança, por ser um método de estimação pontual, nos fornecerá um valor numérico de uma estatística $T(\mathbf{X})$ para estimar $\tau(\theta)$.

Dada uma amostra $\mathbf{x} = (x_1, ..., x_n)$, o Método de Máxima Verossimilhança consiste em maximizar a função de verossimilhança definida por $L(\theta|\mathbf{x}) = p(\mathbf{x}|\theta) = p(x_1|\theta).p(x_2, \theta).....p(x_n, \theta)$, $\theta \in \Theta$ onde Θ representa o espaço paramétrico de θ, isto é, o método consiste em encontrar $\hat{\theta}$ tal que $\hat{\theta} = argmax_{\theta \in \Theta} L(\theta|\mathbf{x})$ onde $\hat{\theta}$ é o vetor das estimativas de θ e $\hat{\theta} = \hat{\theta}(\mathbf{X})$ é o vetor dos estimadores de máxima verossimilhança de θ.

Como o valor de θ que maximiza $L(\theta|\mathbf{x})$ é o mesmo que maximiza $l(\theta|\mathbf{x}) = Ln(L(\theta|\mathbf{x}))$ e $l(\theta|\mathbf{x})$ sempre está definida pois $L(\theta|\mathbf{x})$ é positiva (produto de funções de probabilidade) então,

por simplicidade, maximiza-se $l(\theta|\mathbf{x})$.

O Método de Máxima Verossimilhança pode apresentar algumas dificuldades práticas:

- O cálculo de derivadas é muito complicado ;
- A resolução $\frac{\partial}{\partial\theta}L(\theta|\mathbf{x}) = 0$ não possui solução explícita;
- $L(\theta|\mathbf{x})$ não possui ponto de máximo ou $L(\theta|\mathbf{x})$ possui máximos locais.

Devido às dificuldades práticas, deve-se recorrer à métodos numéricos de otimização para o cálculo de um estimador de máxima verossimilhança.

A seguir, serão apresentadas algumas propriedades dos estimadores de máxima verossimilhança Mas, primeiramente, será definida a Matriz de Informação de Fisher para o caso uniparamétrico e multiparamétrico e os conceitos de suficiência, consistência e eficiência.

Suponha que o modelo paramétrico seja regular [1] e que Θ seja um subconjunto aberto em $\mathbb{R}$. Por simplicidade, será assumido que $p(\mathbf{x}|\theta)$ seja uma função densidade. A discussão e os resultados são essencialmente idênticos para o caso discreto. As seguintes condições serão consideradas.

1. O conjunto $A = \{\mathbf{x} : p(\mathbf{x}|\theta) > 0\}$ não depende de θ onde $p(\mathbf{x}, \theta)$ é uma densidade. E, para todo $\mathbf{x} \in A$, $\theta \in \Theta$, $\frac{\partial}{\partial\theta}Ln(p(\mathbf{x}|\theta))$ existe e é finita;
2. Se T é alguma estatística tal que $\mathbb{E}_\theta(|T|) < \infty$, para todo $\theta \in \Theta$, então

$$\frac{\partial}{\partial\theta}\int T(\mathbf{x})p(\mathbf{x}|\theta)d\mathbf{x} = \int T(\mathbf{x})\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)d\mathbf{x} \tag{3.18}$$

sempre que o lado direito de (3.18) for finito.

Definição 2.2. *Se a condição 1. for válida, o Número de Informação de Fisher, $I(\theta)$, é definido por*

$$I(\theta) = \mathbb{E}_\theta\left(\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right]^2\right) = \int\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right]^2 p(\mathbf{x}|\theta)d\mathbf{x}$$

[1] Um modelo paramétrico P_θ é denominado regular se: Todos os modelos P_θ forem contínuos com função densidade $p(\mathbf{x}|\theta)$ ou todos os modelos P_θ forem discretos com função de probabilidade $p(\mathbf{x}|\theta)$ existindo um conjunto $\{x_1, x_2, ...\}$, independente de θ, tal que $\sum_{i=1}^{\infty} p(\mathbf{x_i}|\theta) = 1$, para todo θ

Note que $0 \leq I(\theta) \leq \infty$.

Lema 2.1. *Suponha que as condições 1. e 2. sejam válidas e que*

$$\mathbb{E}\left|\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right| < \infty$$

Então,

$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right) = 0$$

E, portanto,

$$I(\theta) = Var\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right)$$

Demonstração:

$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right) = \int \frac{\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta)}p(\mathbf{x}|\theta)d\mathbf{x} = \int \frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)d\mathbf{x} = \frac{\partial}{\partial\theta}\int p(\mathbf{x}|\theta)d\mathbf{x} = 0$$

□

Teorema 2.4. *Seja $T(\mathbf{X})$ uma estatística tal que $Var_\theta(T(\mathbf{X})) < \infty$ para todo θ. Denotando $\mathbb{E}_\theta(T(\mathbf{X}))$ por $\psi(\theta)$, supondo a validade das condições 1. e 2. e $0 < I(\theta) < \infty$ temos que, para todo θ, $\psi(\theta)$ é diferenciável e*

$$Var_\theta(T(\mathbf{X})) \geq \frac{[\dot{\psi}(\theta)]^2}{I(\theta)} \tag{3.19}$$

onde $\dot{\psi}(\theta) = \frac{\partial}{\partial\theta}\psi(\theta)$.

Demonstração: Utilizando as condições 1. e 2. obtém-se

$$\dot{\psi}(\theta) = \frac{\partial}{\partial\theta}\int T(\mathbf{X})p(\mathbf{x}|\theta)d\mathbf{x} = \int T(\mathbf{X})\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)d\mathbf{x} = \int T(\mathbf{X})\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{x}|\theta))\right]p(\mathbf{x}|\theta)d\mathbf{x} \tag{3.20}$$

Pelo Lema 2.1, $\mathbb{E}_\theta(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))) = 0$. Então,

$$\begin{aligned}
Cov\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta)), T(\mathbf{X})\right) &= \mathbb{E}\left(\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right]T(\mathbf{X})\right) - \mathbb{E}\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right)\mathbb{E}(\mathbf{X}) = \\
&= \mathbb{E}\left(\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right]T(\mathbf{X})\right) = \\
&= \int T(\mathbf{X})\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{x}|\theta))\right]p(\mathbf{x}|\theta)d\mathbf{x} = \dot{\psi}(\theta)
\end{aligned}$$

Utilizando a Desigualdade de Cauchy-Schwarz [**?**] para as variáveis aleatórias $\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))$ e $T(\mathbf{X})$, isto é, $\left|Cov\left(\left[\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right], T(\mathbf{X})\right)\right| \leq \sqrt{Var(T(\mathbf{X}))Var\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right)}$, obtém-se

$$|\dot{\psi}(\theta)| \leq \sqrt{Var(T(\mathbf{X}))Var\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right)}$$

Pelo Lema 2.1, $Var\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{X}|\theta))\right) = I(\theta)$. Portanto,

$$Var_\theta(T(\mathbf{X})) \geq \frac{[\dot{\psi}(\theta)]^2}{I(\theta)}$$

□

Proposição 2.1. *Suponha que* $\mathbf{X} = (X_1, ..., X_n)$ *é uma amostra aleatória de uma população que possui densidade* $p(\mathbf{x}|, \theta)$, $\theta \in \Theta$, *e que as condições do Teorema 2.4 sejam válidas. Seja* $I_1(\theta) = \mathbb{E}\left(\left[\frac{\partial}{\partial\theta}Ln(p(X_1|\theta))\right]^2\right)$. *Então,*

$$I(\theta) = nI_1(\theta)$$

e

$$Var_\theta(T(\mathbf{X})) \geq \frac{[\dot{\psi}(\theta)]^2}{nI_1(\theta)}$$

Demonstração: Esta proposição é consequência do Lema 2.1 e da expressão

$$I(\theta) = Var\left(\frac{\partial}{\partial\theta}Ln(p(\mathbf{x}|\theta))\right) = Var\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}Ln(p(X_i|\theta))\right) = \sum_{i=1}^{n}Var\left(\frac{\partial}{\partial\theta}Ln(p(X_i|\theta))\right) = nI_1(\theta)$$

□

Para o caso multiparamétrico, seja $\theta = (\theta_1, ..., \theta_d)$. Suponha que Θ seja um subconjunto aberto em $\mathbb{R}^d$ e que $\{p(\mathbf{x}|\theta) : \theta \in \Theta\}$ seja um modelo paramétrico regular com as condições 1. e 2. satisfeitas quando a diferenciação é aplicada à θ_j, $j = 1, ..., d$.

Definição 2.3. *A Matriz de Informação de Fisher é definida como sendo*

$$I_{p\times p}(\theta) = (I_{jk}(\theta))_{1\leq j\leq d, 1\leq k\leq d}$$

onde

$$I_{jk}(\theta) = \mathbb{E}\left(\frac{\partial}{\partial\theta_j}Ln(p(\mathbf{X}|\theta))\frac{\partial}{\partial\theta_k}Ln(p(\mathbf{X}|\theta))\right)$$

Proposição 2.2. *Sob as suposições supra-citadas*

$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta_j}Ln(p(\mathbf{X}|\theta))\right) = 0$$

e

$$I_{jk}(\theta) = Cov_\theta\left(\frac{\partial}{\partial\theta_j}Ln(p(\mathbf{X}|\theta)), \frac{\partial}{\partial\theta_k}Ln(p(\mathbf{X}|\theta))\right)$$

Demonstração: Para a demonstração consulte [**?**].

Teorema 2.5. *Supondo que as condições citadas anteriormente sejam válidas e que a matriz $I(\theta)$ seja não-singular. Então, para todo θ, $\dot{\psi}(\theta)$ existe e $Var_\theta(T(\mathbf{X})) \geq \dot{\psi}(\theta)I(\theta)^{-1}[\dot{\psi}(\theta)]^T$.*

Demonstração: Para a demonstração consulte [**?**].

Por simplicidade, no decorrer da seção, será assumido o caso uniparamétrico. Os conceitos e resultados são análogos para o caso multiparamétrico.

Definição 2.4. *Seja $T(\mathbf{X})$ um estimador para o parâmetro $\tau(\theta)$. O estimador $T(\mathbf{X})$ é não-viciado se $\mathbb{E}_\theta(T(\mathbf{X})) = \tau(\theta)$.*

Para o caso em que $T(\mathbf{X})$ é um estimador não-viciado, a desigualdade (3.19) torna-se,

$$Var_\theta(T(\mathbf{X})) \geq \frac{[\tau^{'}(\theta)]^2}{I(\theta)}$$

pois $\psi(\theta) = \mathbb{E}_\theta(T(\mathbf{X})) = \tau(\theta)$.

Um estimador, em geral, depende do tamanho da amostra. Então, indexando o estimador pelo tamanho amostral, tem-se uma sequência de estimadores. É intuitivo esperar que quanto maior a amostra, melhor seja o estimador. Para tal situação, considera-se a propriedade de consistência dos estimadores a qual descreve o comportamento de um dado estimador quando o tamanho amostral torna-se muito grande.

Definição 2.5. *Sejam $T_1, ..., T_n, ...$ uma sequência de estimadores de $\tau(\theta)$ onde $T_n = T_n(\mathbf{X})$. A sequência $\{T_n\}$ é uma sequência consistente de estimadores do parâmetro θ se, para cada ϵ>0 e cada $\theta \in \Theta$, $lim_{n\to\infty} P_\theta(|T_n - \theta| \geq \epsilon)$=0.*

Um estimador pode alcançar a igualdade na desigualdade (3.19) ou na desigualdade mostrada no Teorema (2.5). Neste caso, o estimador possui variância mínima na classe de todos os estimadores de $\tau(\theta)$. Para o caso assintótico define-se:

Definição 2.6. *Uma sequência de estimadores $T_n(\mathbf{X})$ é assintoticamente eficiente para o parâmetro $\tau(\theta)$ se*

$$lim_{n\to\infty} \frac{Var_\theta(T(\mathbf{X}))}{\frac{[\tau^{'}(\theta)]^2}{nI(\theta)}} = 1$$

Quando um experimento é realizado, depara-se com amostras. Em geral, o conjunto de dados não estão suficientemente organizados para que se possa obter alguma informação. Nesse ponto, uma análise exploratória de dados é realizada e os dados são reduzidos de forma que se possa entender o fenômeno de interesse. Com isso, há uma perda de informação. Então, pergunta-se: Os dados podem ser reduzidos em uma coleção de estatísticas sem perda de informação sobre o parâmetro?

No problema de estimação tem-se um conjunto de dados observados. O objetivo é condensar a informação contida na amostra em alguns números sem perder informação sobre o parâmetro de interesse. Daí, considera-se a propriedade de suficiência das estatísticas.

Sejam χ e κ sendo, respectivamente, o espaço amostral e o domínio de $T(\mathbf{X})$.

Definição 2.7. *Seja* $\mathbf{X} = (X_1, ..., X_n)$ *uma amostra que possui distribuição de probabilidade* $p(\mathbf{X}|\theta)$. *Uma estatística* $T(\mathbf{X})$ *é suficiente se e somente se a distribuição condicional de* $X_1, ..., X_n$ *dado* $T = t$ *não depender de* θ *para qualquer valor* $t \in \kappa$.

Teorema 2.6. *Em um modelo regular, a estatística* $T(\mathbf{X})$ *é suficiente para* θ, *se e somente se, existir uma função* $g(t|\theta)$ *definida para* $t \in \kappa$ *e* $\theta \in \Theta$ *e uma função* h *definida em* χ *tal que* $p(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$ *para todo* $\mathbf{x} \in \chi$, $\theta \in \Theta$.

Demostração: Para a demonstração consulte [**?**].

Dada as definições e resultados anteriores, serão enunciadas algumas propriedades pertinentes aos estimadores de máxima verossimilhança.

Os estimadores de máxima verossimilhança são invariantes pela aplicação de uma função no estimador de máxima verossimilhança, ou seja, esta propriedade nos indica um meio de encontrar um estimador de máxima verossimilhança de uma função $\tau(\theta)$ através do estimador de máxima verossimilhança de θ como mostra o teorema seguinte [**?**].

Teorema 2.7. *Seja* $\hat{\theta} = \hat{\theta}(\mathbf{X})$ *o estimador de máxima verossimilhança de* θ. *Se* $\tau(\theta)$ *for uma transformação no espaço paramético* Θ, *então o estimador de máxima verossimilhança de* $\tau(\theta)$ *será* $\tau(\hat{\theta})$.

Demonstração: Seja $\hat{\theta}$ o estimador de máxima verossimilhança de θ e $L^*(\tau|\mathbf{x}) = sup_{(\theta|\tau(\theta)=\tau)}L(\theta|\mathbf{x})$ a função de verossimilhança de $\tau(\theta)$. É suficiente mostrar que $L^*(\tau(\hat{\theta})|x_1, ..., x_n) \geq L^*(\tau(\theta)|x_1, ..., x_n)$ para algum $\tau \in T$, a qual segue imediatamente da inequação

$$\begin{aligned} L^*(\tau(\theta)|x_1, ..., x_n) &= sup_{(\theta|\tau(\theta)=\tau)}L(\theta|x_1, ..., x_n) \leq sup_{(\theta\in\Theta)}L(\theta|x_1, ..., x_n) = \\ &= L(\hat{\theta}|x_1, ..., x_n) = sup_{(\theta|\tau(\theta)=\tau)}L(\theta|x_1, ..., x_n) = L^*(\tau(\hat{\theta}); x_1, ..., x_n) \end{aligned}$$

□

Temos, também, que se existir uma estatística suficiente então os estimadores de máxima verossimilhança serão função da estatística suficiente como mostra o seguinte resultado [**?**].

Teorema 2.8. *Se existir uma estatística suficiente $T(\mathbf{X})$ associada à uma distribuição de probabilidade $p(\mathbf{x}|\theta)$ então o estimador de máxima verossimilhança $\hat{\theta}$ de θ será função da estatística suficiente, $\hat{\theta} = \hat{\theta}(T(\mathbf{X}))$.*

Demonstração: Se a estatística suficiente $T(\mathbf{X})$ existir então ela satisfará o Teorema 2.6, ou seja, $p(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$.

Como

$$\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta) = 0 \Rightarrow \frac{\partial}{\partial\theta}g(\mathbf{x}|\theta) = 0$$

pois $h(T(\mathbf{x}))$ independe de θ.

Com isso, temos que o estimador de máxima verossimilhança, o qual é obtido pela maximização de $L(\theta|\mathbf{x})$, é função da estatística suficiente $T(\mathbf{X})$. □

Os estimadores de máxima verossimilhança possuem propriedades assintóticas. Cramér [**?**], mostrou que os estimadores de máxima verossimilhança são consistentes, eficientes e possuem normalidade assintótica. Considerando, $L(\mathbf{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$, ele supôs algumas condições:

1. As derivadas $\frac{\partial}{\partial\theta}Ln(p(x|\theta))$, $\frac{\partial^2}{\partial\theta^2}Ln(p(x|\theta))$ e $\frac{\partial^3}{\partial\theta^3}Ln(p(x|\theta))$ existem para todo θ pertencente a um intervalo não-degenerado A e para quase todos os x;

2. Para todo θ pertencente à A, tem-se

$$\left|\frac{\partial}{\partial\theta}p(x|\theta)\right| < F_1(x)$$
$$\left|\frac{\partial^2}{\partial\theta^2}p(x|\theta)\right| < F_2(x)$$
$$\left|\frac{\partial^3}{\partial\theta^3}p(x|\theta)\right| < H(x)$$

onde as funções F_1 e F_2 são integráveis em $(-\infty, \infty)$ enquanto que $\int_{-\infty}^{\infty} H(x)p(x|\theta)dx < M$, M independente de θ;

3. Para todo θ pertencente à A, a integral $\int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\theta}ln(p(x|\theta))\right]^2 p(x|\theta)dx$ é finita e positiva.

Teorema 2.9. *Sob as condições supra-citadas, tem-se:*

1. *Seja* $\mathbf{X} = (X_1, ..., X_n)$ *uma amostra i.i.d. possuindo função de distribuição* $p(\mathbf{x}|\theta)$ *sendo* $L(\theta|\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\theta)$ *a sua função de verossimilhança. Sejam* $\hat{\theta}$ *o estimador de máxima verossimilhnaça de* θ *e* $\tau(\theta)$ *uma função contínua. Então, para cada* $\epsilon>0$ *e cada* $\theta \in \Theta$,

$$lim_{n\to\infty} P_\theta(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0$$

2. *Os estimadores de máxima verossimilhança são eficientes;*

3. *Seja* $\tau(\hat{\theta})$ *o estimador de máxima verossimilhança de* $\tau(\theta)$. *Se* $\tau(\theta)$ *for diferenciável então temos que* $\tau(\hat{\theta})$ *tem assintoticamente uma distribuição Normal com média* $\tau(\theta)$ *e variância* $\frac{[\tau'(\theta)]^2}{nI(\theta)}$.

Demonstração: Para a demonstração consulte [**?**].

Bickel e Doksum [**?**] consideraram condições mais gerais e demonstraram a eficiência e a normalidade assintótica para os M-estimadores [**?**, **?**] sendo que a eficiência e a normalidade assintótica dos estimadores de máxima verossimilhança segue como um caso particular no contexto dos M-estimadores.

Mas há casos em que as propriedades assintóticas dos estimadores de Máxima Verossimilhança não são garantidas. Smith [?] considerou o estudo de funções densidade da forma $f(x|\theta,\phi) = (x-\theta)^{\kappa-1}g(x-\theta|\phi)$ com $\theta < x < \infty$ onde θ e ϕ são vetores formados por componentes desconhecidas, g é uma função que tende para a constante κc quando $x \downarrow \theta$. Assume-se que $\kappa \equiv \kappa(\phi)$ é uma função de ϕ duas vezes diferenciável com $0 < \kappa < \infty$. Esta formulação engloba os casos em que κ é uma componente de ϕ e κ é uma constante conhecida. Neste artigo, o autor apresenta três resultados que permitem o estudo da validade das propriedades dos estimadores de máxima verossimilhança da função $f(x|\theta,\phi)$. Mas, para tal propósito, serão consideradas certas suposições.

Assumindo-se que θ é um número real e os valores de ϕ pertencem ao conjunto $\Phi \subseteq \mathbb{R}^p$, para algum $p \geq 1$, $\phi_1, \phi_2, ...$ e $\phi^1, \phi^2, ...$ representam, respectivamente, elementos de Φ e componentes de um dado ϕ sendo ϕ_j^i, $i = 1, ..., p$, a i-ésima componente de ϕ_j e o símbolo $|\phi|$ denota a norma L^2 de ϕ, assume-se as seguintes condições:

1. As derivadas parciais de segunda ordem de $g(x|\phi)$ existem e são contínuas em $0 < x < \infty$, $\phi \in \Phi$. Além disso, $c(\phi) = \kappa^{-1}\lim g(x|\phi)$, quando $x \downarrow 0$, existe, é positiva e finita para cada ϕ e é duas vezes continuamente diferenciável como uma função de ϕ;

2. Se ϕ_1 e ϕ_2 forem elementos distintos de Φ, então o conjunto $\{x : f(x|0,\phi_1) \neq f(x|0,\phi_2)\}$ possuirá medida de Lebesgue positiva;

3. Para algum $\delta^* \geq 0$ fixo e para cada $\eta > 0$, $\phi_1 \in \Phi$, tem-se

$$Ln(g(x-\epsilon|\phi)) - Ln(g(x-\epsilon_1|\phi_1)) < H_\eta(x|\phi_1)$$

sempre que $|\epsilon| < \eta$, $|\epsilon_1| < \eta$, $|\phi - \phi_1| < \eta$, $x - \epsilon > \delta^*$, $x - \epsilon_1 > \delta^*$ e a função H_η satisfazendo

$$\lim_{\eta \to 0} \int_0^\infty H_\eta(x+y|\phi_1) f(x|0,\phi_o) dx = 0$$

para cada $y \geq 0$, $\phi_o \in \Phi$;

4. Existindo uma sequência fixa crescente de subconjuntos compactos $\{K_m, m \geq 1\}$ de $\mathbb{R}^{p+1}$ e uma constante fixa δ' tal que $\bigcup_m K_m = \mathbb{R} \times \phi$ e, para θ, θ_o, ϕ, ϕ_o satisfazendo $\theta < \theta_o - \delta'$ com $\phi, \phi_o \in \Phi$, (θ, ϕ) não pertencendo à K_m, tem-se

$$Ln(f(x|\theta,\phi)) - Ln(f(x|\theta_o,\phi_o)) < H_m^o(x|\theta_o,\phi_o)$$

com $x > \theta_o$ onde

$$-\infty \leq \limsup_{m\to\infty} \int_{\theta_o}^{\infty} H_m^o(x|\theta_o, \phi_o) f(x|\theta_o, \phi_o) dx < 0;$$

5. Existindo uma sequência fixa crescente de subconjuntos compactos $\{K_m^{'}, m \geq 1\}$ de Φ com $\bigcup_m K_m^{'} = \Phi$, e uma constante fixa $\delta^{'} > 0$, tal que para $\phi \in \Phi - K_m^{'}$, $\theta_o \in \Phi$, $|\epsilon| < \delta^{'}$, $max(0, \epsilon) < x < \infty$ e algum $\eta^{'} > 0$, tem-se

$$\eta^{'}\kappa(\phi) + \{\kappa(\phi) - \kappa(\phi_o)\}Ln(x) + Ln(g(x - \epsilon|\phi)) - Ln(g(x|\phi_o)) < H_m^{'}(x|\phi_o)$$

onde

$$-\infty \leq \limsup_{m\to\infty} \int_{0}^{\infty} H_m^{'}(x|\phi_o) f(x|0, \phi_o) dx < 0;$$

6. Se $\mathbb{E}_\phi$ denotar a esperança com respeito à $f(.|0, \phi)$, então, para cada $\phi \in \Phi$, $i, j = 1, ..., p$,

   (a)

$$\mathbb{E}_\phi \left(\frac{\partial}{\partial \phi^i} Ln(f(X|0, \phi)) \right) = 0$$

   e

$$\mathbb{E}_\phi \left(\left(\frac{\partial}{\partial \phi^i} Ln(f(X|0, \phi)) \right) \left(\frac{\partial}{\partial \phi^j} Ln(f(X|0, \phi)) \right) \right) = -\mathbb{E}_\phi \left(\frac{\partial^2}{\partial \phi^i \partial \phi^j} Ln(f(X|0, \phi)) \right) = m_{ij}(\phi);$$

   (b) Se $\kappa > 1$ então

$$\mathbb{E}_\phi \left(\frac{\partial}{\partial x} Ln(f(X|0, \phi)) \right) = 0$$

   e

$$-\mathbb{E}_\phi \left(\left(\frac{\partial}{\partial x} Ln(f(X|0, \phi)) \right) \left(\frac{\partial}{\partial \phi^i} Ln(f(X|0, \phi)) \right) \right) = \mathbb{E}_\phi \left(\frac{\partial^2}{\partial x \partial \phi^i} Ln(f(X|0, \phi)) \right) =$$

$$= m_{i0}(\phi) = m_{0i}(\phi);$$

(c) Se $\kappa > 2$ então

$$\mathbb{E}_\phi\left(\left[\frac{\partial}{\partial x}Ln(f(X|0,\phi))\right]^2\right) = -\mathbb{E}_\phi\left(\frac{\partial^2}{\partial x^2}Ln(f(X|0,\phi))\right) = m_{00}(\phi) > 0.$$

7. Considerando $h(x|\phi)$ igual à expressão

$$\frac{\partial^2}{\partial x \partial \phi^i}Ln(g(x|\phi))$$

ou à expressão

$$\frac{\partial^2}{\partial \phi^i \partial \phi^j}Ln(g(x|\phi))$$

então, quando $\theta \to \theta_o$ e $\phi \to \phi_o$,

$$\mathbb{E}_o(|h(X-\theta|\phi) - h(X-\theta_o|\phi_o)|) \to 0$$

onde $\mathbb{E}_o$ é a esperança em relação à $f(.|\theta_o, \phi_o)$. Se $\kappa(\phi_o) > 2$, será requerido o mesmo para $h(x|\phi) = \frac{\partial^2}{\partial x^2}Ln(g(x|\phi))$;

8. Para cada $\epsilon > 0$, $\delta > 0$, existe uma função $h_{\epsilon,\delta}$ tal que

$$\left|\frac{\partial^2}{\partial x^2}Ln(g(x|\phi))\right| < \frac{\epsilon}{x^2} + h_{\epsilon,\delta}(y|\phi_o)$$

sempre que $|\phi - \phi_o| < \delta$, $|x - y| < \delta$ e $h_{\epsilon,\delta}$ satisfazendo $\int_0^\infty h_{\epsilon,\delta}(x|\phi_o)f(x|0,\phi_o)dx < \infty$;

9. Para cada $\delta > 0$, $\phi \in \Phi$,

$$\int_\delta^\infty \left[\frac{\partial}{\partial x}Ln(g(x,\phi))\right]^2 f(x|0,\phi)dx < \infty.$$

As suposições 1. e 6. são as mais importantes para as demonstrações dos teoremas que serão enunciados posteriormente. As demais são suposições impostas por razões técnicas. As suposições 2. à 5. são similares, porém mais complexas, às suposições clássicas de Wald [**?**]. A suposição 9. é uma suposição assumida por Woodroofe [**?**].

Considere θ_o e ϕ_o como sendo os verdadeiros valores, respectivamente, de θ e ϕ. Os símbolos κ e c, a menos que se mencione o contrário, denotarão, respectivamente, $\kappa(\phi_o)$ e $c(\phi_o)$.

Seja $\mathbf{X} = (X_1, ..., X_n)$ uma amostra aleatória i.i.d. com densidade $f(.|\theta_o, \phi_o)$ e $(X_{1,n}, ..., X_{n,n})$ a amostra de $\mathbf{X}$ ordenada. A função $L_n(\theta, \phi) = n^{-1} \sum_{i=1}^{n} Ln(f(X_i|\theta, \phi))$ representa a função de verossimilhança dividida por n. Note que $L_n(\theta, \phi)$ está definido somente para $\theta < X_{1,n}$.

Para o caso especial em que $\theta = \theta_o$ for conhecido, $\overline{\phi}_n = \overline{\phi}_n(\theta_o)$ denotará o estimador de máxima verossimilhança de ϕ satisfazendo

$$\frac{\partial L_n}{\partial \phi^i}\left(\theta_o, \overline{\phi}_n\right) = 0.$$

Similarmente, $\overline{\theta}_n = \overline{\theta}_n(\phi_o)$ denotará o estimador de máxima verossimilhança de θ quando $\phi = \phi_o$ for conhecido.

Além disso, seja

$$\xi_{n,\beta} = \begin{cases} 1 & \beta > 1 \\ Ln(n) & \beta = 1 \\ n^{\frac{1}{\beta}-1} & 0 < \beta < 1 \end{cases}$$

onde $n \geq 1$.

Define-se $Y_n <_p r_n$, para v.a. $\{Y_n\}$ e constantes positivas $\{r_n\}$, se

$$\lim_{a \to \infty} \limsup_{n \to \infty} P(|Y_n| > a r_n) = 0.$$

O teorema seguinte é um resultado que abordará o comportamento local da função de verossimilhança perto dos verdadeiros valores dos parâmetros. Não haverá garantia de que o máximo local, cuja existência será garantida pelo próximo teorema quando $\kappa > 1$, seja único, sabendo-se que o máximo local não é um máximo global.

Teorema 2.10. *Assumindo-se as suposições 1. e 6. à 8. e que $\kappa > 1$ e supondo-se que M é estritamente definita positiva onde:*

1. *Para $\kappa > 2$, M é uma matriz $(p+1) \times (p+1)$ com entradas $m_{ij}(\phi_o)$, $i, j = 0, ..., p$;*

2. *Para $1 < \kappa \leq 2$, M é uma matriz $p \times p$ com entradas $m_{ij}(\phi_o)$, $i, j = 0, ..., p$, onde a função m_{ij} foi definida na suposição 6. .*

Então, existe uma sequência $(\hat{\theta_n}, \hat{\phi_n})$ de soluções das equações de verossimilhança tal que $\hat{\theta_n} - \theta_o <_p (n\xi_{n,\kappa})^{-1/2}$ e $\hat{\phi_n} - \phi_o <_p n^{-1/2}$. Além disso, se $\kappa = 2$, $\hat{\theta_n} - \overline{\theta_n} <_p n^{-1/2}(Ln(n))^{-1}$ e $\hat{\phi_n} - \overline{\phi_n} <_p (nLn(n))^{-1/2}$ enquanto que, para $1 < \kappa < 2$, tem-se $\hat{\theta_n} - \overline{\theta_n} <_p n^{-\frac{2}{\kappa}+1/2}$ e $\hat{\phi_n} - \overline{\phi_n} <_p n^{-1/\kappa}$.

Demonstração: Para a demonstração consulte [**?**].

O próximo teorema, através da sua demonstração, mostrará que a região na qual $X_{1,n} - \theta$ é exponencialmente pequena e $\kappa(\phi) \leq 1$ será, assintoticamente, a única região onde a função de verossimilhança é mal comportada embora o principal interesse seja os resultados para o caso $\kappa > 1$, quando os ítens 1. e 2., dados no teorema a seguir, forem válidos.

Teorema 2.11. *Suponha que, para cada n, $\mathbf{X}$ seja uma amostra aleatória com n observações i.i.d com função densidade $f(.|\theta_o, \phi_o)$ onde $(X_{1,n}, ..., X_{n,n})$ representa a amostra de $\mathbf{X}$ ordenada. Fixados $\delta > 0$, $\epsilon > 0$, $\gamma < \infty$, define-se as seguintes regiões do espaço paramétrico $\mathbb{R} \times \Phi$: U é o conjunto de (θ, ϕ) para o qual $\theta < \theta_o - \delta$, $\phi \in \Phi$, e V_n é o conjunto de (θ, ϕ) para o qual $|\theta - \theta_o| < \delta$, $|\phi - \phi_o| > \delta$ e, ou $\kappa(\phi) > 1 + \epsilon$ ou $\theta < X_{1,n} - n^{-\gamma}$.*

1. *Supondo que as suposições 1. à 6. sejam válidas com $\delta^* = \delta$ na suposição 3., $\delta^1 = \delta$ nas suposições 4. e 5. . Então,*

$$\lim_{n\to\infty} P(\sup_U L_n(\theta, \phi) \geq L_n(\theta_o, \phi_o)) = 0;$$

2. *Supondo que as suposições 1. à 6. sejam válidas com $\delta^* = 0$ na suposição 3., $\delta^{'} = \delta$ nas suposições 4. e 5. e $\kappa > 1$. Então,*

$$\lim_{n\to\infty} P\left(\sup_{V_n} L_n(\theta, \phi) \geq L_n(\theta_o, \phi_o)\right) = 0.$$

Demonstração: Para a demonstração consulte [**?**].

Teorema 2.12. *Sob as suposições do Teorema 2.10, seja* $(\hat{\theta}_n, \hat{\phi}_n)$ *uma sequência de estimadores de máxima verossimilhança satisfazendo as conclusões do Teorema 2.10.*

1. *Se* $\kappa > 2$ *então* $n^{1/2}(\hat{\theta}_n - \theta_o, \hat{\phi}_n - \phi_o)$ *converge, em distribuição, para a Distribuição Normal* $N\left(\mathbf{0}, \sum^{-1}\right)$ *onde* M *foi dada no ítem 1. do Teorema 2.10 e* $\sum$ *é uma matriz de covariância;*

2. *Se* $\kappa = 2$ *então* $\{(ncLn(n))^{1/2}(\hat{\theta}_n - \theta_o), n^{1/2}(\hat{\phi}_n - \phi_o)\}$ *converge, em distribuição, para a Distribuição Normal com matriz de covariância da forma*

$$\begin{bmatrix} 1 & 0 \\ 0 & M^{-1} \end{bmatrix}$$

   onde M *foi dada no ítem 2. do Teorema 2.10;*

3. *Se* $1 < \kappa < 2$ *então* $\{(nc)^{1/\kappa}(\hat{\theta_n} - \theta_o), n^{1/2}(\hat{\phi}_n - \phi_o)\} \to (Y, Z)$*, em distribuição, onde* $Y \in \mathbb{R}$ *e* $Z \in \mathbb{R}^p$ *são v.a. independentes,* Z *tendo Distribuição Normal* $N(\mathbf{0}, M^{-1})$*, com* M *dada no ítem 2. deste teorema, e* Y *tendo função de densidade* H *onde* H *foi definida no enunciado do Teorema 2.4 de Woodroofe [**?**].*

Demonstração: Para a demonstração consulte [**?**].

O próximo corolário mostrará que a variância dos estimadores poderá ser estimada, assintoticamente, pelas médias da informação observada, no caso em que a informação esperada não exista.

Corolário 2.1. *Se* $\kappa = 2$ *então*

1. $$(\overline{\theta}_n - \theta_o)\left\{-n\frac{\partial^2 L_n(\overline{\theta}_n, \phi_o)}{\partial\theta^2}\right\}^{1/2} \to Z$$

2. $$(\hat{\theta}_n - \theta_o)\left\{-n\frac{\partial^2 L_n(\hat{\theta}_n, \hat{\phi}_n)}{\partial\theta^2}\right\}^{1/2} \to Z$$

   onde a convergência é em distribuição e, para cada caso, Z *possui Distribuição Normal Padrão.*

Demonstração: Para a demonstração consulte [**?**].

2.3.1 Estimação por Máxima Verossimilhança dos parâmetros da Distribuição GEV

Nesta seção serão obtidos os estimadores dos parâmetros μ, σ e α da Distribuição GEV utilizando-se o Método de Máxima Verossimilhança. Também, serão discutidos os casos em que as propriedades assintóticas de tais estimadores permanecem válidas já que o suporte da Distribuição GEV depende dos seus parâmetros.

Para a Distribuição GEV, com θ=(μ, σ, α), temos a seguinte função de verossimilhança:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \left[\frac{1}{\sigma} \left(1 - \alpha \left(\frac{x_i - \mu}{\sigma} \right) \right)^{(1-\alpha)/\alpha} exp \left(- \left(1 - \alpha \left(\frac{x_i - \mu}{\sigma} \right) \right)^{\frac{1}{\alpha}} \right) \right]$$

Tomando o logaritmo natural da expressão acima temos:

$$l(\theta|\mathbf{x}) = Ln(L) = -nLn(\sigma) - (1-\alpha) \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} e^{-z_i} \tag{3.21}$$

onde $z_i = -\frac{1}{\alpha} ln(1 - \alpha \left(\frac{x_i - \mu}{\sigma} \right))$.

Pela expressão (3.21), a resolução de $\frac{\partial}{\partial \theta} L(\theta) = 0$ não possui solução explícita. Assim, será utilizado o método iterativo Nelder-Mead [**?**, **?**] (veja o anexo sobre este método).

A função densidade da Distribuição GEV é dada por

$$g(x|\mu, \sigma, \alpha) = \frac{1}{\sigma} \left(1 - \alpha \left(\frac{x - \mu}{\sigma} \right) \right)^{(1-\alpha)/\alpha} exp \left(- \left(1 - \alpha \left(\frac{x - \mu}{\sigma} \right) \right)^{\frac{1}{\alpha}} \right) \tag{3.22}$$

cujo suporte é o conjunto $\left\{ x : \alpha \left(\frac{x-\mu}{\sigma} \right) < 1 \right\}$ sendo $\kappa = \frac{1}{\alpha}$.

Observa-se que o suporte da Distribuição GEV depende dos seus parâmetros.

Por causa disso, as propriedades assintóticas, como normalidade assintótica, consistência, eficiência, dos estimadores de máxima verossimilhança da Distribuição GEV ficam comprometidas.

Smith [**?**] mostrou que os estimadores de máxima verossimilhança dos parâmetros da Distribuição GEV não mantêm as suas boas propriedades para $\alpha \geq \frac{1}{2}$.

Para $\alpha > 0$, os resultados anteriores são diretamente aplicáveis e, mostram, em particular, que as propriedades assintóticas dos estimadores de máxima verossimilhança permanecem válidas para $0 < \alpha < 1/2$, mas não para $\alpha \geq 1/2$, isto é, a existência de um ponto de máximo e a convergência assintótica são garantidas, respectivamente, pelos Teoremas 2.10 e 2.12, para $\frac{1}{\alpha} = \kappa > 2$ e $\kappa = 2$, respectivamente, $\alpha \in (0, 1/2)$ e $\alpha = \frac{1}{2}$. E, para $1 < \kappa < 2$ ($\alpha \in (\frac{1}{2}, 1)$), o Teorema 2.10 garante a existência de um ponto de máximo mas o Teorema 2.12 não garante uma convergência assintótica usual.

Para $\alpha < 0$, o suporte da distribuição novamente depende de parâmetros desconhecidos, com função densidade positiva quando $x > \mu + \frac{\sigma}{\alpha}$. Utilizando $\beta = -1/\alpha$, $\theta = \mu + \frac{\sigma}{\alpha}$, a função densidade (3.22) reparametrizada é dada por

$$g(x|\beta, \theta, \alpha) = \frac{1}{\sigma}\left(\frac{x-\theta}{\beta\sigma}\right)^{-\beta-1} exp\left(-\left(\frac{x-\theta}{\beta\sigma}\right)^{-\beta}\right) \tag{3.23}$$

para $x > \theta$ sendo que ela converge à zero mais rapidamente do que alguma potência de $x - \theta$ quando $x \downarrow \theta$. Ainda que isto não caia dentro do escopo do Teorema 2.10, os argumentos que foram aplicados lá, para o caso $\kappa > 2$, são aplicáveis aqui também, e mostram a existência do estimador de máxima verossimilhança assintoticamente normal e eficiente.

Então, conclui-se que as propriedades asssintóticas dos estimadores de máxima verossimilhança permanecem válidas para $-\infty < \alpha < 1/2$ sendo que a matriz de informação é finita no intervalo $-\infty < \alpha < \frac{1}{2}$.

2.4 Comparação entre os Métodos dos L-momentos e Máxima Verossimilhança

Nesta seção, os Métodos de Máxima Verossimilhança (MV) e L-momentos (LM) serão comparados considerando-se perturbações dos parâmetros populacionais como aproximações iniciais dos métodos numéricos utilizados.

Para a compararação dos Métodos de Máxima Verossimilhança e L-momentos, gerou-se 30 e 100 amostras independentes de tamanho 15, 50, 100, 500 e 1000 de uma Distribuição GEV. Para cada

amostra independente, obteve-se dois vetores de estimativas dos parâmetros, μ, σ, α, da Distribuição GEV sendo que cada vetor corresponde a um método de estimação. Através dos vetores de estimativas, estatísticas sumárias e o erro quadrático médio correspondente à cada método e parâmetro estimado foram calculados.

Em relação ao Método dos L-momentos, para a obtenção de $\hat{\alpha}$, utilizou-se o Método da Bissecção (veja o anexo sobre este método) para obter-se a raiz da equação (2.16). E, em relação ao Método de Máxima Verossimilhança, para a obtenção das estimativas $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\alpha}$, utilizou-se o Método de Nelder-Mead (veja o anexo sobre este método) a fim de que a equação (3.21) fosse maximizada. Estes métodos numéricos foram escolhidos por não dependerem da primeira derivada de (2.16) e do gradiente e da hessiana de (3.21), respectivamente, em relação aos Métodos da Bissecção e Nelder-Mead.

Para o Método da Bissecção foi utilizado o intervalo $[\alpha - \varepsilon, \alpha + \varepsilon]$ como aproximação inicial para o método. Para o Método de Nelder-Mead utilizou-se, como estimativas iniciais, $[\mu + \varepsilon, \sigma + \varepsilon, \alpha + \varepsilon]$ onde μ, σ e α são valores populacionais e ε é uma perturbação dos valores populacionais citados. O valor escolhido de ε foi o mesmo para cada método sendo que tal escolha dependeu do êxito da convergência dos métodos numéricos utilizados pois, para uma dada aproximação inicial, apenas um dos métodos convergiu ou ambos não convergiram. Este parâmetro foi incluído nas simulações para que as aproximações iniciais ficassem, relativamente, próximas dos valores populacionais dos parâmetros da Distribuição GEV para que os métodos de estimação pudessem ser comparados pois as aproximações iniciais influem nas estimativas fornecidas pelos métodos numéricos já que elas determinam a região inicial da busca da raiz de uma equação ou do máximo de uma função e, assim, se um método numérico fosse inicializado numa região muito distante em relação ao outro método, as estimativas obtidas poderiam tornar-se inadequadas. Como na prática, não se conhece os verdadeiros parâmetros de uma distribuição, deve-se obter resultados numéricos considerando-se vários valores das condições iniciais.

Foram geradas 30 e 100 amostras independentes provenientes da Distribuição GEV de parâmetros $\mu = 0, \sigma = 1, \alpha = -0,6$ com $\varepsilon = 0,55$. As estatísticas sumárias e os erros quadráticos médios amostrais (EQM) obtidos encontam-se nas Tabelas 4.5, 4.6, 4.7, 4.8 e 4.9.

Também, foram geradas 30 e 100 amostras independentes provenientes de uma GEV de parâmetros $\mu = 0, \sigma = 1, \alpha = 0,5$ com $\varepsilon = 0,45$. As estatísticas sumárias e os erros quadráticos médios

amostrais (EQM) obtidos encontam-se nas Tabelas 4.10, 4.11, 4.12, 4.13 e 4.14.

Pela Tabela 4.5 (tamanho amostral 15) observa-se que os erros quadráticos médios amostrais de $\hat{\alpha}$ e $\hat{\sigma}$, estimados por L-momentos, apresentaram-se, respectivamente, menores e maiores em relação aos erros quadráticos médios das estimativas obtidas pelo Método de Máxima Verossimilhança. Já o erro quadrático médio de $\hat{\mu}$ apresentou-se praticamente igual nos dois métodos para 100 amostras e, para 30 amostras, o EQM correspondente ao Método de Máxima Verossimilhança apresentou-se menor.

Pela Tabela 4.6 (tamanho amostral 50) observa-se que os erros quadráticos médios amostrais de $\hat{\alpha}$ e $\hat{\sigma}$, estimados por L-momentos, apresentaram-se, ligeiramente, maiores. Já o erro quadrático médio de $\hat{\mu}$ apresentou-se praticamente igual nos dois métodos para 100 amostras e, para 30 amostras, o EQM correspondente ao Método de Máxima Verossimilhança apresentou-se menor.

Pelas Tabelas 4.7 (tamanho amostral 100), 4.8 (tamanho amostral 500) e 4.9 (tamanho amostral 1000) observa-se que os erros quadráticos médios amostrais de $\hat{\alpha}$, $\hat{\mu}$ e $\hat{\sigma}$, estimados por L-momentos, apresentaram-se maiores.

Pela Tabela 4.10 (tamanho amostral 15) observa-se que os erros quadráticos médios amostrais de $\hat{\alpha}$ e $\hat{\sigma}$, estimados por L-momentos, apresentaram-se menores. Já o erro quadrático médio de $\hat{\mu}$ apresentou-se praticamente igual nos dois métodos.

Pelas Tabelas 4.11 (tamanho amostral 50), 4.12 (tamanho amostral 100), 4.13 (tamanho amostral 500), 4.14 (tamanho amostral 1000) observa-se que os erros quadráticos médios amostrais de $\hat{\alpha}$, $\hat{\mu}$ e $\hat{\sigma}$ apresentaram-se praticamente iguais nos dois métodos.

Em geral, pelas Tabelas 4.5 e 4.10, o Método dos L-momentos mostrou-se mais eficiente, respectivamente, para estimar o parâmetro de forma e os parâmetros de escala e forma. Para grandes amostras, o Método de Máxima Verossimilhança apresentou-se, ligeiramente, mais apropriado que o Método dos L-momentos para $\alpha < 0$ e ambos os métodos de estimação apresentaram-se, praticamente, equivalentes nos valores das estimativas de todos os parâmetros para $\alpha > 0$.

Vale observar que as estimativas $\hat{\alpha}$, $\hat{\mu}$ e $\hat{\sigma}$ foram obtidas utilizando-se os Métodos da Bissecção e Nelder-Mead. Outros métodos numéricos poderiam ser utilizados como, por exemplo, o Método de Newton [**?**], para obtenção de raiz de equações, em vez do Método da Bissecção e o Método de Newton [**?**], para maximização de funções, em vez do Método de Nelder-Mead sendo que os Métodos de Newton citados necessitam de derivadas de primeira e segunda ordem. Adicionalmente,

no Anexo, serão apresentados e comparados os resultados obtidos a partir de algumas simulações feitas utilizando-se os métodos numéricos de Nelder-Mead e Newton para a maximização de uma função e os métodos da Bissecção e Newton-Raphson para a obtenção das raizes de uma equação.

Do fato que os Métodos de L-momentos e Máxima Verossimilhança necessitaram de rotinas numéricas, os valores das estimativas fornecidas pelos métodos depende dos valores iniciais dados a eles. A fim de observar-se o fato descrito, gerou-se 100 amostras independentes de tamanho amostral 1000 da Distribuição GEV com parâmetros $\mu = 0, \sigma = 1, \alpha = 0,7$. Com isso, foi observado o efeito de ε nas estimativas de μ, σ e α. As estatísticas sumárias e o erros quadráticos médios amostrais (EQM) obtidos encontam-se nas Tabelas 4.15 e 4.16.

Pelas Tabelas 4.15 e 4.16 observa-se que o Método de Máxima Verossimilhança, através do Método de Nelder-Mead, apresentou-se mais sensível aos valores de ε pois, para ε=0,27; 0,28 e 2, a média das estimativas obtidas de α afastou-se, consideravelmente, do valor $\alpha = 0,7$. Já para ε=0,25 e 0,26, a média das estimativas obtidas de α apresentou-se próxima do valor $\alpha = 0,7$. Cabe salientar que o Método dos L-momentos, através do Método da Bissecção, apresentou-se mais resistente às perturbações ε obtendo estimativas mais próximas dos valores reais dos parâmetros.

Os métodos numéricos utilizados, em geral, não convergiram para tamanhos amostrais pequenos como tamanho 15, no contexto das simulações apresentadas na seção. Este mesmo problema foi enfrentado por Hosking [**?**].

Além dos erros quadráticos médios das estimativas dos parâmetros, obtidas através do Método dos L-momentos para tamanhos amostrais pequenos como tamanho 15, serem menores, este método convergiu a maioria das simulações em relação ao Método de Máxima Verossimilhança. Frequentemente, o Método de Máxima Verossimilhança não convergiu, principalmente, para tamanhos amostrais pequenos.

Em relação às propriedades assintóticas como, por exemplo, normalidade assintótica, o Método de Máxima Verossimilhança aplicado à Distribuição GEV possui tais propriedades asseguradas quando $-\infty < \alpha < 1/2$ sendo que não há restrições quanto à normalidade assintótica no Método de L-momentos. Além disso, o Método de Máxima Verossimilhança apresentou-se mais sensível às condições inicias.

Vale observar que os dois métodos de estimação apresentados podem ser utilizados simultaneamente, ou seja, as estimativas obtidas pelo Método dos L-momentos, por ser mais resistente em

relação às aproximações iniciais, podem ser utilizadas como estimativas iniciais no Método de Máxima Verossimilhança e, também, pode-se obter, para diferentes valores das aproximações iniciais, as estimativas dos parâmetros e observar a concordância destas em ambos os métodos.

As expressões dos estimadores de μ, σ e α dados pelas equações (2.15) e (2.16) foram obtidos pelos cáculos dos L-momentos λ_1, λ_2 e λ_3 utilizando-se a Distribuição GEV com $\alpha \neq 0$. Da mesma forma, a equação de log-verossimilhança (3.21) foi obtida utilizando-se a Distribuição GEV com $\alpha \neq 0$. Isso é devido ao fato de que, na prática, não se conhece os valores dos parâmetros da distribuição populacional, isto é, deve ser utilizada a expressão da Distribuição GEV que forneça os valores $\alpha \neq 0$. No caso em que α é um valor próximo de zero, deve ser feito um teste estatístico para verificar se α pode ser considerado nulo ou não. A Tabela 4.17 mostra que, para amostras geradas a partir da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$ e $\alpha = 0$, as estimativas obtidas para o parâmetro α são próximas de zero.

Para testar a hipótese $\alpha = 0$ em uma Distribuição GEV (teste para verificar se a distribuição é a Distribuição Gumbel), pode-se utilizar o Teste de Hosking [**?**], cuja hipótese nula é $H_o : \alpha = 0$. Neste teste, a estatística é $Z = (\hat{\alpha})(\frac{n}{0.5633})^{\frac{1}{2}}$ onde $\hat{\alpha}$ é a estimativa obtida pelo Método dos L-momentos de α. Esta estatística deve ser comparada com os valores de uma Normal Padrão sendo que valores positivos significantes de Z rejeitam a hipótese nula em favor da Distribuição Weibull enquanto que valores negativos significantes de Z rejeitam a hipótese nula em favor da Distribuição Fréchet. Se os parâmetros da Distribuição GEV forem estimados utilizando-se o Método de Máxima Verossimilhança, pode-se utilizar o Teste de Razão das Verossimilhanças [**?**] cuja estatística é dada por $\Lambda = -2(L_{GUMBEL} - L_{GEV})$ onde L_{GUMBEL} e L_{GEV} são as funções de log-verossimilhança, respectivamente, das distribuições Gumbel $G_{\hat{\mu},\hat{\sigma}}$ e GEV ($G_{\hat{\mu},\hat{\sigma},\hat{\alpha}}$) com as respectivas estimativas por máxima verossimilhança. A estatística do teste Λ deve ser comparada com uma distribuição qui-quadrado com grau de liberdade 1 (χ^2_1) para algum nível de significância pré-estabelecido.

Tabela 4.5: Estatísticas provenientes de amostras de tamanho 15 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,485	-0,135	-0,028	0,290	1,238	0,104	0,384	0,153
30	LM	$\hat{\sigma}$	0,455	0,735	0,984	1,128	3,327	1,138	0,673	0,457
		$\hat{\alpha}$	-0,977	-0,588	-0,441	-0,327	-0,051	-0,468	0,201	0,056
		$\hat{\mu}$	-0,578	-0,172	-0,014	0,156	0,470	-0,0051	0,245	0,058
30	MV	$\hat{\sigma}$	0,446	0,669	0,881	1,087	1,80	0,895	0,322	0,111
		$\hat{\alpha}$	-1,61	-1,01	-0,731	-0,458	-0,203	-0,774	0,371	0,163
		$\hat{\mu}$	-0,470	-0,173	0,0054	0,243	0,719	0,032	0,274	0,075
100	LM	$\hat{\sigma}$	0,398	0,769	0,960	1,252	2,345	1,042	0,383	0,147
		$\hat{\alpha}$	-0,885	-0,593	-0,468	-0,322	-0,051	-0,454	0,208	0,064
		$\hat{\mu}$	-0,644	-0,232	-0,039	0,158	0,558	-0,030	0,275	0,076
100	MV	$\hat{\sigma}$	0,368	0,711	0,896	1,048	1,530	0,889	0,256	0,077
		$\hat{\alpha}$	-2,037	-0,937	-0,672	-0,489	-0,103	-0,716	0,357	0,140

Tabela 4.6: Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,235	-0,058	-0,0022	0,097	0,209	0,0028	0,115	0,012
30	LM	$\hat{\sigma}$	0,699	0,892	0,996	1,129	1,378	0,991	0,167	0,027
		$\hat{\alpha}$	-0,713	-0,515	-0,462	-0,419	-0,324	-0,474	0,096	0,024
		$\hat{\mu}$	-0,199	-0,142	-0,025	0,036	0,232	-0,036	0,109	0,01
30	MV	$\hat{\sigma}$	0,692	0,843	0,894	0,989	1,151	0,915	0,126	0,022
		$\hat{\alpha}$	-0,833	-0,672	-0,6107	-0,512	-0,343	-0,597	0,121	0,014
		$\hat{\mu}$	-0,381	-0,103	0,015	0,107	0,530	0,015	0,167	0,028
100	LM	$\hat{\sigma}$	0,522	0,865	0,975	1,135	1,698	1,007	0,205	0,041
		$\hat{\alpha}$	-0,924	-0,628	-0,504	-0,387	-0,198	-0,522	0,167	0,033
		$\hat{\mu}$	-0,436	-0,127	-0,022	0,086	0,469	-0,015	0,168	0,028
100	MV	$\hat{\sigma}$	0,485	0,827	0,906	1,070	1,287	0,945	0,164	0,029
		$\hat{\alpha}$	-0,978	-0,741	-0,622	-0,518	-0,281	-0,630	0,162	0,027

Tabela 4.7: Estatísticas provenientes de amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,182	-0,049	0,0087	0,058	0,293	0,014	0,095	0,0090
30	LM	$\hat{\sigma}$	0,717	0,953	1,011	1,100	1,495	1,017	0,156	0,023
		$\hat{\alpha}$	-0,834	-0,562	-0,476	-0,425	-0,300	-0,511	0,123	0,022
		$\hat{\mu}$	-0,179	-0,098	-0,011	0,047	0,179	-0,018	0,096	0,0093
30	MV	$\hat{\sigma}$	0,723	0,870	0,952	1,043	1,228	0,958	0,120	0,015
		$\hat{\alpha}$	-0,822	-0,669	-0,611	-0,556	-0,418	-0,605	0,098	0,0094
		$\hat{\mu}$	-0,277	-0,069	-0,0021	0,076	0,371	0,0064	0,113	0,012
100	LM	$\hat{\sigma}$	0,706	0,907	1,003	1,106	1,455	1,010	0,152	0,023
		$\hat{\alpha}$	-0,852	-0,639	-0,521	-0,447	-0,180	-0,534	0,135	0,022
		$\hat{\mu}$	-0,256	-0,081	-0,0205	0,046	0,278	-0,020	0,092	0,0088
100	MV	$\hat{\sigma}$	0,706	0,897	0,969	1,035	1,226	0,964	0,107	0,012
		$\hat{\alpha}$	-0,831	-0,692	-0,608	-0,5309	-0,216	-0,611	0,113	0,012

Tabela 4.8: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,116	-0,031	0,0103	0,0607	0,112	0,0083	0,063	0,0039
30	LM	$\hat{\sigma}$	0,825	0,988	1,051	1,083	1,136	1,029	0,078	0,0068
		$\hat{\alpha}$	-0,681	-0,611	-0,568	-0,535	-0,455	-0,571	0,057	0,0040
		$\hat{\mu}$	-0,112	-0,034	-0,0029	0,035	0,079	-0,0060	0,051	0,0025
30	MV	$\hat{\sigma}$	0,872	0,964	1,021	1,044	1,088	1,005	0,055	0,0030
		$\hat{\alpha}$	-0,691	-0,637	-0,606	-0,579	-0,513	-0,609	0,042	0,0017
		$\hat{\mu}$	-0,109	-0,033	0,0093	0,051	0,158	0,011	0,055	0,0031
100	LM	$\hat{\sigma}$	0,858	0,962	1,029	1,075	1,175	1,023	0,071	0,0055
		$\hat{\alpha}$	-0,802	-0,615	-0,568	-0,525	-0,428	-0,571	0,064	0,0048
		$\hat{\mu}$	-0,130	-0,034	-0,0025	0,025	0,121	-0,0044	0,049	0,0024
100	MV	$\hat{\sigma}$	0,870	0,960	0,995	1,034	1,114	0,996	0,050	0,0025
		$\hat{\alpha}$	-0,755	-0,642	-0,613	-0,586	-0,497	-0,611	0,047	0,0023

Tabela 4.9: Estatísticas provenientes de amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,094	-0,018	-0,0021	0,035	0,078	0,0043	0,040	0,0016
30	LM	$\hat{\sigma}$	0,882	0,994	1,031	1,053	1,114	1,019	0,057	0,0035
		$\hat{\alpha}$	-0,734	-0,614	-0,583	-0,542	-0,480	-0,584	0,058	0,0035
		$\hat{\mu}$	-0,0554	-0,027	-0,0031	0,015	0,069	-0,0035	0,032	0,0010
30	MV	$\hat{\sigma}$	0,942	0,984	1,003	1,031	1,070	1,006	0,034	0,0011
		$\hat{\alpha}$	-0,675	-0,627	-0,604	-0,585	-0,544	-0,606	0,036	0,0013
		$\hat{\mu}$	-0,100	-0,022	0,0056	0,038	0,098	0,0047	0,045	0,0020
100	LM	$\hat{\sigma}$	0,802	0,970	1,012	1,059	1,161	1,012	0,066	0,0045
		$\hat{\alpha}$	-0,869	-0,618	-0,577	-0,540	-0,453	-0,588	0,075	0,0057
		$\hat{\mu}$	-0,089	-0,028	-0,0063	0,020	0,774	-0,0040	0,036	0,0013
100	MV	$\hat{\sigma}$	0,898	0,972	1,001	1,023	1,080	0,999	0,039	0,0015
		$\hat{\alpha}$	-0,708	-0,631	-0,607	-0,584	-0,551	-0,608	0,034	0,0012

Tabela 4.10: Estatísticas provenientes de amostras de tamanho 15 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,405	-0,188	0,060	0,198	0,694	0,026	0,267	0,070
30	LM	$\hat{\sigma}$	0,517	0,888	0,984	1,023	1,221	0,945	0,163	0,028
		$\hat{\alpha}$	0,247	0,361	0,447	0,651	0,802	0,491	0,162	0,025
		$\hat{\mu}$	-0,374	-0,137	0,106	0,248	0,711	0,070	0,264	0,072
30	MV	$\hat{\sigma}$	0,480	0,854	0,950	0,992	1,187	0,917	0,169	0,034
		$\hat{\alpha}$	0,230	0,445	0,544	0,699	0,918	0,559	0,181	0,035
		$\hat{\mu}$	-0,580	-0,194	0,014	0,147	0,735	-0,0040	0,265	0,069
100	LM	$\hat{\sigma}$	0,517	0,86	0,988	1,065	1,512	0,967	0,182	0,033
		$\hat{\alpha}$	0,180	0,364	0,472	0,594	0,806	0,482	0,157	0,024
		$\hat{\mu}$	-0,505	-0,132	0,064	0,197	0,721	0,0531	0,258	0,069
100	MV	$\hat{\sigma}$	0,480	0,814	0,954	1,041	1,355	0,939	0,179	0,035
		$\hat{\alpha}$	0,230	0,445	0,586	0,699	0,918	0,571	0,164	0,031

Tabela 4.11: Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,343	-0,086	0,019	0,084	0,416	0,014	0,162	0,025
30	LM	$\hat{\sigma}$	0,717	0,913	0,964	1,043	1,295	0,9827	0,122	0,014
		$\hat{\alpha}$	0,296	0,417	0,501	0,575	0,710	0,495	0,114	0,012
		$\hat{\mu}$	-0,325	-0,065	0,044	0,118	0,385	0,033	0,159	0,025
30	MV	$\hat{\sigma}$	0,691	0,915	0,957	1,034	1,262	0,976	0,120	0,014
		$\hat{\alpha}$	0,322	0,442	0,523	0,588	0,781	0,526	0,108	0,011
		$\hat{\mu}$	-0,332	-0,026	0,056	0,131	0,458	0,048	0,143	0,022
100	LM	$\hat{\sigma}$	0,724	0,890	0,978	1,047	1,178	0,972	0,104	0,011
		$\hat{\alpha}$	0,255	0,417	0,502	0,575	0,780	0,499	0,109	0,0118
		$\hat{\mu}$	-0,329	0,00061	0,087	0,158	0,493	0,074	0,140	0,024
100	MV	$\hat{\sigma}$	0,702	0,889	0,970	1,040	1,192	0,968	0,107	0,012
		$\hat{\alpha}$	0,282	0,475	0,545	0,612	0,842	0,543	0,099	0,0116

Tabela 4.12: Estatísticas provenientes de amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,164	-0,062	0,046	0,087	0,264	0,023	0,105	0,011
30	LM	$\hat{\sigma}$	0,823	0,937	0,980	1,036	1,138	0,977	0,075	0,0059
		$\hat{\alpha}$	0,309	0,439	0,495	0,529	0,616	0,483	0,071	0,0052
		$\hat{\mu}$	-0,165	-0,045	0,057	0,119	0,232	0,039	0,101	0,011
30	MV	$\hat{\sigma}$	0,819	0,933	0,978	1,035	1,140	0,979	0,078	0,0063
		$\hat{\alpha}$	0,378	0,482	0,508	0,551	0,619	0,513	0,054	0,0030
		$\hat{\mu}$	-0,222	-0,057	0,023	0,087	0,217	0,0107	0,1004	0,0101
100	LM	$\hat{\sigma}$	0,775	0,945	1,001	1,053	1,156	0,991	0,082	0,0068
		$\hat{\alpha}$	0,283	0,441	0,495	0,565	0,669	0,501	0,086	0,0073
		$\hat{\mu}$	-0,202	-0,041	0,027	0,098	0,238	0,023	0,1004	0,0105
100	MV	$\hat{\sigma}$	0,769	0,940	1,000	1,049	1,147	0,991	0,081	0,0066
		$\hat{\alpha}$	0,313	0,490	0,526	0,561	0,721	0,524	0,062	0,0044

Tabela 4.13: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,095	-0,027	0,0032	0,028	0,127	-0,0016	0,048	0,0022
30	LM	$\hat{\sigma}$	0,901	0,965	0,989	1,018	1,085	0,994	0,043	0,0018
		$\hat{\alpha}$	0,415	0,468	0,495	0,522	0,575	0,495	0,036	0,0013
		$\hat{\mu}$	-0,104	-0,014	0,0047	0,036	0,117	0,0054	0,046	0,0020
30	MV	$\hat{\sigma}$	0,902	0,971	0,993	1,024	1,088	0,997	0,042	0,0017
		$\hat{\alpha}$	0,446	0,493	0,508	0,522	0,584	0,508	0,030	0,0009
		$\hat{\mu}$	-0,156	-0,027	0,0018	0,027	0,127	-0,00085	0,047	0,0022
100	LM	$\hat{\sigma}$	0,901	0,965	0,9900	1,018	1,085	0,993	0,0397	0,0016
		$\hat{\alpha}$	0,408	0,470	0,497	0,522	0,577	0,495	0,038	0,0014
		$\hat{\mu}$	-0,147	-0,019	0,0064	0,031	0,117	0,00579	0,045	0,0021
100	MV	$\hat{\sigma}$	0,902	0,966	0,9918	1,020	1,088	0,994	0,0395	0,0015
		$\hat{\alpha}$	0,446	0,490	0,506	0,523	0,584	0,506	0,026	0,0007

Tabela 4.14: Estatísticas provenientes de amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,45$

n	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,052	-0,0051	0,023	0,041	0,086	0,020	0,035	0,00158
30	LM	$\hat{\sigma}$	0,940	0,973	0,994	1,008	1,047	0,991	0,024	0,0006
		$\hat{\alpha}$	0,441	0,494	0,507	0,518	0,544	0,503	0,023	0,0005
		$\hat{\mu}$	-0,049	-0,0109	0,015	0,049	0,076	0,018	0,034	0,00151
30	MV	$\hat{\sigma}$	0,943	0,977	0,995	1,010	1,041	0,993	0,022	0,0005
		$\hat{\alpha}$	0,468	0,494	0,503	0,511	0,532	0,503	0,016	0,0002
		$\hat{\mu}$	-0,080	-0,022	0,0055	0,036	0,093	0,0066	0,037	0,0014
100	LM	$\hat{\sigma}$	0,9246	0,981	1,001	1,017	1,058	0,998	0,025	0,00066
		$\hat{\alpha}$	0,441	0,487	0,507	0,518	0,568	0,503	0,025	0,00067
		$\hat{\mu}$	-0,083	-0,018	0,0041	0,033	0,093	0,0064	0,035	0,0013
100	MV	$\hat{\sigma}$	0,9243	0,983	1,000	1,014	1,054	0,998	0,024	0,00060
		$\hat{\alpha}$	0,464	0,493	0,503	0,517	0,537	0,504	0,016	0,00028

Tabela 4.15: Estimativas provenientes de 100 amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,7$

ε	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,090	-0,029	0,0022	0,029	0,097	-0,0018	0,039	0,0015
0,25	LM	$\hat{\sigma}$	0,929	0,980	0,997	1,016	1,088	0,998	0,029	0,0008
		$\hat{\alpha}$	0,637	0,682	0,703	0,719	0,774	0,700	0,029	0,0008
		$\hat{\mu}$	-0,083	-0,027	-0,0011	0,031	0,179	0,0006	0,041	0,0016
0,25	MV	$\hat{\sigma}$	0,936	0,982	0,999	1,021	1,209	1,002	0,035	0,0012
		$\hat{\alpha}$	0,658	0,688	0,701	0,716	0,982	0,704	0,033	0,0011
		$\hat{\mu}$	-0,081	-0,016	0,0069	0,034	0,091	0,0064	0,034	0,0012
0,26	LM	$\hat{\sigma}$	0,929	0,981	1,002	1,020	1,068	1,000	0,029	0,00083
		$\hat{\alpha}$	0,619	0,683	0,705	0,722	0,775	0,704	0,031	0,0009
		$\hat{\mu}$	-0,086	-0,014	0,0049	0,032	0,091	0,0082	0,033	0,0011
0,26	MV	$\hat{\sigma}$	0,936	0,980	1,002	1,020	1,067	1,001	0,028	0,00081
		$\hat{\alpha}$	0,666	0,693	0,706	0,720	0,762	0,707	0,018	0,0004
		$\hat{\mu}$	-0,076	-0,019	0,008	0,0296	0,099	0,0056	0,035	0,0012
0,27	LM	$\hat{\sigma}$	0,929	0,983	1,002	1,020	1,069	1,001	0,028	0,0008
		$\hat{\alpha}$	0,633	0,685	0,706	0,728	0,775	0,706	0,031	0,0010
		$\hat{\mu}$	-0,077	0,204	0,208	0,209	0,212	0,185	0,069	0,039
0.27	MV	$\hat{\sigma}$	0,946	1,210	1,213	1,215	1,341	1,194	0,064	0,041
		$\hat{\alpha}$	0,660	1,001	1,002	1,003	1,007	0,973	0,092	0,083

Tabela 4.16: Estimativas provenientes de 100 amostras de tamanho 1000 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,7$

ε	Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
		$\hat{\mu}$	-0,090	-0,020	0,010	0,032	0,085	0,0064	0,037	0,0014
0,28	LM	$\hat{\sigma}$	0,929	0,982	0,998	1,016	1,058	0,999	0,027	0,0007
		$\hat{\alpha}$	0,638	0,684	0,707	0,728	0,771	0,706	0,030	0,0009
		$\hat{\mu}$	0,198	0,212	0,214	0,216	0,220	0,214	0,0035	0,045
0,28	MV	$\hat{\sigma}$	1,207	1,219	1,221	1,222	1,224	1,220	0,0029	0,048
		$\hat{\alpha}$	1,008	1,012	1,013	1,014	1,017	1,013	0,0018	0,097
		$\hat{\mu}$	-0,093	-0,016	-0,0017	0,018	0,0812	0,0	0,031	0,0009
2	LM	$\hat{\sigma}$	0,918	0,985	1,003	1,020	1,091	1,002	0,028	0,0008
		$\hat{\alpha}$	0,577	0,682	0,7002	0,715	0,767	0,708	0,030	0,0009
		$\hat{\mu}$	0,283	0,303	0,308	0,312	0,318	0,307	0,0070	0,094
2	MV	$\hat{\sigma}$	2,443	2,452	2,454	2,455	2,458	2,453	0,0029	2,111
		$\hat{\alpha}$	2,188	2,202	2,206	2,212	2,222	2,207	0,0074	2,270

Tabela 4.17: Estimativas provenientes de 100 amostras de tamanho 100 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0$ e perturbação $\varepsilon = 0,15$

Método	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
	$\hat{\mu}$	-0,201	-0,063	0,0078	0,091	0,278	0,014	0,105	0,0112
LM	$\hat{\sigma}$	0,814	0,937	0,985	1,033	1,258	0,986	0,073	0,0054
	$\hat{\alpha}$	-0,144	-0,0508	-0,0129	0,046	0,146	-0,0038	0,067	0,0045
	$\hat{\mu}$	-0,193	-0,051	0,023	0,107	0,276	0,022	0,106	0,0116
MV	$\hat{\sigma}$	0,842	0,938	0,982	1,025	1,201	0,981	0,067	0,0047
	$\hat{\alpha}$	-0,168	-0,034	-0,0082	0,034	0,171	0,0029	0,068	0,0046

Capítulo 3

Cópula

3.1 Introdução

Em Estatística é desejável o conhecimento da função de distribuição acumulada de uma variável aleatória a fim de se estudar o seu comportamento. Quando estamos interessados em mais que uma variável aleatória devemos conhecer a função de distribuição acumulada conjunta que, geralmente, é mais difícil de se obter, em comparação às funções de distribuição acumulada marginais associadas à tais variáveis aleatórias.

Pelo resultado obtido por Sklar [**?**], a função de distribuição acumulada conjunta $F_{\mathbf{X}}$ pode ser obtida através de suas funções de distribuição acumulada marginais e da cópula, o objetivo deste capítulo será a estimação da estrutura de dependência dada por uma distribuição $C_{\mathbf{X}}$ associada à $\mathbf{X}$, isto é, $C_{\mathbf{X}}$ é uma função de distribuição acumulada conjunta de $\mathbf{U} = (U_1, ..., U_n)$ onde $U_i = F_{X_i}(X_i)$, $i = 1, ..., n$ sendo F_{X_i} a função de distribuição acumulada marginal associada à v.a. X_i, $i = 1, ..., n$, e $C_{\mathbf{X}}$ a estrutura de dependência denominada cópula [**?**].

Vale observar que a cópula é uma função de $(U_1, ..., U_n)$ onde $U_i = F(X_i)$, $i = 1, ..., n$. Isto significa, por exemplo, quando se tem valores observados de duas v.a. X e Y, a cópula modela a ordem dos valores observados das v.a. X e Y pois considera-se a f.d.a. de cada v.a. sendo que a ordem da amostra é algo inerente à esta função. Neste caso, pode-se pensar que tais valores observados transmitem a informação de dependência através, por exemplo, de um Gráfico de Dispersão. Mas,

Figura 3.1: À esquerda: Gráfico de Dispersão dos 1000 valores gerados das v.a. independentes $X \sim Exp(2)$ e $Y \sim Exp(10)$. À direita: Gráfico de Dispersão dos 1000 valores gerados das v.a. $X \sim Exp(2)$ e $Y \sim Exp(10)$ avaliados em suas respectivas f.d.a.

com isto pode-se obter uma falsa impressão de dependência. Gerando-se, por exemplo, 1000 valores das v.a. independentes $X \sim Exp(2)$ e $Y \sim Exp(10)$, o Gráfico de Dispersão dos valores gerados de X versus os valores de Y indica uma dependência pois os valores estarão concentrados em uma mesma região já que a esperança das v.a. X e Y são, respectivamente, 0,5 e 0,1. Agora, com o Gráfico de Dispersão dos 1000 valores gerados das v.a. $X \sim Exp(2)$ e $Y \sim Exp(10)$ avaliados em suas respectivas f.d.a., $F_X(x) = 1 - e^{-2x}$ e $F_Y(y) = 1 - e^{-10y}$, não haverá indicação de dependências entre os valores gerados como mostra os gráficos a seguir [**?**].

Há vários modelos de cópula que podem ser utilizados dependendo do tipo de dependência e do número de variáveis aleatórias a serem descritas. Por isso, nas próximas seções, serão apresentados alguns conceitos, propriedades e exemplos de dependência.

3.2 Tipos de Dependência entre Variáveis Aleatórias

O estudo dos vários tipos de dependência é importante pois um dado modelo de cópula pode ser mais adequado para um tipo de dependência do que para outro. Por exemplo, há modelos de cópula que modelam dependências positivas (limitante superior de Fréchet [**?**]) enquanto que há modelos de cópula que modelam dependências negativas (limitante inferior de Fréchet [**?**]).

Há várias formas para caracterizar dependências. Para uma dada amostra, pode ser mais conveniente a modelagem utilizando-se um conceito mais forte de dependência (por exemplo, MTP_2 ou MRR_2) do que um conceito de dependência mais fraco (por exemplo, PQD ou NQD).

3.2.1 Funções Totalmente Positivas e Totalmente Negativas

Nesta seção, por simplicidade, serão dadas as definições e propriedades utilizando-se funções densidade. O caso discreto é análogo.

Definição 3.1. *Seja uma função $f(\mathbf{x})$, definida em $\chi = \chi_1 \times \chi_2 \times ... \times \chi_n$ onde cada χ_i é totalmente ordenado, satisfazendo*

$$f(\mathbf{x} \vee \mathbf{y}) f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) f(\mathbf{y}) \tag{2.1}$$

onde $\vee$ e $\wedge$ são operações definidas como sendo

$$\mathbf{x} \vee \mathbf{y} = (max(x_1, y_1), ..., max(x_n, y_n))$$

e

$$\mathbf{x} \wedge \mathbf{y} = (min(x_1, y_1), ..., min(x_n, y_n));$$

*$\mathbf{x} = (x_1, ..., x_n)$ e $\mathbf{y} = (y_1, ..., y_n)$. A função densidade que satisfizer (2.1) será denominada função multivariada totalmente positiva de ordem 2 ("multivariate totally positive of order 2 ") denotada por MTP_2. Um vetor aleatório $\mathbf{X} = (X_1, ..., X_n)$ de n componentes será chamado MTP_2 se sua função densidade for MTP_2 [**?**].*

Definição 3.2. *A função $f(\mathbf{x})$ que satisfizer a inequação (2.1) com a desigualdade trocada ($\leq$ no lugar de $\geq$) será chamada função multivariada totalmente negativa de ordem 2 ("multivariate reverse rule of order 2 ") denotada por MRR_2 [**?**].*

Para a análise das definições acima será considerado o caso $n = 2$. Para este caso, será dada uma interpretação para cada um dos conceitos citados.

Definição 3.3. *Uma função $f(x, y)$, não-negativa, de duas variáveis reais, definidas em $\chi_1 \times \chi_2$ com χ_1 e χ_2 totalmente ordenados será totalmente positiva de ordem 2 ("totally positive of order 2 "), denotado por TP_2, se o determinante da matriz quadrada de ordem 2, $|f(x_i, y_i)|_{i,j=1,2}$, for não-negativo, ou seja, $f(x_1, x_2)f(y_1, y_2) - f(x_1, y_2)f(y_1, x_2) \geq 0$ para toda escolha $x_1 < x_2$, $y_1 < y_2$, $x_i \in \chi_1$, $y_i \in \chi_2$, $i = 1, 2$ [**?**].*

O determinante da matriz quadrada de ordem r, $|f(x_i, y_i)|_{i,j=1,...,m}$, $1 \leq m \leq r$, é definido como sendo [**?**]:

Definição 3.4. *Seja uma função densidade de duas variáveis $f(x, y)$ definidas em $\chi = \chi_1 \times \chi_2$ com χ_1 e χ_2 totalmente ordenados. A função será totalmente positiva de ordem r, denotada por TP_r, se para todo $x_1 < ... < x_m$, $y_1 < .... < y_m$, $x_i \in \chi_1$, $y_i \in \chi_2$, $1 \leq m \leq r$, satisfizer*

$$|f(x_i, y_i)|_{i,j=1,2,...,m} = \begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) & ... & f(x_1, y_m) \\ f(x_2, y_1) & f(x_2, y_2) & ... & f(x_2, y_m) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ f(x_m, y_1) & f(x_m, y_2) & ... & f(x_m, y_m) \end{vmatrix} \geq 0$$

Para a verificação da expressão (2.1), é suficiente mostrar que $f(x_1, ..., x_n) > 0$ é TP_2 para cada par de componentes X_i, X_j do vetor $(X_1, ..., X_n)$ considerando-se as demais componentes fixas [**?**].

Por outro lado, para uma função, o determinante da matriz quadrada de ordem 2, $|f(x_i, y_i)|_{i,j=1,2}$, pode ser não-positivo. Então, define-se:

Definição 3.5. *Uma função $f(x, y)$, não-negativa, de duas variáveis reais, definidas em $\chi_1 \times \chi_2$ com χ_1 e χ_2 totalmente ordenados será totalmente negativa de ordem 2 ("reverse rule of order 2 "), denotado por RR_2, se o determinante da matriz quadrada de ordem 2, $|f(x_i, y_i)|_{i,j=1,2}$, for não-positivo, ou seja, $f(x_1, x_2)f(y_1, y_2) - f(x_1, y_2)f(y_1, x_2) \leq 0$ para toda escolha $x_1 < x_2$ e $y_1 < y_2$, $x_i \in \chi_1$, $y_i \in \chi_2$, $i = 1, 2$ [**?**].*

Quando a função f for uma função densidade, o seu domínio será real, ou seja, toma-se $\chi_1, ..., \chi_n$ como sendo $\mathbb{R}$.

A condição de dependência positiva $f(x_1, x_2)f(y_1, y_2) \geq f(x_1, y_2)f(y_1, x_2)$ significa que é mais provável que ocorram dois pares com componentes assumindo valores grande-grande ou pequeno-pequeno do que dois pares com componentes assumindo valores grande-pequeno ou pequeno-grande.

Analogamente, $f(x_1, x_2)f(y_1, y_2) - f(x_1, y_2)f(y_1, x_2) \leq 0$ é uma condição de dependência negativa.

Nas definições das dependências TP_2 e RR_2, assume-se a existência de uma função densidade. Pode-se definir tais conceitos utilizando-se o conceito de medida, ou seja, considere μ uma medida de probabilidade nos conjuntos de Borel em $\mathbb{R}^n$. Se $I_1, ..., I_n$ forem intervalos em $\mathbb{R}$, define-se $\widetilde{\mu}(I_1, ..., I_n) = \mu(I_1 \times ... \times I_n)$. Se I e J forem intervalos em $\mathbb{R}$, $I < J$ se $x \in I$, $y \in J$ implicar em $x < y$. Com isso, define-se [**?**, **?**]:

Definição 3.6. *Seja $\widetilde{\mu}$ uma medida de probabilidade em $\mathbb{R}^2$. Esta medida será RR_2 se $\widetilde{\mu}(I_1, I_2)\widetilde{\mu}(I_1^{'}, I_2^{'}) \leq \widetilde{\mu}(I_1, I_2^{'})\widetilde{\mu}(I_1^{'}, I_2)$ para todos os intervalos $I_1 < I_1^{'}$, $I_2 < I_2^{'}$ em $\mathbb{R}$.*

e

Definição 3.7. *Seja $\widetilde{\mu}$ uma medida de probabilidade em $\mathbb{R}^n$, $n \geq 2$. Esta medida será RR_2 aos pares se $\widetilde{\mu}(I_1, ..., I_n)$ for RR_2 nos pares I_i, I_j para todo $1 \leq i < j \leq n$ sendo que as demais variáveis são mantidas fixas. As v.a. $X_1, ..., X_n$ (ou o vetor aleatório $\mathbf{X}$ ou sua f.d.a F) serão RR_2 aos pares se sua correspondente medida de probabilidade em $\mathbb{R}^n$ for RR_2 aos pares.*

A seguir, serão apresentadas algumas propriedades referentes ao conceito MTP_2. A demonstração das propriedades podem ser encontradas em [**?**]. As seguintes propriedades podem ser utilizadas para a obtenção de distribuições MTP_2.

1. Sejam $f_1(\mathbf{x}), f_2(\mathbf{x})$ funções densidade em χ satisfazendo, para todo $\mathbf{x}, \mathbf{y} \in \chi$,

$$f_1(\mathbf{x})f_2(\mathbf{y}) \leq f_3(\mathbf{x} \vee \mathbf{x})f_4(\mathbf{x} \wedge \mathbf{y}) \tag{2.2}$$

   Então, para alguma função crescente φ em χ, tem-se

$$\int_{\chi} \varphi(\mathbf{x}) f_1(\mathbf{x}) d\mathbf{x} \leq \int_{\chi} \varphi(\mathbf{x}) f_2(\mathbf{x}) d\mathbf{x}$$

2. Seja f uma função densidade MTP_2 em χ. Então a função densidade marginal, φ, em $\prod_{i=1}^{k} \chi_i$ dada por
$$\varphi(x_1, ..., x_k) = \int_{\chi_n} .... \int_{\chi_{k+1}} f(x_1, ..., x_k, x_{k+1}, ..., x_n) dx_{k+1} ... dx_n$$
é MTP_2;

3. Sejam f e g forem funções densidade MTP_2. Então, a função fg será MTP_2;

4. Sejam $\chi = \prod_{i=1}^{n} \chi_i$, $\xi = \prod_{i=1}^{n} \xi_i$, $\zeta = \prod_{i=1}^{n} \zeta_i$ onde χ_i, ξ_i e ζ_i são espaços totalmente ordenados. Sejam f e g funções densidade MTP_2, respectivamente, em $\xi \times \chi$ e $\chi \times \zeta$. Define-se, $h(\mathbf{y}, \mathbf{z}) = \int_{\chi} f(\mathbf{y}, \mathbf{x}) g(\mathbf{x}, \mathbf{z}) d\sigma(\mathbf{x})$, $\sigma = \sigma_1 \times ... \times \sigma_n$. Então, h é uma função MTP_2 em $\xi \times \zeta$;

5. Se $f(\mathbf{x})$ for MTP_2, $\mathbf{x} \in \chi$, $\varphi_1, ..., \varphi_n$ forem todas funções crescentes (ou todas funções decrescentes), respectivamente, em $\chi_1, ..., \chi_n$ e $\{\phi_k(x_k)\}_{k=1,...,n}$ forem funções positivas. Então, a função $\psi(\mathbf{x}) = \psi(x_1, ..., x_n) = (\prod_{i=1}^{n} \phi_i(x_k)) f(\varphi_1(x_1), ..., \varphi_n(x_n))$ será MTP_2 em χ;

6. Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por variáveis aleatórias $X_1, ..., X_n$ independentes onde cada X_i, $i = 1, ..., n$, possui função densidade, f_{X_1}, PF_2[1]. Seja $\mathbf{Y} = (Y_1, ..., Y_n)$ um vetor aleatório com função densidade conjunta, $f_{\mathbf{Y}}$, MTP_2 em $\mathbb{R}^n$ e suponha que $\mathbf{X}$ e $\mathbf{Y}$ são independentes. Então, $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ possui uma função densidade MTP_2;

7. Seja $\mathbf{X}$ um vetor aleatório definido de acordo com o ítem 6 e seja X_o uma v.a. independente de $\mathbf{X}$ possuindo função densidade f_{X_o}. Define-se $\mathbf{Z_i} = \mathbf{X_i} + \mathbf{X_o}$, $i = 1, ..., n$. Então, a função densidade conjunta de $\mathbf{Z} = (Z_1, ..., Z_n)$ é MTP_2;

8. Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por variáveis aleatórias $X_1, ..., X_n$ independentes onde cada X_i, $i = 1, ..., n$ possui função densidade, f_{X_i}. Seja X_o uma variável aleatória positiva. Se, para $i = 1, ..., n$, ou $f_{X_i}(u/v)$ for TP_2 em $-\infty < u < \infty$ e $v > 0$, ou $f_{X_i}(uv)$ for

[1] Uma função $f(x)$ definida em $(-\infty, \infty)$ é PF_r ("Pólya frequency function of order r) se $f(x - y)$ for TP_r, $-\infty < x, y < \infty$ [?]

TP_2 em $-\infty < u < \infty$ e $v > 0$, então ambos os vetores aleatórios $\mathbf{Z} = (X_1X_o, ..., X_nX_o)$ e $\mathbf{Z} = (X_1/X_o, ..., X_n/X_o)$ terão funções densidade MTP_2;

9. Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório possuindo densidade conjunta MTP_2. Sejam ψ e φ, ambas funções crescentes (ou ambas funções decrescentes), em $\mathbb{R}^n$. Então,

$$\mathbb{E}[\varphi(\mathbf{X})\psi(\mathbf{X})] \geq (\mathbb{E}[\varphi(\mathbf{X})])(\mathbb{E}[\psi(\mathbf{X})]).$$

Um caso particular da propriedade 1 pode ser obtido tomando-se uma função positiva $\psi(\mathbf{x})$ e definindo-se $f_2(\mathbf{x}) = \frac{\psi(\mathbf{x})f(\mathbf{x})}{\gamma}$ com $\gamma = \int_\chi \psi(\mathbf{x})f(\mathbf{x})d\sigma$ e $f_1(\mathbf{x}) = f(\mathbf{x})$, quando $\psi(\mathbf{x})$ for uma função crescente para que f_1 e f_2 satisfaçam a expressão (2.2): Seja $f(\mathbf{x})$ uma função densidade com respeito à medida $d\sigma(\mathbf{x})$ em χ satisfazendo, para todo $\mathbf{x}, \mathbf{y} \in \chi$, $f(\mathbf{x})f(\mathbf{y}) \leq f(\mathbf{x} \vee \mathbf{x})f(\mathbf{x} \wedge \mathbf{y})$ então, para algum par φ e ψ em χ de funções crescentes (ou descrescentes), tem-se que

$$\int \varphi(\mathbf{x})\psi(\mathbf{x})f(\mathbf{x})d\sigma(\mathbf{x}) \geq \left(\int \varphi(\mathbf{x})f(\mathbf{x})d\sigma(\mathbf{x})\right)\left(\int \psi(\mathbf{x})f(\mathbf{x})d\sigma(\mathbf{x})\right)$$

Vale observar que, utilizando-se as propriedades 2 e 3, obtém-se a propriedade 4 e que da propriedade 3, obtém-se: Variáveis aleatórias independentes possuem função densidade conjunta MTP_2.

Da propriedade 2 obtém-se: Se $\mathbf{X} = (X_1, ..., X_n)$ for MTP_2 então qualquer subconjunto formado pelas componentes de $\mathbf{X}$, por exemplo, $(X_i, ..., X_k)$ será MTP_2 onde $2 \leq i, k < n$

As propriedades 2 à 5 são úteis para a obtenção de funções densidade MTP_2 a partir de outras funções densidade MTP_2 e as propriedades 6 à 8 são úteis para a obtenção de funções densidade MTP_2 a partir de funções densidade PF_2 ou TP_2.

E, da propriedade 9, obtém-se que

$$Cov(\varphi(\mathbf{X}), \psi(\mathbf{X})) \geq 0 \tag{2.3}$$

a qual sugere a seguinte definição [**?**, **?**]:

Definição 3.8. *Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório satisfazendo a expressão (2.3) para qualquer par de funções crescentes (ou decrescentes) ψ e φ. Os componentes de $\mathbf{X}$, $X_1, ..., X_n$, são denominados associados.*

Utilizando-se a definição anterior, obtém-se o seguinte resultado:

Teorema 3.1. *Seja* $\mathbf{X} = (X_1, ..., X_n)$ *um vetor aleatório formado por variáveis aleatórias associadas e sejam* $\varphi_1, ..., \varphi_k$ *funções não-negativas em* $\mathbb{R}^n$ *sendo todas crescentes (ou todas decrescentes). Então,*

$$\mathbb{E}\left[\prod_{i=1}^{k} \varphi_i(\mathbf{X})\right] \geq \prod_{i=1}^{k} \mathbb{E}\left[\varphi_i(\mathbf{X})\right] \tag{2.4}$$

Demonstração: Para a demonstração consulte [**?**].

Em particular, se $\varphi_i(\mathbf{X}) = \varphi_i(X_i)$, $i = 1, ..., n$, a expressão (2.4) do teorema anterior fornece

$$\mathbb{E}[\varphi_1(X_1)...\varphi_n(X_n)] \geq \prod_{i=1}^{n} \mathbb{E}[\varphi_i(X_i)] \tag{2.5}$$

Especialmente, a expressão (2.5) fornece $P(X_1 \geq c_1, ..., X_n \geq c_n) \geq \prod_{i=1}^{n} P(X_i \geq c_i)$ e $P(X_1 \leq c_1, ..., X_n \leq c_n) \geq \prod_{i=1}^{n} P(X_i \leq c_i)$. Na próxima seção, tais desigualdades designarão outros tipos de dependência.

A seguir, serão apresentadas algumas propriedades referentes à v.a. associadas [**?**].

1. Um subconjunto de v.a. associadas é associado;
2. Se dois conjuntos de v.a. associadas forem independentes, então a união será um conjunto de v.a. associadas;
3. Funções não-decrescentes de v.a. associadas são associadas;
4. Um conjunto formado por uma única v.a. é associado;
5. Variáveis aleatórias independentes são associadas;
6. Se $\mathbf{T}^{(k)} = (T_1^{(k)}, ..., T_n^{(k)})$ for um vetor aleatório com $T_i^{(k)}$, $i = 1, ..., n$, associadas para cada k e $\mathbf{T}^{(k)} \to \mathbf{T}$, em distribuição, então as componentes do vetor aleatório $\mathbf{T} = (T_1, ..., T_n)$ serão associadas.

Serão apresentados, a seguir, alguns exemplos de funções MTP_2 [**?**].

- Seja $\mathbf{X} = (X_1, ..., X_n)$ uma amostra aletória formada por variáveis aleatórias i.i.d. X_i, possuindo função densidade f. Então, a densidade conjunta das estatísticas de ordem $X_{1,n}, ..., X_{n,n}$ é MTP_2;

- Considere que $\mathbf{X} = (X_1, ..., X_n) \sim N(0, \sum)$. Esta distribuição será MTP_2 se e somente se $-\sum^{-1}$ possuir elementos fora da diagonal não-negativos;

- A função densidade da Distribuição Logística Multivariada é definida como sendo

$$f(x_1, ..., x_n) = n! exp\left\{-\sum_{i=1}^{n} x_i\right\}\left\{1 + \sum_{i=1}^{n} e^{-x_i}\right\}^{-(m+1)}$$

  O Núcleo Generalizado de Cauchy é dado por

$$k(\mathbf{y}) = \frac{1}{(1 + \sum_{i=1}^{n} y_i)^{\alpha}}$$

  onde $y_i > 0$, $i = 1, ..., n$. Este núcleo é TP_2 em cada par de variáveis. Com isso, a Distribuição Logística Multivariada é MTP_2 pelas propriedades 3 e 5 referentes à função densidade MTP_2;

- Seja $X_1, ..., X_n$ v.a. independentes com $X_i \sim Gama(\alpha_i, \beta_i)$, $\alpha_i \geq 1$, $\beta_i > 0$, $i = 1, ..., n$. A função densidade $f_{X_i}(x) = c_i x^{\alpha_i - 1} e^{-\beta_i x}$, $x > 0$ é PF_2. Se X_o for independente de $X_1, ..., X_n$, $X_o \sim Gama(\alpha_o, \beta_o)$, então o vetor $\mathbf{Z} = (X_1 + X_o, ..., X_n + X_o)$ possui Distribuição Gamma Multivariada e, pela propriedade 7 referente à função densidade MTP_2, tem-se que a sua densidade conjunta será MTP_2;

- Seja $X_1, ..., X_n$ v.a. independentes com $X_i \sim Gama(\alpha_i, \beta_i)$, $\alpha_i > 0$, $\beta_i > 0$, $i = 1, ..., n$. A função densidade $f_{X_i}(u/v) = c_i (u/v)^{\alpha_i - 1} e^{-\beta_i (u/v)}$, é TP_2 para u e v positivos. Se X_o for independente de $X_1, ..., X_n$ e $X_i \sim \chi^2_{\nu_i}$, então o vetor $\mathbf{Z} = ((X_1/\nu_1)(X_o/\nu_o)^{-1}, ..., (X_n/\nu_n)(X_o/\nu_o)^{-1})$ possui Distribuição F Multivariada e, pela propriedade 8 referente à função densidade MTP_2, tem-se que a sua densidade conjunta será MTP_2;

- Seja $\mathbf{X} = (X_1, ..., X_n) \sim N(0, I)$ e $S \sim \chi^2$, $\mathbf{X}$ e S independentes. Seja $\mathbf{Z} = (Z_1, ..., Z_n) = (X_1/S, ..., X_n/S)$. O vetor aleatório $\mathbf{Z}$ possui Distribuição de Cauchy multivariada. Então,

$(|Z_1|, ..., |Z_n|)$ possui densidade conjunta MTP_2 considerando o mesmo argumento do exemplo prévio desde que e^{-u^2/v^2} é TP_2 com $u > 0$, $v > 0$.

A seguir, serão apresentadas algumas propriedades referentes às funções MRR_2 [**?**].

1. Se $f(x_1, x_2)$ for uma função densidade RR_2 e h_1, h_2 forem ambas funções crescentes (ou ambas decrescentes) e absolutamente integráveis com respeito à f. Então,

$$\int\int h_1(x_1)h_2(x_2)f(x_1, x_2)dx_1dx_2 \leq \int h_1(x_1)f_1(x_1)dx_1 \int h_2(x_2)f_2(x_2)dx_2$$

   onde $f_1(x_1) = \int f(x_1, x_2)dx_2$ e $f_2(x_2) = \int f(x_1, x_2)dx_1$, ou seja, $f_i(x_i)$, $i = 1, 2$, são funções densidade marginais;

2. Sejam f e g forem funções MRR_2. Então, a função fg será MRR_2;

3. Se $f(\mathbf{x})$ for MRR_2 com $\mathbf{x} \in \chi$ e $\varphi_1, ..., \varphi_n$ forem todas funções crescentes (ou todas funções decrescentes), respectivamente, em $\chi_1, ..., \chi_n$. Então, a função $\psi(\mathbf{x}) = \psi(x_1, ..., x_n) = f(\varphi_1(x_1), ...\varphi_n(x_n))$ será MRR_2 em χ;

A seguir, serão apresentados alguns exemplos de distribuições MRR_2 [**?**].

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes tais que $X_i \sim Binomial(n_i, p_i)$, $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n X_i = N$. A distribuição de $\mathbf{X}$ é dada por

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{N!}{\prod_{i=1}^n x_i! \, (N - \sum_{i=1}^n x_i)!} \prod_{i=1}^n p_i^{x_i} \left(1 - \sum_{i=1}^n p_i\right)^{N - \sum_{i=1}^n x_i} \tag{2.6}$$

  onde $\mathbf{x} = (x_1, ..., x_n)$ representa um dado valor que o vetor aleatório $\mathbf{X}$ assume e $p_i \geq 0$, $i = 1, ..., n$. Esta distribuição é denominada Distribuição Multinomial e, pode-se mostrar, que esta é MRR_2;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes. A Distribuição Hipergeométrica Multivariada é dada por

$$P_{\mathbf{X}}(\mathbf{x}) = \left[\prod_{i=1}^{n} \begin{pmatrix} M_i \\ x_i \end{pmatrix}\right] \begin{pmatrix} M - \sum_{i=1}^{n} M_i \\ m - \sum_{i=1}^{n} x_i \end{pmatrix} \begin{pmatrix} M \\ m \end{pmatrix}^{-1} \tag{2.7}$$

com $\sum_{i=1}^{n} x_i = m, 0 \leq x_i \leq M_i$, $i = 1, ..., n$ e $\sum_{i=1}^{n} M_i = M$ onde $\mathbf{x} = (x_1, ..., x_n)$ representa um dado valor que o vetor aleatório $\mathbf{X}$ assume. Pode-se mostrar que esta distribuição é MRR_2;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes. A Distribuição de Dirichlet é dada

$$f(\mathbf{x}) = \frac{\Gamma(\sum_{j=0}^{n} \theta_j)}{\prod_{j=0}^{n} \Gamma(\theta_j)} \left(1 - \sum_{j=1}^{n} x_j\right)^{\theta_0 - 1} \prod_{j=1}^{n} x_j^{\theta_j - 1} \tag{2.8}$$

onde $\sum_{j=1}^{n} x_j \leq 1$, $0 \leq x_j$ e $\theta_j \geq 1$, $j = 0, .., n$ sendo $\Gamma(.)$ a Função Gama.

Pode-se mostrar que a Distribuição de Dirichlet é MRR_2.

3.2.2 Dependência do Quadrante Positivo (e Negativo) e do Octante

Considere $\mathbf{X}$ um vetor aleatório de dimensão n ($n \geq 2$) com f.d.a. F. Define-se a dependência do octante superior positivo ("positive upper orthant dependent") e a dependência do octante inferior positivo ("positive lower orthant dependent") [**?**] como sendo:

Definição 3.9. $\mathbf{X}$ *ou* F *possuem dependência do octante superior positivo (PUOD) se* $P(X_i > a_i, i = 1, ..., n) \geq \prod_{i=1}^{n} P(X_i > a_i)$, $\forall \mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$.

Definição 3.10. $\mathbf{X}$ *ou* F *possuem dependência do octante inferior positivo (PLOD) se* $P(X_i \leq a_i, i = 1, ..., n) \geq \prod_{i=1}^{n} P(X_i \leq a_i)$, $\forall \mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$.

Se as expressões que compõem as Definições 3.9 e 3.10 são ambas válidas então $\mathbf{X}$ ou F possuem dependência do octante positivo("positive orthant dependent") denotada por POD.

Intuitivamente, a expressão da Definição 3.9 significa que é mais provável que $X_1, ..., X_n$ assumam, simultaneamente, valores grandes comparado com o vetor de v.a. independentes com as mesmas correspondentes distribuições marginais univariadas.

Similarmente, serão enunciados os conceitos de dependência do octante inferior negativo ("negative lower orthant dependent"), dependência do octante superior negativo ("negative upper orthant dependent") e dependência do octante negativo ("negative orthant dependent") [**?**].

Definição 3.11. **X** *ou* F *possuem dependência do octante superior negativo (NUOD) se* $P(X_i > a_i, i = 1, ..., n) \leq \prod_{i=1}^{n} P(X_i > a_i)$, $\forall \mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$.

Definição 3.12. **X** *ou* F *possuem dependência do octante inferior negativo (NLOD) se* $P(X_i \leq a_i, i = 1, ..., n) \leq \prod_{i=1}^{n} P(X_i \leq a_i)$, $\forall \mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$.

Se as expressões que compõem as Definições 3.11 e 3.12 são ambas válidas então **X** ou F possuem dependência do octante negativo (NOD).

As expressões que compõem as definições 3.9 e 3.10, em geral, para $n \geq 3$, não se equivalem (veja um contra-exemplo na seção 3.2.5). Porém, para $n = 2$, elas se equivalem. Isto, também, vale para as definições 3.11 e 3.12. Com isso, definições acima se reduzem, para $n = 2$, às seguintes:

Seja $\mathbf{X} = (X_1, X_2)$ um vetor aleatório bivariado com f.d.a. F. Então, define-se a dependência do quadrante positivo ("positive quadrant dependent") [**?**, **?**] como sendo:

Definição 3.13. **X** *ou* F *possuem dependência do quadrante positivo (PQD) se* $P(X_1 > a_1, X_2 > a_2) \geq P(X_1 > a_1)P(X_2 > a_2)$, $\forall a_1, a_2 \in \mathbb{R}$, *ou equivalentemente,* $P(X_1 \leq a_1, X_2 \leq a_2) \geq P(X_1 \leq a_1)\ P(X_2 \leq a_2)$, $\forall a_1, a_2 \in \mathbb{R}$.

A expressão da definição 3.13 representa uma condição de dependência positiva e significa que é mais provável que X_1 e X_2 assumam, conjuntamente, valores grandes ou pequenos comparado com X_1' e X_2' onde $X_1 \stackrel{d}{=} X_1'$ e $X_2 \stackrel{d}{=} X_2'$ sendo X_1' e X_2') v.a. independentes.

Similarmente, será apresentada a dependência do quadrante negativo("negative quadrant dependent") [**?**, **?**] definido a seguir.

Definição 3.14. **X** *ou* F *possuem dependência do quadrante negativo (NQD) se* $P(X_1 \leq a_1, X_2 \leq a_2) \leq P(X_1 \leq a_1)P(X_2 \leq a_2)$, $\forall a_1, a_2 \in \mathbb{R}$.

A seguir, serão apresentadas algumas propriedades referentes ao conceito de dependência PQD e NQD [**?**].

1. O vetor de v.a. (X, X) é PQD para todo X;

2. Se o vetor de v.a. (X, Y) for PQD, então o vetor $(X, -Y)$ será NQD;

3. Sejam $(X_1, Y_1), ..., (X_n, Y_n)$ pares independentes de v.a. com funções de distribuição $F_1, ..., F_n$. Sejam r e s funções de n variáveis e sejam $X = r(X_1, ..., X_n)$ e $Y = s(Y_1, ..., Y_n)$. Então:

   (a) (X, Y) será PQD se, para cada i, as seguintes condições forem válidas:

      i. F_i é PQD e r, s são concordantes [2] para a i-ésima coordenada ou;

      ii. F_i é NQD e r, s são discordantes [3] para a i-ésima coordenada.

   (b) Similarmente, (X, Y) será NQD se, para cada i, as seguintes condições forem válidas:

      i. F_i é PQD e r, s são discordantes para a i-ésima coordenada ou;

      ii. F_i é NQD e r, s são concordantes para a i-ésima coordenada.

   (c) Sejam U e V v.a. independentes e independentes de $(X_1, Y_1), ..., (X_n, Y_n)$ e sejam $X = r(U, X_1, ..., X_n)$ e $Y = s(V, X_1, ..., X_n)$. Então, as conclusões dos ítens (a) e (b) continuam válidas.

4. Se(X, Y) for PQD e se existir $\mathbb{E}(XY)$, $\mathbb{E}(X)$ e $\mathbb{E}(Y)$, então $\mathbb{E}(XY) \geq \mathbb{E}(X)\mathbb{E}(Y)$.

A seguir, serão apresentados algums exemplos de v.a. PQD [**?**]???????. As justificativas dos exemplos baseiam-se nas propriedades anteriores.

- Para alguma v.a. X e para alguma função não-decrescente s, o vetor $(X, s(X))$ será PQD;
- $X = U + aZ$ e $Y = V + bZ$ serão PQD se a e b tiverem o mesmo sinal onde U, V e Z são v.a. independentes PQD;
- Para X e V v.a. independentes e PQD, X e $Y = X + V$ serão PQD;
- $X = r(U, Z)$ e $Y = s(V, Z)$ serão PQD se U, V e Z forem v.a. independentes PQD e r e s forem funções não-decrescentes em Z (arbitrárias para as demais v.a.).

[2] Duas funções reais r e s de n variáveis independentes serão concordantes para a i-ésima coordenada se, considerando-as funções da i-ésima coordenada, deixando as demais fixas, ambas são funções não-decrescentes ou ambas são funções não-crescentes

[3] Duas funções reais r e s de n variáveis independentes serão concordantes para a i-ésima coordenada se, considerando-as funções da i-ésima coordenada, deixando as demais fixas, uma das funções é não-decrescente e a outra é não-crescente

Agora, serão apresentadas algumas propriedades relativas ao conceito NOD [**?**]:

1. Qualquer conjunto de v.a. independentes será NOD;

2. Qualquer subconjunto de tamanho ≥ 2 formado por v.a. NOD será NOD;

3. Se $X_1, ..., X_n$ forem v.a. NOD e $\varphi_1, ..., \varphi_n$ forem funções Borel-mensuráveis crescentes com valores reais. Então, $\varphi_1(X_1), ..., \varphi_n(X_n)$ serão v.a. NOD;

4. A união de conjuntos disjuntos de v.a. NOD é NOD.

Também, serão apresentados alguns exemplos de distribuições NOD [**?**].

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes tais que $X_i \sim Binomial(n_i, p_i)$, $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} X_i = N$, $p_i \geq 0$, $i = 1, ..., n$. A Distribuição Multinomial de $\mathbf{X}$ dada por (2.6) é NOD;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes. A Distribuição Hipergeométrica Multivariada dada por (2.7) é NOD;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes. A Distribuição de Dirichlet dada por (2.8) é NOD.

3.2.3 Variáveis Aleatórias Estocasticamente Crescentes e Decrescentes, Dependência Crescente na Cauda à Direita e Dependência Decrescente na Cauda à Esquerda

Denotando-se $\Im(F_1, ..., F_n)$ como a classe das distribuições multivariadas que possuem funções de distribuição marginais $F_1, ..., F_n$, considere $\mathbf{X} = (X_1, X_2)$ um vetor aleatório bivariado possuindo f.d.a $F \in \Im(F_1, F_2)$. Define-se:

Definição 3.15. *A variável aleatória X_2 será estocasticamente crescente ("stochastically increasing"), denotada por SI, em X_1 ou a distribuição condicional $F_{2|1}$ será estocasticamente crescente se $P(X_2 > x_2 | X_1 = x_1) = 1 - F_{2|1}(x_2|x_1) \uparrow x_1, \forall x_2$ [**?**].*

Se, na Definição 3.15, trocarmos os índices 1 por 2 e 2 por 1, ter-se-á que X_1 será SI em X_2 ou $F_{1|2}$ será SI.

A expressão da definição 3.15 representa uma condição de dependência positiva e significa que a probabilidade de que X_2 ultrapasse um limiar x_2 é crescente quando X_1 aumentar.

Se, na Definição 3.15, trocarmos $\uparrow$ por $\downarrow$, a variável aleatória X_2 será estocasticamente decrescentes ("stochastically decreasing"), denotada por SD, em X_1 [**?**].

Em $P(X_2 > x_2|X_1 = x_1)$, ao invés, da v.a. X_2 ser condicionada a um valor fixo, x_1, da v.a. X_1, pode-se considerar X_2 condicionada em $X_1 > x_1$ ou $X_1 \leq x_1$. Com isso, tem-se as seguintes definições, considerando-se $(X_1, ..., X_n)$ [**?**].

Definição 3.16. *As v.a. $X_1, ..., X_n$ serão crescentes na cauda à direita em sequência ("right-tail increasing in sequence"), denotada por RTIS, se, para todo $x_i \in \mathbb{R}$, $i = 1, ..., n$, $P(X_i > x_i|X_1 > x_1, ..., X_{i-1} > x_{i-1})$ for crescente em $x_1, ..., x_{i-1}$.*

Para $n = 2$, a Definição 3.16 torna-se:

Definição 3.17. *A variável aleatória X_2 será crescente na cauda à direita ("right-tail increasing "), denotada por RTI, em X_1 se $P(X_2 > x_2|X_1 > x_1) = \frac{1-F(x_1,x_2)}{1-F_1(x_1)} \uparrow x_1, \forall x_2$ [**?**].*

Também, define-se [**?**]:

Definição 3.18. *As v.a. $X_1, ..., X_n$ serão decrescentes na cauda à esquerda em sequência ("left-tail decreasing in sequence"), denotada por RTDS, se, para todo $x_i \in \mathbb{R}$, $i = 1, ..., n$, $P(X_i \leq x_i|X_1 \leq x_1, ..., X_{i-1} \leq x_{i-1})$ for decrescente em $x_1, ..., x_{i-1}$.*

Para $n = 2$, a Definição 3.18 torna-se:

Definição 3.19. *A variável aleatória X_2 será decrescente na cauda à esquerda ("left-tail decreasing "), denotada por LTD, em X_1 se $P(X_2 \leq x_2|X_1 \leq x_1) = \frac{F(x_1,x_2)}{F_1(x_1)} \downarrow x_1, \forall x_2$ [**?**].*

As expressões nas Definições 3.17 e 3.19 são condições de dependência positiva. Na Definição 3.17 tem-se que é mais provável que X_2 assuma valores grandes quando X_1 aumentar. E, na Definição 3.19 tem-se que é mais provável que X_2 assuma valores pequenos quando X_1 diminuir. Se a direção da monocidade nas Definições 3.17 e 3.19 forem trocadas, ter-se-á conceitos referentes à dependência negativa.

As Definições 3.16 e 3.18 possuem as seguintes versões para dependência negativa [**?**]:

Definição 3.20. *As v.a.* $X_1, ..., X_n$ *serão decrescentes na cauda à direita em sequência ("right-tail decreasing in sequence"), denotada por RTDS, se, para todo* $x_i \in \mathbb{R}$, $i = 1, ..., n-1$, $P(X_{i+1} > x_{i+1} | \bigcap_{j=1}^{i}(X_j > x_j))$ *for decrescente em* $x_1, ..., x_i$.

Para $n = 2$, a partir da Definição 3.20, obtém-se que X_2 é decrecente na cauda à direita ("right-tail decresing"), denotado por RTD, em X_1.

Definição 3.21. *As v.a.* $X_1, ..., X_n$ *serão crescentes na cauda à esquerda em sequência ("right-tail decreasing in sequence"), denotada por LTIS, se, para todo* $x_i \in \mathbb{R}$, $i = 1, ..., n-1$, $P(X_{i+1} \leq x_{i+1} | \bigcap_{j=1}^{i}(X_j \leq x_j))$ *for crescente em* $x_1, ..., x_i$.

Para $n = 2$, a partir da Definição 3.21, obtém-se que X_2 é crecente na cauda à esquerda ("left-tail incresing"), denotado por LTI, em X_1.

A seguir, algumas propriedades referentes às dependências RTDS e RTD serão dadas [**?**]:

1. Um conjunto de v.a. independentes é RTDS;

2. Um subconjunto de v.a. RTDS permanece RTDS;

3. Se $X_1, ..., X_n$ forem v.a. RTDS e $\varphi_1, ..., \varphi_n$ forem funções Borel-mensuráveis crescentes, então $\varphi_1(X_1), ..., \varphi_n(X_n)$ será RTDS;

4. Seja $\mathbf{X} = (X_1, X_2)$. Então, X_2 é RTD em $X_1 \Leftrightarrow P(X_2 \geq x_2 | X_1 > x_1) \downarrow x_1$ para todo $x_2 \Leftrightarrow \mathbb{E}(\varphi(X_2) | X_1 > x_1) \downarrow x_1$ para toda função real φ crescente;

5. Seja X_2 v.a. RTD na v.a. X_1 e seja Z uma v.a. independente de (X_1, X_2). Definindo $X = X_1 + aZ$ e $Y = X_2 + bZ$, a e b constantes, tem-se que Y é RTD em X;

6. Suponha a sequência de f.d.a. H_n convergente à H e considere a sequência de v.a. multivariadas relativas à H_n sendo RTDS. Então, a v.a. multivariada relativa à H será RTDS;

7. Seja (U, V) RTD e seja Z uma v.a. independente de (U, V). Definindo φ_1 e φ_2, funções Borel-mensuráveis que mapeiam $\mathbb{R}^2$ em $\mathbb{R}$ com $\varphi_1(u, .)$ crescente em u e $\varphi_2(., v)$ crescente em v. Então, (X, Y) é RTD onde $X = \varphi_1(U, Z)$ e $Y = \varphi_2(V, Z)$;

8. Sejam $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório formado por v.a. RTDS, $\varphi_l : \mathbb{R} \to \mathbb{R}$ uma função Borel-mensurável crescente para cada $l = 1, ..., n$. Sejam $\mathbf{Z} = (Z_1, ...Z_n)$ um vetor aleatório formado por v.a. RTDS sendo $\mathbf{X}$ e $\mathbf{Z}$ independentes. Definindo $Y_l = \varphi_l(X_l) + Z_l$, $l = 1, ..., n$, tem-se que $Y_1, ..., Y_n$ são RTDS;

9. Sejam

   (a) Seja $(X_1, ..., X_n)$ um vetor aleatório tal que $\overline{F}_n(x_1, ..., x_n) = P(X_1 > x_1, ..., X_n > x_n)$ satisfaça:

      i.
$$\begin{vmatrix} \overline{F}_n(x_1, ..., x_i, ..., x_j, ..., x_n) & \overline{F}_n((x_1, ..., x_i, ..., x_j^{'}, ..., x_n) \\ \overline{F}_n((x_1, ..., x_i^{'}, ..., x_j, ..., x_n) & \overline{F}_n((x_1, ..., x_i^{'}, ..., x_j^{'}, ..., x_n) \end{vmatrix} \leq 0 \qquad (2.9)$$

      para cada par de variáveis permanecendo as demais fixas;

      ii. Se $\overline{F}_k(x_1, ..., x_k) = P(X_1 > x_1, ..., X_k > x_k)$, $1 \leq k < n$, verificar (2.9) para cada par de variáveis permanecendo as demais fixas;

   Então, $(X_1, ..., X_n)$ será RTDS e qualquer permutação de $(X_1, ..., X_n)$ será RTDS;

Há dois conceitos de dependência que podem ser considerados como uma extensão multivariada do conceito SI (Definição 3.15): dependência positiva através da ordenação estocástica ("positive dependence through the stochastic ordering") e dependência condicionalmente crescente em sequência ("conditional increasing in sequence")[**?**].

Definição 3.22. *O vetor aleatório $(X_1, ..., X_n)$ possuirá dependência positiva através da ordenação estocástica , denotada por PDS, se $[X_i, i \neq j]$ condicionado à $X_j = x$ for estocasticamente crescente quando x aumentar, para todo $j = 1, ..., n$.*

Definição 3.23. *O vetor aleatório $(X_1, ..., X_n)$ possuirá dependência condicionalmente crescente em sequência, denotada por CIS, se X_i for estocasticamente crescente em $X_1, ..., X_{i-1}$ para $i = 2, ..., n$, isto é, se $P(X_i > x_i | X_j = x_j, j = 1, ..., i-1)$ for crescente em $x_1, ..., x_{i-1}$ para todo x_i.*

Note que, para $n = 2$, PDS equivale à X_2 SI em X_1 e X_1 SI em X_2. Da mesma forma, CIS equivale à SI.

No caso em que $n = 2$ na Definição de dependência CIS, X_2 é dito ter dependência regressiva positiva ("positive regression dependent"), denotada por PRD, em X_1.

Os conceitos de v.a. PDS e CIS possuem as seguintes versões para dependência negativa: dependência negativa através da ordenação estocástica ("negative dependence through the stochastic ordering") e dependência condicionalmente decrescente em sequência ("conditional decreasing in sequence").

Definição 3.24. *O vetor aleatório $(X_1, ..., X_n)$ possuirá dependência negativa através da ordenação estocástica , denotada por NDS, se $[X_1, ..., X_{i-1}|X_i = x_i]$ for estocasticamente decrescente em x_i, para todo $i = 1, ..., n$ [**?**].*

Definição 3.25. *O vetor aleatório $(X_1, ..., X_n)$ possuirá dependência condicionalmente decrescente em sequência, denotada por CDS, se $P(X_i > x_i|X_j = x_j, j = 1, ..., i-1)$ for decrescente em $x_1, ..., x_{i-1}$ para todo x_i, $i = 2, ..., n$ [**?**].*

No caso em que $n = 2$, na Definição de dependência CIS, X_2 é dito ter dependência regressiva negativa ("negative regression dependent"), denotada por NRD, em X_1.

Para v.a. CIS e NOD, tem-se a seguinte propriedade [**?**]: Sejam $X_o, X_1, ..., X_n$ v.a. independentes possuindo funções densidade PF_2. Então, para x fixo, as v.a. condicionais $(X_1, ..., X_n|X_o + X_1 + ... + X_n = x)$ são RR_2 aos pares e, consequentemente, serão CDS e NOD.

Algumas propriedades relativas aos conceitos PRD e NRD serão dadas a seguir [**?**]:

1. (X_1, X_2) será NRD se e somente se $(-X_1, X_2)$ for PRD. Além disso, (X_1, X_2) será NRD se e somente se $(X_1, -X_2)$ for PRD;

2. Seja o vetor aleatório (X, Y) possuindo função densidade $f(x, y)$ satisfazendo

$$\begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{vmatrix} \leq 0$$

   para cada escolha $x_1 < x_2$, $y_1 < y_2$. Então, (X, Y) será NRD;

3. Sejam:

(a) $(X_1, ..., X_n)$ um vetor aleatório possuindo função densidade $f(x_1, ..., x_n)$ que satisfaz

$$\begin{vmatrix} f(x_1, ..., x_i, ..., x_j, ..., x_n) & f(x_1, ..., x_i, ..., x_j^{'}, ..., x_n) \\ f(x_1, ..., x_i^{'}, ..., x_j, ..., x_n) & f(x_1, ..., x_i^{'}, ..., x_j^{'}, ..., x_n) \end{vmatrix} \leq 0 \tag{2.10}$$

para cada par de variáveis permanecendo as demais fixas;

(b) Todas as funções de densidade marginais $f_k(x_1, ..., x_k)$, $1 \leq k < n$ satisfazendo (2.10) para cada par de variáveis permanecendo as demais fixas.

Então, $(X_1, ..., X_n)$ será CDS e cada permutação de $(X_1, ..., X_n)$ será CDS;

4. Seja o vetor aleatório $(X_1, ..., X_n)$ CDS. Então, $\mathbb{E}(\varphi(X_i)|X_1 = x_1, ..., X_{i-1} = x_{i-1})$ será decrescente em $x_1, ..., x_{i-1}$ para cada função integrável crescente φ;

5. Seja a sequênxia $\{X_n, n \geq 1\}$ de vetores aleatórios CDS p-dimensionais com f.d.a. $\{H_n, n \geq 1\}$ tais que $H_n \to H$ fracamente quando $n \to \infty$ onde H é a f.d.a. do vetor aleatório p-dimensional $\mathbf{X}$. Então, $\mathbf{X}$ será CDS.

A seguir, serão apresentados alguns exemplos relativos à alguns conceitos de dependência apresentados nesta seção [**?**]:

- Seja $\mathbf{X} = (X_1, ..., X_n)$ possuindo Distribuição Multinormal com vetor de médias $\mu = (\mu_1, ..., \mu_n)$ e matriz de variância-covariância $\sum$ definida positiva. Então, $\mathbf{X}$ é CDS;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório possuindo Distribuição Multinomial com função de distribuição de probabilidade conjunta dada por

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{N!}{\prod_{i=1}^{n} x_i! \, (N - \sum_{i=1}^{n} x_i)!} \prod_{i=1}^{n} p_i^{x_i} \left(1 - \sum_{i=1}^{n} p_i\right)^{N - \sum_{i=1}^{n} x_i}$$

onde $\mathbf{x} = (x_1, ..., x_n)$ representa um dado valor que o vetor aleatório $\mathbf{X}$ assume e $p_i \geq 0$, $i = 1, ..., n$. onde $\mathbf{x} = (x_1, ..., x_n)$ e $p_i \geq 0$, $i = 1, ..., n$ e $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} x_i = N$.

Pode-se mostrar que um subconjunto de $(X_1, ..., X_n)$, $(X_1, ..., X_k)$, é CDS;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes possuindo Distribuição de Dirichlet dada por

$$f(\mathbf{x}) = \frac{\Gamma(\sum_{j=0}^{n} \theta_j)}{\prod_{j=0}^{n} \Gamma(\theta_j)} \left(1 - \sum_{j=1}^{n} x_j\right)^{\theta_0 - 1} \prod_{j=1}^{n} x_j^{\theta_j - 1}$$

  para $0 \leq x_j, j = 1, ..., n, \sum_{j=1}^{n} x_j \leq 1$ onde $\theta_j > 1$, $j = 0, .., n$ e $\Gamma(.)$ representa a Função Gama. Pode-se mostrar que $(X_1, ..., X_n)$ é CDS;

- Nos três exemplos anteriores, no caso bivariado, as duas v.a. são mutuamente NRD e, assim, também, serão NQD.

3.2.4 Variáveis Aleatórias Negativamente Associadas

Definição 3.26. *As variáveis aleatórias $X_1, ..., X_n$ possuem associação negativa("negative association"), denotada por NA, se para cada par de subconjuntos disjuntos A_1, A_2 de $\{1, ..., n\}$ tem-se*

$$Cov(\varphi(X_i, i \in A_1), \psi(X_j, j \in A_2)) \leq 0 \tag{2.11}$$

*sempre que φ e ψ forem funções crescentes [**?**].*

Vale observar que (2.11) continua válida se φ e ψ forem funções decrescentes.

A seguir, serão apresentadas algumas propriedades referentes ao conceito de v.a. NA [**?**].

1. Sejam $A_1, ..., A_m$ subconjuntos disjuntos de $\{1, ..., n\}$ e $\phi, ..., \phi_m$ funções crescentes e positivas. Se as variáveis aleatórias, $X_1, ..., X_n$, forem NA, então:

$$\mathbb{E}\left[\prod_{i=1}^{m} \phi_i(X_j, j \in A_i)\right] \leq \prod_{i=1}^{m} \mathbb{E}[\phi_i(X_j, j \in A_i)]$$

   Uma consequência desta propriedade é dada a seguir:

   Se A_1, A_2 forem subconjuntos disjuntos de $\{1, ..., n\}$ e $x_1, ..., x_n \in \mathbb{R}$. Então,

$$P(X_i \leq x_i, i = 1, ..., n) \leq P(X_i \leq x_i, i \in A_i) P(X_j \leq x_j, j \in A_2)$$

e

$$P(X_i > x_i, i = 1, ..., n) \leq P(X_i > x_i, i \in A_i)P(X_j > x_j, j \in A_2);$$

2. Um subconjunto de duas ou mais v.a. NA é NA;

3. Um conjunto de v.a. independentes é NA;

4. Funções crescentes definidas em subconjuntos disjuntos de um conjunto de v.a. NA são NA;

5. A união de conjuntos independentes de v.a. NA é NA;

6. Sejam $X_1, ..., X_n$ v.a. independentes e suponha que a esperança condicional $\mathbb{E}(\phi(X_i, i \in A)|\sum_{i\in A} X_i)$ seja crescente em $\sum_{i\in A} X_i$, para cada função crescente f e para cada subconjunto apropriado A de $\{1, ..., n\}$. Então, a distribuição condicional de $X_1, ..., X_n$ dado $\sum X_i$ é NA, quase certamente;

7. Sejam $X_1, ..., X_n$ v.a. independentes possuindo funções densidade PF_2. Então, a distribuição condicional conjunta de $X_1, ..., X_n$ dado $\sum X_i$ é NA, quase certamente.

Uma imediada consequência da propriedade 1. é que, para A_1, A_2 subconjuntos disjuntos de $\{1, ..., n\}$ e $x_1, ..., x_n \in \mathbb{R}$, tem-se $P(X_i > x_i, i = 1, ..., n) \leq \prod_{i=1}^{n} P(X_i > x_i)$ e $P(X_i \leq x_i, i = 1, ..., n) \leq \prod_{i=1}^{n} P(X_i \leq x_i)$. Em particular, $X_1, ..., X_n$ são NOD.

A seguir serão apresentados alguns exemplos de distribuições NA [**?**].

- Seja $\mathbf{x} = (x_1, ..., x_n)$ um conjunto de n números reais. Uma distribuição de permutação é a distribuição conjunta do vetor $\mathbf{X} = (X_1, ..., X_n)$ que assume os valores de todas a $n!$ permutações de $\mathbf{x}$ com igual probabilidade, sendo cada probabilidade igual à $\frac{1}{n!}$, $n > 1$. Temos que uma distribuição de permutação é NA;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a independentes tais que $X_i \sim Binomial(n_i, p_i)$, $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} X_i = N$, $p_i \geq 0$, $i = 1, ..., n$. A Distribuição Multinomial de $\mathbf{X}$ dada por (2.6) é NA;

- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a independentes. A Distribuição Hipergeométrica Multivariada dada por (2.7) é NA;

- Seja $\mathbf{X} = (X_1, .., X_n)$ uma amostra aleatória de uma população. Seja R_i o posto de X_i, $i = 1, ..., n$. Como $\mathbf{R} = (R_1, ..., R_n)$ possui distribuição de permutação, tem-se que $\mathbf{R}$ será NA;
- Variáveis aleatórias que possuem distribuições normais negativamente correlacionadas são NA;
- Seja $\mathbf{X} = (X_1, ..., X_n)$ um vetor aleatório composto por v.a. independentes. A Distribuição de Dirichlet dada por (2.8) é NA.

3.2.5 Implicações e Contra-exemplos envolvendo os Conceitos de Dependência

Teorema 3.2. *Todos os tipos de dependências, definidos nas seções anteriores, são invariantes com respeito às transformações estritamente crescentes sobre as componentes do vetor aleatório.*

Demonstração: Para a demonstração consulte [**?**].

Por exemplo, para a dependência PQD, o teorema anterior, diz que se (X_1, X_2) for PQD então $(\varphi(X_1), \psi(X_2))$ será PQD se as funções φ, ψ forem estritamente crescentes.

Teorema 3.3. *Relações para o caso bivariado:*

(a) densidade $TP_2 \Rightarrow SI \Rightarrow LTD, RTI;$

(b) LTD ou RTI $\Rightarrow$ associação $\Rightarrow$ PQD;

(c) densidade $TP_2 \Rightarrow$ f.d.a. TP_2 e função de sobrevivência TP_2;

(d) f.d.a $TP_2 \Rightarrow$ LTD e função de sobrevivência $TP_2 \Rightarrow$ RTI.

Demonstração: Para a demonstração consulte [**?**].

Pelo teorema anterior, observa-se que a dependência TP_2 é uma dependência forte pois esta implica nas dependências SI, LTD, RTI, PQD e associação entre v.a. . A dependência mais fraca é

a PQD.

Teorema 3.4. *Relações para o caso multivariado:*

(a) um subvetor aleatório de um vetor aleatório associado é associado;

(b) associação $\Rightarrow$ PUOD e PLOD;

(c) PDS $\Rightarrow$ PUOD e PLOD;

(d) CIS $\Rightarrow$ associação.

Demonstração: Para a demonstração consulte [**?**].

A seguir, serão dadas outras implicações entre os conceitos de dependência:

1. Um par de v.a. NQD implica em NA [**?**];
2. Um par de v.a. NA implica em NOD [**?**];
3. RTDS equivale à NOD [**?**];
4. LTI equivale à NQD. Porém, para $n \geq 3$, LTIS não implica em NOD [**?**];
5. RTIS implica em POD [**?**];
6. Se (X_1, X_2) for NRD, ele será também RTD;
7. Se (X_1, X_2) for NRD, ele será também LTI;

Na sequência, serão dados contra-exemplos de implicações, entre os conceitos de dependência, que não são válidas:

- Para $n \geq 3$, NUOD e NLOD não se equivalem [**?**]

Sejam as v.a. X_1, X_2 e X_3 assumindo valores $(0,1,1),(1,0,1),(1,1,0)$ e $(0,0,0)$ sendo que cada valor assume a probabilidade $1/4$. Então,

$$P(X_1 > 0, X_2 > 0, X_3 > 0) = 0 < \frac{1}{8} = P(X_1 > 0)P(X_2 > 0)P(X_3 > 0)$$

e

$$P(X_1 \leq 0, X_2 \leq 0, X_3 \leq 0) = \frac{1}{4} > \frac{1}{8} = P(X_1 \leq 0)P(X_2 \leq 0)P(X_3 \leq 0);$$

- Nem v.a. NUOD e nem v.a. NLOD implicam em v.a. NA [**?**]

  No contra-exemplo seguinte, $\mathbf{X} = (X_1, X_2, X_3, X_4)$ será NOD mas não será NA.

  Seja o vetor aleatório $\mathbf{X} = (X_1, X_2, X_3, X_4)$ onde cada v.a. X_i possui Distribuição de Bernoulli com $P(X_i = 1) = 0,5$, $i = 1, ..., 4$. Considere os pares (X_1, X_2) e (X_3, X_4) possuindo a mesma distribuição bivariada. A distribuição conjunta de probabilidade de (X_1, X_2, X_3, X_4) é dada pela Tabela 2.1.

  Pode-se verificar que todas as condições NLOD e NUOD são válidas.

  Porém, $P(X_i = 1, i = 1, ..., 4) > P(X_1 = X_2 = 1)P(X_3 = X_4 = 1)$ viola o conceito NA;

- RTD não implica em LTI [**?**]

  Para mostrar isto, sejam (X, Y) duas v.a. tendo distribuição conjunta de probabilidade dada pela Tabela 2.2.

  Então, como

  $P(Y > 0|X \geq 0) = \frac{3}{10}$;

  $P(Y > 0|X \geq 1) = \frac{44}{165}$;

  $P(Y > 0|X \geq 2) = \frac{30}{165}$

  e

$P(Y > 0|X \geq 3) = 0$

tem-se que Y é RTD em X.

Porém, $P(Y = 0|X \leq 0) = \frac{3}{5} > \frac{5}{9} = P(Y = 0|X \leq 1)$ e, assim, Y não é LTI em X;

- LTI não equivale à RTD [**?**]

  Para mostrar que LTI não implica em RTD, considere (X, Y) duas v.a. tendo distribuição conjunta de probabilidade dada pela Tabela 2.3.

  Então, como

  $P(Y = 0|X \leq 0) = \frac{2}{5}$;

  $P(Y = 0|X \leq 1) = \frac{1}{2}$;

  $P(Y = 0|X \leq 2) = \frac{3}{5}$

  e

  $P(Y = 0|X \leq 3) = \frac{3}{5}$.

  tem-se que Y é LTI em X.

  Porém, $P(Y = 1|X > 1) = 0,30 < 0,40 = P(Y = 1|X > 2)$ e, assim, Y não é RTD em X;

- LTIS não equivale à NOD [**?**]

  Considere as v.a. X, Y e Z tais que Z condicionado em $X \leq x$ e $Y \leq y$, possui função densidade

  $$P(Z \leq z|X \leq x, Y \leq y) = 1 - exp\{-z(x + y)\} \tag{2.12}$$

  onde $z > 0$. As v.a. X e Y possuem distribuição conjunta de probabilidade dada pela Tabela 2.3.

Desde que lado direito de (2.12) é crescente em x e y e, a partir do exemplo supra-citado, (X, Y) é LTI e a sequência (X, Y, Z) é LTIS.

Porém, desde que

$$P(Z \leq z | X \leq x) = 1 - exp\{-z(x+1)\}$$

e

$$P(Z \leq z | Y \leq y) = 1 - exp\{-z(y+3)\}$$

segue que

$$P(X > 2, Y > 0, Z > 2) = e^{-8} - 1,35e^{-6} + 0,45e^{-4} > 0,1e^{-8} = P(X > 2)P(Y > 0)P(Z > 2);$$

- CIS não implica RTIS [**?**];
- RTIS não implica CIS [**?**];
- CDS não implica em RTDS [**?**]

  Para mostrar que CDS não implica em RTDS, considere X_1, X_2 e X_3 v.a. tais que

$$f_{X_3|X_1,X_2}(x_3|x_1, x_2) = x_2 e^{-x_2 x_3}$$

com $x_3 > 0$. As v.a. X_1 e X_2 possuem distribuição conjunta de probabilidade dada pela Tabela 2.4.

Então, como

$P(X_2 > 1 | X_1 = 0) = 1;$

$P(X_2 > 1 | X_1 = 1) = \frac{1}{3};$

$P(X_2 > 1 | X_1 = 2) = \frac{1}{4};$

$P(X_2 > 2 | X_1 = 0) = \frac{1}{3};$

$P(X_2 > 2 | X_1 = 1) = \frac{1}{3}$

e

$P(X_2 > 2|X_1 = 2) = 0$

tem-se que X_2 é NRD em X_1.

Também, como $P(X_3 > x|X_1 = x_1, X_2 = x_2) = exp\{-x_2 x\} \downarrow x_2$, (X_1, X_2, X_3) é CDS.

Por outro lado, utilizando-se a identidade

$$P(C|A \cup B) = P(C|A)\frac{P(A)}{P(A) + P(B)} + P(C|B)\frac{P(B)}{P(A) + P(B)}$$

quando $P(A \cap B) = 0$, $P(A) > 0$ e $P(B) > 0$, obtém-se

$$\begin{aligned} P(X_3 > x|X_1 > 0, X_2 > 1) &= \frac{1}{2}P(X_3 > x|X_1 = 1, X_2 = 3) + P(X_3 > x|X_1 = 2, X_2 = 2) = \\ &= \frac{1}{2}\left(e^{-3x} + e^{-2x}\right) \end{aligned}$$

enquanto que $P(X_3 > x|X_1 > 1, X_2 > 1) = P(X_3 > x|X_1 = 2, X_2 = 2) = e^{-2x}$.

Como $P(X_3 > x|X_1 > 1, X_2 > 1) > P(X_3 > x|X_1 > 0, X_2 > 1)$ para todo $x > 0$, (X_1, X_2, X_3) não é RTDS;

- NA não implica em RR_2 aos pares, CDS ou NDS [**?**]

  No contra-exemplo seguinte, será mostrado que $\mathbf{Y}$ é NA mas não é RR_2 aos pares, CDS ou NDS.

  Seja $\mathbf{X} = (X_1, X_2, X_3)$ um vetor aleatório possuindo uma função de probabilidade multinomial trivariada f com probabilidades p_1, p_2 e p_3 estritamente positivas e $X_1 + X_2 + X_3 = 3$. Considere o vetor aleatório induzido $\mathbf{Y} = (Y_1, Y_2)$ onde $Y_1 = X_1 X_2$ e $Y_2 = X_3$. A distribuição conjunta de probabilidade de (Y_1, Y_2) será denotada por g e é dada pela Tabela 2.5.

  A Distribuição Multinomial é RR_2 aos pares, CDS e NDS [**?**, **?**]. A função de probabilidade conjunta g não é nem RR_2, nem CDS e nem NDS.

  Pela função de distribuição de probabilidade conjunta de g, obtém-se que

$$P(\mathbf{x}) = \begin{vmatrix} g(0,0) & g(0,1) \\ g(1,0) & g(1,1) \end{vmatrix} > 0$$

Então, conclui-se que g não é RR_2.

Notando que $P(Y_2 > 0|Y_1 = 0) = 1 - P(Y_2 = 0|Y_1 = 0) < 1$ enquanto que $P(Y_2 > 0|Y_1 = 0) = 1 - P(Y_2 = 0|Y_1 = 0) = 1$, conclui-se que $\mathbf{Y}$ não é CDS.

Como, para um vetor bivariado, CDS equivale à NDS, segue que $\mathbf{Y}$ não é NDS.

Pelas propriedades 7 e 4 da seção 3.2.4, conclui-se, respectivamente, que $\mathbf{X}$ é NA. e $\mathbf{Y}$ é NA.

Portanto, NA não implica em RR_2 aos pares, CDS ou NDS.

Tabela 2.1: Distribuição conjunta de probabilidade de (X_1, X_2, X_3, X_4)

Y/X	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	0,0577	0,0623	0,0623	0,0577
(0,1)	0,0623	0,0677	0,0677	0,0623
(1,0)	0,0623	0,0677	0,0677	0,0623
(1,1)	0,0577	0,0623	0,0623	0,0577

Tabela 2.2: Distribuição conjunta de probabilidade de X e Y

Y/X	0	1	2	3
0	0,15	0,10	0,20	0,25
1	0,10	0,10	0,10	0

Tabela 2.3: Distribuição conjunta de probabilidade de X e Y

Y/X	0	1	2	3
0	0,10	0,15	0,20	0,15
1	0,15	0,10	0,05	0,10

Tabela 2.4: Distribuição conjunta de probabilidade de X e Y

Y/X	0	1	2
1	0	0,20	0,30
2	0,20	0	0,10
3	0,10	0,10	0

Tabela 2.5: Distribuição conjunta de probabilidade de (Y_1, Y_2)

Y/X	0	1	2
0	$f(0,3,0) + f(3,0,0)$	0	$f(1,2,0) + f(2,1,0)$
1	$f(0,2,1) + f(2,0,1)$	$f(1,1,1)$	0
2	$f(0,1,2) + f(1,0,2)$	0	0
3	$f(0,0,3)$	0	0

3.3 Construção de Cópulas com Dependência Positiva

Para distribuições contínuas multivariadas, em termos de modelagem, pode-se utilizar as suas funções densidade marginais univariadas e uma estrutura multivariada de dependência sendo que tal estrutura multivariada de dependência pode ser representada por uma cópula.

A cópula é a distribuição multivariada de um vetor aleatório composto por distribuições marginais univariadas $U(0,1)$. Para uma distribuição m-variada $F \in \Im(F_1(x_1), ..., F_m(x_m))$, onde F_j é a j-ésima função de distribuição marginal univariada, a cópula associada à F é uma função distribuição $C : [0,1]^m \longrightarrow [0,1]$ que satisfaz

$$F(\mathbf{x}) = C(F_1(x_1), ..., F_m(x_m)) \tag{3.13}$$

onde $\mathbf{x} \in \mathbb{R}^m$ [**?**].

Se F é uma função de distribuição contínua m-variada com marginais univariadas $F_1, ..., F_m$ e funções quantis $F_1^{-1}, ..., F_m^{-1}$, então $C(\mathbf{u}) = F(F_1^{-1}(u_1), ..., F_m^{-1}(u_m))$ é a escolha única em (3.13). Este resultado segue essencialmente a partir de duas propriedades.

1. Se H é f.d.a. univariada com inversa H^{-1} e $U \sim U(0,1)$ então $H^{-1}(U) \sim H$;

2. Se H é f.d.a. univariada contínua e $X \sim H$ então $H(X) \sim U(0,1)$.

Por isso, se $\mathbf{X} \sim F$ e F é contínua então $(F_1(X_1), ..., F_m(X_m)) \sim C$ e, se $\mathbf{U} \sim C$, $\mathbf{U} = (U_1, ..., U_m)$, então $(F_1^{-1}(U_1), ..., F_m^{-1}(U_m)) \sim F$.

Vale observar que se F for uma f.d.a. m-variada de uma v.a. discreta, a cópula associada à esta não será única pois se uma f.d.a H não for contínua ou for uma f.d.a. univariada discreta e $X \sim H$, então $H(X)$ não possui distribuição $U(0,1)$ [**?**].

Oportunamente, alguns modelos de cópula serão discutidos.

A construção de uma cópula C está vinculada com a Transformada de Laplace que será definida a seguir:

Definição 3.27. *Seja M uma função de distribuição acumulada de uma v.a. não-negativa. A Transformada de Laplace é definida por*

$$\phi(s) = \int_0^\infty e^{-sw} dM(w) \tag{3.14}$$

onde $s \geq 0$ *[?].*

Se π_o for a massa de M em 0, então $lim_{s\to\infty}\phi(s) = \phi(\infty) = \pi_o$. Durante o capítulo, será assumido que as Transformadas de Laplace correspondem à transformadas de v.a. positivas, isto é, $M(0) = 0$ ou, em termos de $\phi(s)$, $\phi(\infty) = 0$. Isto é devido ao fato de que $exp\{-\phi^{-1}(F(x))\}$ será uma f.d.a. quando F for uma f.d.a. univariada.

Observa-se que a transformada ϕ é contínua, estritamente decrescente e satisfaz $\phi(0) = 1$. Então, o funcional inverso ϕ^{-1} é estritamente decrescente e satisfaz $\phi^{-1}(0) = \infty$ e $\phi^{-1}(1) = 0$.

Além disso, ϕ possui derivadas contínuas de todas as ordens e derivadas com sinais alternados, isto é, $(-1)^i\phi^{(i)}(s) \geq 0$ para todo $s \geq 0$ onde $\phi^{(i)}$ representa a i-ésima derivada. Esta propriedade de alternância de sinais na derivada é denominada completamente monótona.

Seja $L_m = \{\phi : [0,\infty) \to [0,1] | \phi(0) = 1, \phi(\infty) = 0, (-1)^j\phi^{(j)} \geq 0, j = 1, ..., m\}$, $m = 1, 2, ...$, a classe das funções diferenciáveis, estritamente decrescentes. Considerando a classe $L_n^* = \{w : [0,\infty) \to [0,\infty) | w(0) = 0, w(\infty) = \infty, (-1)^{j-1}w^{(j)} \geq 0, j = 1, ..., n\}$, $n = 1, 2, ...$, tem-se que as funções em L_n^* são, usualmente, composições da forma $\psi^{-1} \circ \phi$ com $\psi, \phi \in L_1$. As composições do tipo $\psi^{-1} \circ \phi \in L_\infty^*$ aparecerão na construção de cópulas trivariadas.

A seguir serão apresentados alguns teoremas sobre a classe L_∞^*. Estes são úteis pois nos fornecem a imagem das funções que são o domínio das Transformadas de Lapace. Isto nos possibilita verificar se o domínio das expresões utilizadas na construção de cópulas é $[0,\infty)$.

Teorema 3.5. *Seja* χ *uma Transformada de Laplace. Então* χ^α *é completamente monótona para todo* $\alpha > 0$ *se e somente se* $-Ln(\chi) \in L_\infty^*$.

Demonstração: Para a demonstração consulte [**?**].

Teorema 3.6. *Se* ψ *é uma Transformada de Laplace tal que* $-Ln(\psi) \in L_\infty^*$ *e* ϕ *é outra Transformada de Laplace, então* $\eta(s) = \phi(-Ln(\psi(s)))$ *é uma Transformada de Laplace.*

Demonstração: Para a demonstração consulte [**?**].

Agora, seja F uma dada f.d.a. univariada. Então, existe uma única f.d.a. G tal que

$$F(x) = \int_0^\infty G^\alpha(x) dM(\alpha) \tag{3.15}$$

Comparando (3.14) e (3.15) podemos escrever

$$F(x) = \int_0^\infty G^\alpha(x) dM(\alpha) = \phi(-Ln(G(x))) \tag{3.16}$$

Da expressão (3.16) obtemos $G(x) = e^{-\phi^{-1}(F)}$.

Considerando a classe bivariada $\Im(F_1, F_2)$, seja $G_j = e^{-\phi^{-1}(F_j)}$, $j = 1, 2$. Então, a expressão a seguir é uma f.d.a. em $\Im(F_1, F_2)$:

$$\int_0^\infty G_1^\alpha G_2^\alpha dM(\alpha) = \phi(-Ln(G_1) - Ln(G_2)) = \phi(\phi^{-1}(F_1) + \phi^{-1}(F_2)) \tag{3.17}$$

A cópula obtida considerando que F_1 e F_2 assumem valores da distribuição $U(0, 1)$ é

$$C(u_1, u_2) = \phi(\phi^{-1}(u_1) + \phi^{-1}(u_2)) \tag{3.18}$$

A forma (3.18) é denominada Cópula Arquimediana. e possui a seguinte propriedade:

Teorema 3.7. *A cópula (3.18) possui densidade TP_2.*

Demonstração: Para a demonstração consulte [**?**].

O teorema anterior fornece a dependência da cópula arquimediana. Como ela é TP_2, que é uma dependência forte, então, ela, também, possui dependência SI, LTD, RTI, PQD. Porém, ela poderá ser adequada na modeladem de variáveis aleatórias TP_2 e não ser adequada na modelagem de variáveis aleatórias que possuem algum tipo de dependência mais fraco, por exemplo, v.a. SI.

Para m f.d.a. univariadas, $F_1, ..., F_m$, uma extensão simples é dada pela f.d.a. multivariada $F = \phi(\sum_{j=1}^m \phi^{-1}(F_j))$ cuja Cópula Arquimediana é,

$$C(\mathbf{u}) = \phi(\sum_{j=1}^m \phi^{-1}(u_j)) \tag{3.19}$$

Para se conseguir mais tipos de dependência, considera-se

$$C(\mathbf{u}) = \int_0^\infty K[exp(-\alpha\phi^{-1}(u_1)), ..., exp(-\alpha\phi^{-1}(u_m))]dM(\alpha) \tag{3.20}$$

$$C(\mathbf{u}) = \int_0^\infty ... \int_0^\infty K(G_1^{\alpha_1}, ..., G_m^{\alpha_m})M_m(d\alpha_1, ..., d\alpha_m)$$

onde K é uma cópula multivariada, $G_j = e^{-\phi_j^{-1}(u_j)}$ e M_m é uma distribuição multivariada tal que a j-ésima função de distribuição marginal univariada possui Transformada de Laplace ϕ_j, $j = 1, ..., m$. A cópula (3.20) possui a seguinte propriedade:

Teorema 3.8. *A cópula (3.20) será associada se K e M_m forem f.d.a. de v.a. associadas.*

Demonstração: Para a demonstração consulte [**?**].

Uma generalização trivariada de (3.18), no sentido de obter-se mais tipos de dependência pois a expressão seguinte envolverá duas Transformadas de Laplace diferentes e, consequentemente, envolverá mais que um parâmetro, ao contrário da expressão (3.19) que envolve apenas uma Transformada de Laplace, é dada por

$$C(\mathbf{u}) = \psi(\psi^{-1} \circ \phi[\phi^{-1}(u_1) + \phi^{-1}(u_2)] + \psi^{-1}(u_3)) \tag{3.21}$$

onde ψ, ϕ são Transformadas de Laplace e $\nu = \psi^{-1} \circ \phi \in L_\infty^*$.

Note que (3.21) possui funções de distribuição marginais bivariadas $(1, 2)$ da forma (3.18) com Transformada de Laplace ϕ e funções de distribuição marginais bivariadas $(1, 3)$ e $(2, 3)$ da forma (3.18) com Transformada de Laplace ψ. Observe que (3.19) é um caso especial de (3.21) quando $\psi = \phi$. A representação em forma de mistura de funções de distribuição de (3.21), que generaliza (3.17), é

$$C(\mathbf{u}) = \int_0^\infty \int_0^\infty G_1^\beta(u_1)G_2^\beta(u_2)dM_2(\beta, \alpha)G_3^\alpha(u_3)dM_1(\alpha) \tag{3.22}$$

onde $G_1 = G_2 = e^{-\phi^{-1}}$ e $G_3 = e^{-\psi^{-1}}$, M_1 é a distribuição correspondente à ψ, $M_2(.;\alpha)$ é a distribuição com Transformada de Laplace χ_α definida por $\chi_\alpha^{-1}(z) = \nu^{-1}(-\alpha^{-1}Ln(z))$.

A expressão (3.22) segue da seguinte representação:

$$\int_0^\infty G_{12}^\alpha(u_1, u_2) G_3^\alpha(u_3) dM_1(\alpha)$$

onde M_1 e G_3 foram definidas acima e $G_{12}(u_1, u_2) = exp(-\nu[\phi^{-1}(u_1) + \phi^{-1}(u_2)])$.

As seguintes famílias uniparamétrica de Transformada de Laplace (TL) podem ser utilizadas para a construção de cópulas com o uso das expressões apresentadas anteriormente:

- LTA: $\phi_\theta(s) = exp(-s^{1/\theta})$, $\theta \geq 1$;
- LTB (gamma): $\phi_\theta(s) = (1+s)^{-1/\theta}$, $\theta \geq 0$;
- LTC (série de potência): $\phi_\theta(s) = 1 - (1 - e^{-s})^{1/\theta}$, $\theta > 0$;
- LTD (série logarítmica): $\phi_\theta(s) = -\theta^{-1}Ln(1 - (1 - e^{-\theta})e^{-s})$, $\theta > 0$.

As correspondentes transformadas inversas são dadas por:

- LTA: $\phi_\theta^{-1}(t) = (-Ln(t))^\theta$;
- LTB: $\phi_\theta^{-1}(t) = t^{-\theta} - 1$;
- LTC: $\phi_\theta^{-1}(t) = -Ln(1 - (1-t)^\theta)$;
- LTD: $\phi_\theta^{-1}(t) = -Ln\left(\frac{1-e^{-\theta t}}{1-e^{-\theta}}\right)$.

onde $t = \phi(s)$.

A família LTD é obtida através de v.a. cuja função de probabilidade possui a expressão $\frac{(1-e^{-\theta})^i}{i\theta}$, $i = 1, 2, ...$Similarmente, a família LTC é obtida através de v.a. cuja função de probabilidade é θ^{-1} para $i = 1$ e $\theta^{-1}\prod_{j=1}^{i-1}(j - \theta^{-1})$ para $i = 2, 3, ...$.

Considerando-se as famílias LTA, LTB, LTC e LTC, pode-se mostrar que $-Ln(\phi_\theta) \in L^*_\infty$. Para as famílias LTA e LTB, a demonstração é direta. E, para as famílias LTC e LTD, demonstra-se utilizando o Teorema 3.5. Para as demonstrações destes fatos consulte [**?**].

Joe [**?**] apresenta outras famílias de Transformadas de Laplace.

Utilizando as famílias de Transformada de Laplace supra-citadas e a equação (3.18), obtém-se algums modelos de cópulas bivariadas.

- Família B_1: **Modelo de Frank [?]**

  Este modelo utiliza a família LTD para ϕ. A cópula associada à este modelo é dada por

$$C(u, v; \delta) = -\delta^{-1} Ln\left(\frac{\eta - (1 - e^{-\delta u})(1 - e^{-\delta v})}{\eta}\right)$$

  onde $\eta = 1 - e^{-\delta}$, $0 \leq \delta < \infty$.

  A função densidade associada à este modelo de cópula é,

$$c(u, v; \delta) = \frac{\delta \eta e^{-\delta(u+v)}}{(\eta - (1 - e^{-\delta u})(1 - e^{-\delta v}))^2}$$

  Esta família possui as seguintes propriedades: SI, densidade TP_2, reflexão simétrica [4] [**?**].

  Este modelo engloba o caso das variáveis aleatórias serem independentes quando $\delta \to 0$.

- Família B_2: **Modelo de Kimeldorf e Sampson [?]**

  Este modelo utiliza a família LTB para ϕ. A cópula associada a este modelo é dada por

[4] $c(u, v|\delta = c(1 - u, 1 - v|\delta), 0 < u, v < 1$

$$C(u,v;\delta) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}$$

para $0 \leq \delta < \infty$.

A sua função densidade é dada por:

$$c(u,v;\delta) = (1+\delta)(uv)^{-\delta-1}(u^{-\delta} + v^{-\delta} - 1)^{-2-1/\delta}$$

Esta família possui as seguintes propriedades: SI, densidade TP_2 [**?**].

Este modelo engloba o caso das v.a. serem independentes quando $\delta \to 0$.

- Família B_3: **Modelo de Joe [?]**

Este modelo utiliza a família LTC para ϕ. A cópula associada a este modelo é dada por

$$C(u,v;\delta) = 1 - (\overline{u}^{\delta} + \overline{v}^{\delta} - \overline{u}^{\delta}\overline{v}^{\delta})^{1/\delta}$$

para $1 \leq \delta < \infty$ onde $\overline{u} = 1 - u$ e $\overline{v} = 1 - v$.

A sua função densidade é,

$$c(u,v;\delta) = \overline{u}^{\delta-1}\overline{v}^{\delta-1}[\delta - 1 + \overline{u}^{-\delta} + \overline{v}^{-\delta} - \overline{u}^{\delta}\overline{v}^{\delta}](\overline{u}^{\delta} + \overline{v}^{\delta} - \overline{u}^{\delta}\overline{v}^{\delta})^{-2+1/\delta}$$

Esta família possui as seguintes propriedades: SI, densidade TP_2 [**?**].

Este modelo engloba o caso das v.a. serem independentes quando $\delta = 1$.

Figura 3.2: Gráficos da família B_2 correspondentes aos parâmetros $\delta = 0,5$ e $\delta = 2$

Figura 3.3: Gráficos da família B_4 correspondentes aos parâmetros $\delta = 2$, $\delta = 5$ e $\delta = 9$

- Família B_4: **Modelo de Gumbel [?]**

  Este modelo utiliza a família LTA para ϕ. A cópula associada a este modelo é dada por

$$C(u, v; \delta) = exp(-(\widetilde{u}^{\delta} + \widetilde{v}^{\delta})^{1/\delta})$$

  para $1 \leq \delta < \infty$ onde $\widetilde{u} = -Ln(u)$ e $\widetilde{v} = -Ln(v)$.

  A função densidade associada à este modelo de cópula é dada por:

$$c(u, v; \delta) = C(u, v; \delta)(uv)^{-1} \frac{(\widetilde{u}\widetilde{v})^{\delta-1}}{(\widetilde{u}^{\delta} + \widetilde{v}^{\delta})^{2-1/\delta}} [(\widetilde{u}^{\delta} + \widetilde{v}^{\delta})^{1/\delta} + \delta - 1]$$

  Esta família possui as seguintes propriedades: SI, densidade TP_2 [**?**].

  Este modelo engloba o caso das v.a. serem independentes quando $\delta = 1$.

As figuras 3.2 e 3.3 mostram, respectivamente, alguns gráficos do Modelo de Kimeldorf e Sampson e do Modelo de Gumbel.

Joe [**?**] apresenta outros modelos de cópulas bivariadas além de extendê-los para variáveis aleatórias com dependência negativa.

Neste capítulo, serão apresentadas as versões trivariadas das famílias B_2 e B_4 utilizando-se as famílias de Transformada de Laplace LTA e LTB e a equação (3.21). Inicialmente, estes modelos foram escolhidos por apresentarem expressões, relativamente, mais fáceis de serem manipuladas. Em trabalhos posteriores, outros modelos podem ser utilizados.

Joe [**?**], também, apresenta outros modelos de cópula trivariada além de extendê-los para variáveis aleatórias com dependência negativa.

- Família M_1: **Generalização da família** B_2

  Este modelo utiliza a família LTB para ψ_{θ_1} e ϕ_{θ_2}. A cópula associada a este modelo é dada por

$$C(u_1, u_2, u_3; \theta_1, \theta_2) = \left[\left(u_1^{-\theta_2} + u_2^{-\theta_2} - 1 \right)^{\theta_1/\theta_2} + u_3^{-\theta_1} - 1 \right]^{-1/\theta_1}$$

  para $\theta_1 \leq \theta_2$, $\theta_1 \geq 0$ e $\theta_2 \geq 0$.

- Família M_2: **Generalização da família** B_4

  Este modelo utiliza a família LTA para ψ_{θ_1} e ϕ_{θ_2}. A cópula associada a este modelo é dada por

$$C(u_1, u_2, u_3; \theta_1, \theta_2) = exp \left\{ - \left(\left([-Ln(u_1)]^{\theta_2} + [-Ln(u_2)]^{\theta_2} \right)^{\theta_1/\theta_2} + [-Ln(u_3)]^{\theta_1} \right)^{1/\theta_1} \right\}$$

  para $\theta_1 < \theta_2$, $\theta_1 \geq 1$ e $\theta_2 \geq 1$.

Como os modelos de cópula M_1 e M_2 dependem de u_1, u_2 e u_3, o gráfico da função densidade de $C(u_1, u_2, u_3; \theta_1, \theta_2)$ é quadridimensional. Por isso, a seguir, serão apresentadas as projeções do gráfico da função densidade de $C(u_1, u_2, u_3; \theta_1, \theta_2)$, dos dois modelos citados, para diferentes valores dos parâmetros θ_1 e θ_2 e de u_1, u_2 e u_3.

Figura 3.4: Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1

Figura 3.5: Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1

Figura 3.6: Gráficos correspondentes aos parâmetros $\theta_1 = 0,1$ e $\theta_2 = 0,3$ no modelo M_1

Figura 3.7: Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1

Figura 3.8: Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1

Figura 3.9: Gráficos correspondentes aos parâmetros $\theta_1 = 0,5$ e $\theta_2 = 0,6$ no modelo M_1

Figura 3.10: Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1

Figura 3.11: Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1

Figura 3.12: Gráficos correspondentes aos parâmetros $\theta_1 = 0,8$ e $\theta_2 = 0,9$ no modelo M_1

Figura 3.13: Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2

Figura 3.14: Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2

Figura 3.15: Gráficos correspondentes aos parâmetros $\theta_1 = 1$ e $\theta_2 = 3$ no modelo M_2

Figura 3.16: Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2

Figura 3.17: Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2

Figura 3.18: Gráficos correspondentes aos parâmetros $\theta_1 = 5$ e $\theta_2 = 6$ no modelo M_2

Figura 3.19: Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2

Figura 3.20: Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2

Figura 3.21: Gráficos correspondentes aos parâmetros $\theta_1 = 8$ e $\theta_2 = 9$ no modelo M_2

3.4 Implementação de um Modelo de Cópula

Considere as variáveis aleatórias $X_1, ..., X_n$. Dada as realizações $x_{j,1}, ..., x_{j,m}$ da variável X_j, $j = 1, ..., n$ e tomando-se blocos de tamanho k desta amostra tem-se:

$$\begin{aligned} M_{j1} &= max\{x_{j,1}, ..., x_{j,k}\} \\ M_{j2} &= max\{x_{j,k+1}, ..., x_{j,2k}\} \\ &\vdots \\ M_{jb} &= max\{x_{j,bk-k+1}, ..., x_{j,bk}\} \end{aligned}$$

Por exemplo, considerando a Distribuição GEV dada por

$$F_{M_j}(x) = \begin{cases} exp\left(-(1-\alpha(\frac{x-\mu}{\sigma}))^{\frac{1}{\alpha}}\right) & \alpha \neq 0 \\ \\ exp(-exp(-\frac{x-\mu}{\sigma})) & \alpha = 0 \end{cases}$$

pode-se, através dessa nova amostra, $M_{j,1}, ..., M_{j,b}$, obter os estimadores de $\hat{\alpha_j}$, $\hat{\mu_j}$ e $\hat{\sigma_j}$ e, assim, determinar a função de distribuição marginal F_{M_j}, j=1, ..., n.

Depois de obter-se os dados $\{(M_{1,1}, ..., M_{n,1}), ..., (M_{1,b}, ..., M_{n,b})\}$, pode-se ajustar um modelo de cópula aos dados a partir das funções de distribuição marginais, construindo-se $\{(u_{1,1}, ..., u_{1,b}), ..., (u_{n,1}, ..., u_{n,b})\}$ sabendo-se que $U_{ji} = F_{X_j}(X_{ji})$, $j = 1, ..., n$ e $i = 1, ..., b$.

Capítulo 4

Aplicação

4.1 Introdução

Em linguística, o estudo de correlação acústica de classes rítmicas é um problema pesquisado.

Os idiomas diferem em seus ritmos da fala [**?**]. Originalmente, três classes rítmicas foram propostas: "syllable-timed rhythm", "stress-timed languages" e "mora-timing". Mas nenhuma correlação acústica foi encontrada para estas classes [**?**].

Em 1999, Ramus, Nespor e Mehler [**?**], obtiveram evidências de que estatísticas baseadas na segmentação manual de vogais e consoantes de um sinal da fala poderia ser utilizado para a diferenciação de classes rítmicas. Porém, o problema desta metodologia é o trabalho manual feito por um foneticista e, por causa disto, eles usam conjuntos de dados pequenos e os resultados não serão estatisticamente significantes. Com isso, Galves et al. [**?**] obtiverão resultados similares à Ramus com o uso de uma metodogia completamente automatizada sem a necessidade de trabalhos manuais.

energia em duas regiões diferentes do espectograma do sinal.

Considera-se um sinal acústico da fala como uma observação, ou seja, uma pessoa pronuncia uma frase e esta é codificada num sinal, conforme ilustra a Figura 4.1, relativo à pronúncia da frase "I genitori lasciano Marco senza risorse" (língua italiana). Tal codificação é permitida através de um microfone pois este capta as diferenças de pressão durante a pronúncia da frase e as converte em sinais elétricos.

O sinal resultante da pronúncia possui várias frequências. Então, como a Transformada de

Fourier filtra frequências, tal transformada é aplicada ao sinal a fim de obter-se o espectograma o qual mostra o comportamento do sinal nas diversas frequências. A Figura 4.1 ilustra o espectograma do sinal resultante da pronúncia da frase, em italiano, "I genitori lasciano Marco senza risorse".

O espectograma no tempo t e na frequência f, pode ser interpretado como a energia que a frequência f adiciona à energia total num sinal acústico no tempo t. A Figura 4.2 ilustra um espectograma onde as regiões escuras correspondem às altas energias e as seguintes regiões de frequências: 80 Hz à 800 Hz (baixas frequências), 800 Hz à 1500 Hz (médias frequências) e 1500 Hz à 3000 Hz (altas frequências). Vale observar que há frequências acima de 3000 Hz que não serão consideradas neste estudo.

Dado o sinal $x(t)$, o seu espectro de Fourier e a sua energia estão definidos, respectivamente, por

$$X(f) = \int_t x(t)e^{-i2\pi ft}dt$$

e

$$E_x = \int_t |x(t)|^2 dt$$

onde t e f representam, respectivamente, o tempo e a frequência.

O teorema de Parseval fornece que

$$\int_t |x(t)|^2 dt = \int_f |X(f)|^2 df$$

Assim, $|X(f)|^2$ será denominada densidade de energia espectral de $x(t)$. O teorema de Parseval diz que a densidade de energia espectral para cada frequência pode ser interpretada como a contribuição desta frequência à energia total do sinal.

Como o interesse reside na densidade espectral local, será utilizada a Transformada de Fourier local. Definindo uma janela $w(t)$ como uma função real que vai para zero rapidamente à medida que nos distanciamos do zero, o sinal no retorno de t_o é definido como $x_w(t_o, t) = x(t)w(t - t_o)$. Então, o espectro local será definido como

$$X_w(t_o, f) = \int_t x(t)w(t - t_o)e^{-i2\pi ft}dt$$

Define-se o espectograma como a densidade de energia para o espectro local de $x(t)$, $S_x(t, f) = |X_w(t, f)|^2$.

Figura 4.1: Acima: Onda acústica resultante da pronúncia da frase, na língua italiana, "I genitori lasciano Marco senza risorse". Embaixo: Espectograma correspondente à onda acima indicando um tempo t e sua frequência f associada sendo que este espectograma foi obtido considerando-se uma discretização temporal de 2 ms

Figura 4.2: Acima: Indicação de regiões de alta e baixa enegia no espectograma. Embaixo: As regiões de frequência 80 Hz à 800 Hz, 800 Hz à 1500 Hz e 1500 Hz à 3000 Hz são mostradas no espectograma

Neste estudo, o tempo e a frequência foram discretizandos, respectivamente, a cada 25 ms e 20 Hz. No sinal acústico, a cada 25 ms e 20 Hz, utilizou-se uma janela gaussiana para obter-se os coeficientes de Fourier para as regiões de baixa, média e alta frequência.

Em cada intervalo de tempo de 25 ms, considerou-se o seu ponto médio. Para tal ponto, através dos coeficientes de Fourier para cada região de frequência, as energias associadas à cada região foram obtidas pelas expressões

$$
\begin{aligned}
E_1^{i,j} &= \frac{1}{T} \sum_{f=80,100,\ldots,800} S^{i,j}(t,f) \\
E_2^{i,j} &= \frac{1}{T} \sum_{f=820,840,\ldots,1500} S^{i,j}(t,f) \\
E_3^{i,j} &= \frac{1}{T} \sum_{f=1520,1540,\ldots,3000} S^{i,j}(t,f)
\end{aligned}
$$

sendo, neste caso, $x = (i,j)$ onde as energias $E_1^{i,j}$, $E_2^{i,j}$ e $E_3^{i,j}$ representam, respectivamente, as energias nas regiões de baixa, média e alta frequência sendo que os índices i e j correspondem, respectivamente, ao falante i e a frase j.

4.2 Metodologia

O objetivo desta seção será modelar as energias extremas nas regiões de alta, média e baixa frequência utilizando-se a Distribuição GEV e, posteriormente, os modelos de cópula tridimensional M_1 e M_2 serão ajustados a fim de obter-se a função densidade conjunta das energias extremas uniformizadas.

Os dados utilizados consistem em 153 sentenças da língua inglesa e 216 sentenças da língua francesa faladas por 4 falantes do sexo feminino. Para a língua inglesa, uma pessoa pronunciou 33 frases e as demais pronunciaram 40 frases e, para a língua francesa, as quatro pessoas pronunciaram, respectivamente, 50, 58, 54 e 54 frases.

Para cada frase, o espectograma e as energias correspondentes à cada região da frequência foram obtidos. As energias obtidas de uma dada frase, para cada falante, consistem na amostra, isto é, para uma dada frase, obtém-se energias correspondentes à baixa, média e alta frequência sendo que tais energias, separadas de acordo com as regiões de frequência, formam uma amostra.

Para cada amostra, o número de energias obtidas é diferente. Mas, para a estimação da Distribuição GEV, deve haver o mesmo número de obervações em cada amostra. Então, foram retiradas as m últimas observações de cada amostra onde m correponde ao mínimo dos comprimentos das amostras.

Como, para a língua inglesa, o mínimo comprimento obtido foi 47, os blocos seriam formados por 47 observações e, assim, muita informação seria perdida das amostras com mais que 100 observações. Então, foi decidido pela retirada da frase composta por 47 energias e, com isso, 152 amostras foram consideradas. Para a língua francesa, não retirou-se frases.

Ao invés de considerar as energia de cada frase, considerou-se o log-retorno das energias, isto é, $Ln(E^{i,j}_{l+1,k}/E^{i,j}_{l,k})$, $l = 1, ..., n_l$ onde l representa a l-ésima observação da frase j do falante i, n_l é o número de energias da frase i e k, $k = 1, 2, 3$, representa a região de frequência. Com a aplicação do log-retorno, perde-se uma observação em cada amostra.

Como há amostras relativas aos falantes 1, 2, 3 e 4 das línguas inglesa e francesa, modelos autoregressivos [**?**] foram utilizados para que fosse retirada a influência da pronúncia de cada pessoa tornando as observações, relativas à uma mesma frase, independentes, isto é, o resíduo de um modelo autoregressivo será modelado.

Inicialmente, o critério utilizado para a obtenção da ordem do modelo autoregressivo foi a

minimização do AIC (Critério de Informação de Akaike) [**?**, **?**]. Porém, para algumas frases, a ordem obtida foi alta como, por exemplo, 14, 16.

Considerando-se algumas ordens para o modelo autoregressivo, foram obtidas as estimativas dos parâmetros da Distribuição GEV e dos modelos de cópula M_1 e M_2. Por exemplo, para os log-retornos correspondentes à região de baixa frequência de frases da língua inglesa, considerando-se o Método de Máxima Verossimilhança e o critério da minimização do AIC, obteve-se as estimativas 4,9326; 0,9777 e 0,1675 relativas, respectivamente, aos parâmetros μ, σ e α da Distribuição GEV e, também, obteve-se as estimativas $\hat{\theta}_1 = 0,1756$; $\hat{\theta}_2 = 0,1756$ e $\hat{\theta}_1 = 1,0914$; $\hat{\theta}_2 = 1,1523$ relativas, respectivamente, aos modelos de cópula M_1 e M_2. Já, para o modelo autoregressivo de ordem 7, obteve-se as estimativas 5,1729; 1,017 e 0,2043 relativas, respectivamente, aos parâmetros μ, σ e α da Distribuição GEV e, também, obteve-se as estimativas $\hat{\theta}_1 = 0,2038$; $\hat{\theta}_2 = 0,2038$ e $\hat{\theta}_1 = 1,0835$; $\hat{\theta}_2 = 1,2119$ relativas, respectivamente, aos modelos de cópula M_1 e M_2.

Analisando os gráficos de autocorrelação, os valores dos AIC dos modelos ajustados considerando-se várias ordens e o Teste de Box-Pierce [**?**] para as três regiões de frequência, optou-se por um modelo autoregressivo de ordem 7 para ambas as línguas.

Vale observar que ao invés de um modelo autoregressivo, um modelo ARMA foi considerado. Porém, neste caso, o ajuste utilizando-se o "software" R não foi possível para os dados das línguas inglesa e francesa.

Os resíduos obtidos para uma dada frase foram considerados como sendo um bloco a fim de se calcular o seu máximo para que os parâmetros da Distribuição GEV fossem estimados pelos Métodos dos L-momentos e Máxima Verossimilhança para as três regiões de frequência. Como necessita-se de blocos de mesmo tamanho para o cálculo dos máximos, foram considerados blocos de tamanho igual ao mínimo do tamanho de todas as amostras.

Com os blocos de tamanhos iguais, obteve-se o máximo de cada bloco para que os parâmetros μ, σ e α da Distribuição GEV pudessem ser estimados. O Método dos L-momentos foi utilizado para obter estimativas iniciais dos parâmetros μ, σ e α já que ele apresentou-se menos sensível às condições iniciais e à não convergência. Considerando-se as estimativas obtidas pelo método citado como estimativas iniciais, foi utilizado o Método de Máxima Verossimilhança sendo que estes métodos foram utilizados simultaneamente para verificar a concordância das estimativas obtidas. As estimativas dos parâmetros obtidas pelos métodos citados são mostradas nas Tabelas 2.3 e 2.4

correspondentes, respectivamente, às linguas inglesa e francesa.

O Teste de Hosking [?] foi aplicado utilizando-se as estimativas das três regiões de frequência, obtidas por L-momentos. Para as regiões de baixa, média e alta frequência da língua inglesa, o valor da estatística do teste foi, respectivamente, 3,03; 3,05 e 3,35 evidenciando que os dados podem ser ajustados à Distribuição GEV com $\alpha > 0$ (Distribuição Weibull). E, para as regiões de baixa, média e alta frequência da língua francesa, o valor da estatística do teste foi, respectivamente, 3,17; 2,01 e 5,36 evidenciando que os dados podem ser ajustados à Distribuição GEV com $\alpha > 0$ (Distribuição Weibull).

Utilizando-se os máximos uniformizados obtidos considerando-se os parâmetros da Distribuição GEV estimados pelos Métodos dos L-momentos e Máxima Verossimilhança, foram aplicados o Teste de Kolmogorov-Smirnov [?] e o Teste de Wilcoxon [?] para verificar o ajuste da Distribuição GEV com a sua distribuição empírica nas três regiões de frequência e nas línguas inglesa e francesa. Pelas Tabelas 2.1 e 2.2 verifica-se que o ajuste é melhor considerando-se o Método dos L-momentos para a obtenção das estimativas dos parâmetros da Distribuição GEV referentes às línguas inglesa e francesa. Assim, consideraremos as estimativas obtidas pelo Método dos L-momentos para ambas as línguas.

Também, utilizando-se os máximos uniformizados nas três regiões de frequência, foram obtidos os gráficos de dispersão baixa versus média frequência, média versus alta frequência e baixa versus alta frequência, mostrados nas Figuras 4.3 e 4.4, para que fosse verificado o tipo de dependência dos dados. Observou-se que os dados referentes às regiões de baixa, média e alta frequência não possuem dependência negativa para ambas as línguas. Pelas Figuras 4.3 e 4.4, observa-se que os máximos uniformizados sugerem alguma dependência entre os máximos uniformizados.

Por fim, por meio da uniformização dos máximos, foram estimados os parâmetros das cópulas trivariadas citadas no capítulo anterior (Família M_1 e M_2) através do Método de Máxima Verossimilhança para as línguas inglesa e francesa. Neste caso, a estimação dos parâmetros das cópulas foi obtida através do Método de Máxima Verossimilhança por ser de fácil implementação já que, para aplicar-se o Método dos L-momentos, as expressões dos L-momentos devem ser obtidas sendo que os modelos de cópula utilizados são trivariados e isto pode tornar o cálculo dos L-momentos laborioso e sendo necessário a inversão da f.d.a. dos modelos utilizados de cópula.

Para os modelos M_1 e M_2, as estimativas dos parâmetros são mostradas nas Tabelas 2.5 e 2.6,

Figura 4.3: Gráficos de Dispersão dos máximos uniformizados referentes às três regiões de frequência da língua inglesa

correspondentes, respectivamente, às línguas inglesa e francesa.

Tabela 2.1: Testes de Kolmogorov-Smirnov e Wilcoxon para verificar o ajuste da Distribuição GEV aos dados relativos à língua inglesa

Região da Frequência	Método de Estimação	Teste de Kolmogorov-Smirnov	Teste de Wilcoxon
Baixa	MV	0,984	0,889
	LM	0,997	0,955
Média	MV	0,977	0,963
	LM	0,999	0,920
Alta	MV	0,997	0,925
	LM	0,997	0,922

Tabela 2.2: Testes de Kolmogorov-smirnov e Wilcoxon para verificar o ajuste da Distribuição GEV aos dados relativos à língua francesa

Região da Frequência	Método de Estimação	Teste de Kolmogorov-Smirnov	Teste de Wilcoxon
Baixa	MV	0,992	0,858
	LM	0,999	0,955
Média	MV	0,974	0,901
	LM	0,992	0,940
Alta	MV	0,998	0,929
	LM	0,998	0,929

4.3 Comparação entre os Modelos

Para que o ajuste dos modelos M_1 e M_2 fossem comparados, considerou-se a Classe de Fréchet, ou seja, a classe formada pelas funções de distribuição acumulada marginais de uma dada função de distribuição multivariada. Por exemplo, considerando uma função de distribuição trivariada $F_{\mathbf{X}}(\mathbf{x})$ onde $\mathbf{X} = (X_1, X_2, X_3)$, uma das Classe de Fréchet é formada pelas funções de distribuição acumulada marginais $(F_{X_1}, F_{X_2}, F_{X_3}, F_{X_1,X_2}, F_{X_1,X_3}, F_{X_2,X_3}, F_{X_1,X_2,X_3})$.

Se os modelos ajustados forem parecidos, espera-se que as componentes que formam a Classe

Figura 4.4: Gráficos de Dispersão dos máximos uniformizados referentes às três regiões de frequência da língua francesa

de Fréchet, de ambos os modelos, sejam compátiveis.

A partir da cópula $C(u_1, u_2, u_3)$, as cópulas marginais $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ foram obtidas tomando-se, respectivamente, $u_3 = 1$, $u_2 = 1$ e $u_1 = 1$ já que as cópulas são avaliadas em dados uniformizados.

Comparou-se, através do Teste de Wilcoxon, a cópula $C(u_1, u_2, u_3)$ e as cópulas marginais $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$, para ambos os modelos M_1 e M_2 ajustados sendo que as estimativas dos parâmetros da Distribuição GEV foram estimados pelo Método dos L-momentos, com as suas respectivas distribuições empíricas. Também, utilizando-se tais modelos, considerou-se a maior distância, entre as cópulas $C(u_1, u_2, u_3)$, $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ e as suas correspondentes distribuições empíricas. Os resultados correspondentes às línguas inglesa e francesa são apresentados, respectivamente, nas Tabelas 3.7, 3.10, 3.8 e 3.11.

O procedimento supra-citado foi repetido comparando-se as cópulas $C(u_1, u_2, u_3)$, $C(u_1, u_2)$, $C(u_1, u_3)$ e $C(u_2, u_3)$ resultantes do modelo M_1 com as cópulas resultantes do modelo M_2. Os resultados correspondentes às línguas inglesa e francesa são apresentados, respectivamente, nas Tabelas 3.9 e 3.12.

Pelas Tabelas 3.7 e 3.10, observa-se que o modelo de cópula M_1 se ajusta melhor tanto para a língua inglesa como para a francesa utilizando-se como critério o Teste de Wilcoxon.

Pelas Tabelas 3.8 e 3.11, utilizando-se o critério da distância máxima entre o modelo de cópula e a correspondente distribuição empírica, observa-se o modelo M_2 apresentou as menores distâncias considerando-se a língua inglesa enquanto que, para a língua francesa, os resultados não foram conclusivos. No geral, ambos os modelos para ambas as línguas, apresentaram as distâncias máximas pequenas.

Pelas Tabelas 3.9 e 3.12, observa-se que os modelos de cópulas M_1 e M_2 não são discriminados, para ambas as línguas consideradas, podendo-se utilizar qualquer um dos modelos embora o modelo M_1 apresentou p-valores menores no Teste de Wilcoxon para ambos os idiomas conforme Tabelas 3.7 e 3.10. Talvez, utilizando um outro teste, diferenças entre os modelos poderiam ser detectadas.

4.4 Conclusão

Um objetivo ao ajustar-se os modelos de cópulas poderia ser, dado um sinal acústico associado a uma determinada frase, a discriminação desta, por exemplo, determinado se ela pertence à língua

Figura 4.5: Gráfico de θ_2 versus θ_1 considerando-se os modelos de cópula M_1 e M_2 e as línguas inglesa e francesa

inglesa ou francesa. Este problema de reconhecimento sonoro possui grande interesse e aplicabilidade. Por exemplo, uma pessoa que utiliza algum serviço internacional deseja comunicar-se em sua língua. Então, pode-se implementar um sistema em que a pessoa pronuncia alguma frases e o sistema codifica estas frases em sinais acústicos para que o seu idioma seja discriminado a fim de alocar uma atendente que fale sua língua.

O gráfico 4.5 mostra os pontos formados pelas estimativas θ_1 e θ_2 dos modelos M_1 e M_2 aplicados às línguas inglesa e francesa. Observa-se que os pontos formados pelas estimativas das línguas inglesa e francesa estão próximos entre si, apesar de possibilitarem uma discriminação entre os modelos, tanto para o modelo M_1 quanto para o modelo M_2.

O objetivo deste trabalho foi o estudo do ajuste dos modelos de cópula M_1 e M_2 aplicados às línguas inglesa e francesa em relação à modelagem multidimensional de características de estresse do sinal acústico quantificadas pelos máximos das energias observadas nas três regiões de frequência.

O próximo passo seria o estudo de técnicas de agrupamento para que seja criada uma região de confiança em torno do ponto (θ_1, θ_2) correspondente aos modelos de cópula e aos idiomas considerados, de forma que uma dada frase seja discriminada automaticamente e alocada em algum idioma. Para isso, poderia ser estudada a distribuição amostral dos estimadores de (θ_1, θ_2).

Tabela 2.3: Estimativas dos parâmetros da Distribuição GEV correspondentes aos log-retornos extremos das energias nas três regiões de frequência da língua inglesa obtidas através dos Métodos dos L-momentos e Máxima Verossimilhança considerando-se um modelo autoregressivo de ordem 7

Região da Frequência	Estimativas obtidas utilizando-se o Método dos L-momentos	Estimativas obtidas utilizando-se o Método de Máxima Verossimilhança
Baixa	$\hat{\mu}$ = 5,1670 $\hat{\sigma}$=0,9813 $\hat{\alpha}$ = 0,1850	$\hat{\mu}$ = 5,1729 $\hat{\sigma}$=1,017 $\hat{\alpha}$ =0,2043
Média	$\hat{\mu}$ = 6,4811 $\hat{\sigma}$ = 1,5211 $\hat{\alpha}$ = 0,1858	$\hat{\mu}$ =6,4753 $\hat{\sigma}$=1,4758 $\hat{\alpha}$ =0,1688
Alta	$\hat{\mu}$ = 6,1402 $\hat{\sigma}$ = 1,2829 $\hat{\alpha}$ =0,2043	$\hat{\mu}$ =6,1337 $\hat{\sigma}$=1,2833 $\hat{\alpha}$ =0,1936

Tabela 2.4: Estimativas dos parâmetros da Distribuição GEV correspondentes aos log-retornos extremos das energias nas três regiões de frequência da língua francesa obtidas através dos Métodos dos L-momentos e Máxima Verossimilhança considerando-se um modelo autoregressivo de ordem 7

Região da Frequência	Estimativas obtidas utilizando-se o Método dos L-momentos	Estimativas obtidas utilizando-se o Método de Máxima Verossimilhança
Baixa	$\hat{\mu} = 5{,}1658$ $\hat{\sigma}$=0,8286 $\hat{\alpha} = 0{,}1619$	$\hat{\mu} = 5{,}1825$ $\hat{\sigma}$=0,8471 $\hat{\alpha}$ =0,2011
Média	$\hat{\mu} = 6{,}6056$ $\hat{\sigma} = 1{,}2612$ $\hat{\alpha} = 0{,}1027$	$\hat{\mu}$ =6,6218 $\hat{\sigma}$=1,2608 $\hat{\alpha}$ =0,1231
Alta	$\hat{\mu} = 6{,}2067$ $\hat{\sigma} = 1{,}3159$ $\hat{\alpha}$ =0,2738	$\hat{\mu}$ =6,2178 $\hat{\sigma}$=1,2976 $\hat{\alpha}$ =0,2838

Tabela 2.5: Estimativas dos parâmetros dos modelos de cópula M_1 e M_2 obtidas através do Método de Máxima Verossimilhança considerando a uniformização dos máximos da língua inglesa pela aplicação da Distribuição GEV com os parâmetros estimados pelos métodos dos L-momentos e Máxima Verossimilhança

Modelo de Cópula	Método utilizado na estimação dos parâmetros da Distribuição GEV	Estimativas dos parâmetro do do modelo de cópula
M_1	LM	$\hat{\theta}_1$=0,1929 $\hat{\theta}_2$ = 0,1929
M_1	MV	$\hat{\theta}_1$=0,2038 $\hat{\theta}_2$ = 0,2038
M_2	LM	$\hat{\theta}_1$ = 1,0788 $\hat{\theta}_2$ = 1,2094
M_2	MV	$\hat{\theta}_1$ = 1,0835 $\hat{\theta}_2$ = 1,2119

Tabela 2.6: Estimativas dos parâmetros dos modelos de cópula M_1 e M_2 obtidas através do Método de Máxima Verossimilhança considerando a uniformização dos máximos da língua francesa pela aplicação da Distribuição GEV com os parâmetros estimados pelos métodos dos L-momentos e Máxima Verossimilhança

Modelo de Cópula	Método utilizado na estimação dos parâmetros da Distribuição GEV	Estimativas dos parâmetro do do modelo de cópula
M_1	LM	$\hat{\theta}_1$=0,1544 $\hat{\theta}_2 = 0{,}1544$
M_1	MV	$\hat{\theta}_1$=0,1663 $\hat{\theta}_2 = 0{,}1663$
M_2	LM	$\hat{\theta}_1 = 1{,}074$ $\hat{\theta}_2 = 1{,}130$
M_2	MV	$\hat{\theta}_1 = 1{,}0657$ $\hat{\theta}_2 = 1{,}1221$

Tabela 3.7: Teste de Wilcoxon entre os os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua inglesa

Modelo de Cópula	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
M_1	0,261	0,924	0,989	0,232
M_2	0,158	0,795	0,806	0,168

Tabela 3.8: Distância máxima entre os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua inglesa

Modelo de Cópula	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
M_1	0,089	0,060	0,055	0,083
M_2	0,073	0,033	0,047	0,076

Tabela 3.9: Teste de Wilcoxon e distância máxima entre os modelos de cópula M_1 e M_2 considerando-se a língua inglesa

Critério de Decisão	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
Teste de Wilcoxon	0,770	0,699	0,803	0,824
Distância Máxima	0,030	0,029	0,0114	0,0117

Tabela 3.10: Teste de Wilcoxon entre os os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua francesa

Modelo de Cópula	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
M_1	0,500	0,745	0,675	0,160
M_2	0,317	0,634	0,79	0,132

Tabela 3.11: Distância máxima entre os modelos de cópula e as suas respectivas distribuições empíricas considerando-se a língua francesa

Modelo de Cópula	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
M_1	0,038	0,042	0,028	0,062
M_2	0,028	0,029	0,035	0,065

Tabela 3.12: Teste de Wilcoxon e distância máxima entre os modelos de cópula M_1 e M_2 considerando-se a língua francesa

Critério de Decisão	$C(u_1, u_2, u_3)$	$C(u_1, u_2)$	$C(u_1, u_3)$	$C(u_2, u_3)$
Teste de Wilcoxon	0,689	0,854	0,866	0,868
Distância Máxima	0,020	0,018	0,008	0,008

Apêndices

Comparações entre os Métodos de Máxima Verossimilhança e L-momentos utilizando-se os métodos numéricos de Nelder-Mead, Bissecção, Newton e Newton-Raphson

Nesta seção os Métodos de Máxima Verossimilhança e L-momentos serão comparados utilizando-se os métodos numéricos de Newton-Raphson e da Bissecção para a obtenção das raizes de uma equação e os métodos de Nelder-Mead e de Newton para a maximização de uma função.

Para a compararação dos Métodos de Máxima Verossimilhança e L-momentos, gerou-se 10 amostras independentes de tamanho 50 e 500 de uma Distribuição GEV. Para cada amostra independente, obteve-se dois vetores de estimativas dos parâmetros, μ, σ, α, da Distribuição GEV sendo que cada vetor corresponde a um método de estimação. Através dos vetores de estimativas, estatísticas sumárias e o erro quadrático médio correspondente à cada método e parâmetro estimado foram calculados.

Pelas Tabelas , , e observa-se que o Método dos L-momentos utilizando-se os métodos numéricos da Bissecção e Newton-Raphson comportam-se de forma similar. Também, o Método de Máxima Verossimilhança utilizando-se os métodos de Nelder-Mead e Newton comportam-se de forma similar.

U-estatísticas

Seja $\Im$ a classe formada por de todas as f.d.a. e o funcional estatístico $T : \Im \to \mathbb{R}$. Em geral, considera-se uma amostra aleatória de uma distribuição desconhecida F a fim de obter-se alguma

Tabela 4.13: Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

Método Numérico	Método de Estimação	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
Método da Bissecção	LM	$\hat{\mu}$	-0,093	-0,022	0,060	0,100	0,141	0,039	0,081	0,007
		$\hat{\sigma}$	0,732	0,990	1,125	1,182	1,323	1,088	0,177	0,036
		$\hat{\alpha}$	-0,975	-0,713	-0,583	-0,524	-0,411	-0,632	0,174	0,028
Método de Nelder-Mead	MV	$\hat{\mu}$	-0,132	-0,043	0,019	0,085	0,181	0,025	0,095	0,008
		$\hat{\sigma}$	0,844	1,047	1,073	1,109	1,239	1,070	0,099	0,013
		$\hat{\alpha}$	-0,869	-0,835	-0,711	-0,641	-0,484	-0,710	0,134	0,028
Método de Newton-Raphson	LM	$\hat{\mu}$	-0,093	-0,066	-0,010	0,076	0,120	0,005	0,084	0,007
		$\hat{\sigma}$	0,732	0,955	0,988	1,130	1,312	1,017	0,179	0,036
		$\hat{\alpha}$	-0,975	-0,713	-0,514	-0,433	-0,350	-0,589	0,205	0,028
Método de Newton	MV	$\hat{\mu}$	-0,048	-0,025	0,019	0,085	0,181	0,038	0,079	0,007
		$\hat{\sigma}$	1,023	1,068	1,073	1,109	1,239	1,093	0,061	0,011
		$\hat{\alpha}$	-0,869	-0,835	-0,702	-0,639	-0,484	-0,708	0,135	0,028

Tabela 4.14: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = -0,6$ e perturbação $\varepsilon = 0,55$

Método Numérico	Método de Estimação	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
Método da Bissecção	LM	$\hat{\mu}$	-0,108	-0,026	-0,008	0,039	0,099	0,0001	0,059	0,003
		$\hat{\sigma}$	0,823	0,962	0,984	1,017	1,065	0,982	0,066	0,004
		$\hat{\alpha}$	-0,784	-0,689	-0,536	-0,526	-0,443	-0,588	0,116	0,012
Método de Nelder-Mead	MV	$\hat{\mu}$	-0,070	-0,036	-0,023	0,033	0,121	-0,001	0,058	0,003
		$\hat{\sigma}$	0,924	0,959	0,974	1,003	1,056	0,982	0,039	0,001
		$\hat{\alpha}$	-0,672	-0,611	-0,593	-0,579	-0,489	-0,595	0,052	0,002
Método de Newton-Raphson	LM	$\hat{\mu}$	-0,108	-0,026	-0,008	0,039	0,099	0,0001	0,059	0,003
		$\hat{\sigma}$	0,823	0,962	0,984	1,018	1,066	0,982	0,066	0,004
		$\hat{\alpha}$	-0,784	-0,689	-0,536	-0,526	-0,443	-0,588	0,116	0,012
Método de Newton	MV	$\hat{\mu}$	-0,070	-0,036	-0,023	0,033	0,121	-0,001	0,058	0,003
		$\hat{\sigma}$	0,924	0,959	0,974	1,003	1,056	0,982	0,039	0,001
		$\hat{\alpha}$	-0,672	-0,611	-0,593	-0,579	-0,489	-0,595	0,052	0,002

Tabela 4.15: Estatísticas provenientes de amostras de tamanho 50 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = -0,25$

Método Numérico	Método Estimação	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
Método da Bissecção	LM	$\hat{\mu}$	-0,359	-0,187	-0,012	0,038	0,208	-0,051	0,181	0,032
		$\hat{\sigma}$	0,873	0,948	1,016	1,066	1,094	1,000	0,079	0,005
		$\hat{\alpha}$	0,328	0,374	0,430	0,527	0,638	0,459	0,111	0,012
Método de Nelder-Mead	MV	$\hat{\mu}$	-0,338	-0,165	-0,026	0,100	0,199	-0,036	0,177	0,029
		$\hat{\sigma}$	0,859	0,944	1,009	1,035	1,079	0,989	0,072	0,004
		$\hat{\alpha}$	0,430	0,446	0,472	0,493	0,601	0,478	0,049	0,002
Método de Newton-Raphson	LM	$\hat{\mu}$	-0,359	-0,187	-0,012	0,038	0,208	-0,051	0,181	0,032
		$\hat{\sigma}$	0,873	0,948	1,016	1,066	1,094	1,000	0,079	0,005
		$\hat{\alpha}$	0,328	0,374	0,43	0,527	0,638	0,459	0,111	0,012
Método de Newton	MV	$\hat{\mu}$	-0,338	-0,165	0,019	0,108	0,199	-0,024	0,183	0,030
		$\hat{\sigma}$	0,859	0,947	1,024	1,057	1,121	1,003	0,083	0,006
		$\hat{\alpha}$	0,430	0,446	0,472	0,494	0,759	0,505	0,101	0,009

Tabela 4.16: Estatísticas provenientes de amostras de tamanho 500 da Distribuição GEV com parâmetros $\mu = 0$, $\sigma = 1$, $\alpha = 0,5$ e perturbação $\varepsilon = 0,55$

Método Numérico	Método de Estimação	Estimativa	Minímo	Q1	Q2	Q3	Máximo	Média	DP	EQM
Método da Bissecção	LM	$\hat{\mu}$	-0,049	-0,034	-0,005	0,014	0,052	-0,004	0,035	0,001
		$\hat{\sigma}$	0,925	0,989	0,998	1,018	1,052	1,000	0,036	0,001
		$\hat{\alpha}$	0,455	0,472	0,487	0,515	0,554	0,496	0,033	0,001
Método de Nelder-Mead	MV	$\hat{\mu}$	-0,047	-0,033	-0,006	0,031	0,064	0,0008	0,038	0,001
		$\hat{\sigma}$	0,924	0,993	0,997	1,017	1,045	1,000	0,033	0,001
		$\hat{\alpha}$	0,457	0,498	0,507	0,518	0,529	0,505	0,020	0,0004
Método de Newton-Raphson	LM	$\hat{\mu}$	-0,049	-0,0342	-0,005	0,014	0,052	-0,004	0,035	0,001
		$\hat{\sigma}$	0,925	0,989	0,998	1,019	1,053	1,001	0,036	0,001
		$\hat{\alpha}$	0,455	0,472	0,487	0,515	0,554	0,496	0,033	0,001
Método de Newton	MV	$\hat{\mu}$	-0,047	-0,033	-0,005	0,031	0,064	0,0008	0,038	0,001
		$\hat{\sigma}$	0,924	0,993	0,997	1,017	1,045	1,0001	0,033	0,001
		$\hat{\alpha}$	0,457	0,498	0,507	0,518	0,529	0,505	0,020	0,0004

informação a respeito de alguma função envolvendo $F \in \Im$, isto é, $\theta = T(F)$ para um funcional T conhecido como, por exemplo:

- Se $T(F) = F(c)$ para alguma constante c, T é um funcional que mapeia cada F em $P_F(X \leq c)$;
- Se $T(F) = F^{-1}(p)$ para alguma constante p, T mapeia F no seu correspondente p-quantil;
- Se $T(F) = \mathbb{E}(X)$, T mapeia F em sua média.

Seja $(X_1, ..., X_a)$ uma amostra aleatória i.i.d. com f.d.a. F e $\phi(x_1, ..., x_a)$, $a > 1$, uma função real que possui mais que um argumento real. Define-se a seguinte função estatística de grau $a \geq 1$,

$$\theta(F) = T(F) = \mathbb{E}_F(\phi(X_1, ..., X_a)) = \int ... \int \phi(x_1, ..., x_a) dF(x_1).....dF(x_a) \tag{4.1}$$

para $\forall F \in \Im$ onde, neste caso, $\Im = \{F : |\theta(F)| < \infty\}$ e $\phi(x_1, ..., x_a)$ é uma estatística Borel-mensurável.

Como $\mathbb{E}_F(\phi(X_1, ..., X_a)) = \mathbb{E}_F(\phi(X_{\pi_1}, ..., X_{\pi_a}))$ para alguma permutação π que mapeia $\{1, ..., a\}$ em $\{1, ..., a\}$ e há $a!$ permutações, considera-se a função:

$$\phi^*(x_1, ..., x_a) \stackrel{def}{=} \frac{1}{a!} \sum_{\forall \pi} \phi(x_{\pi_1}, ..., x_{\pi_a})$$

Como $\mathbb{E}_F(\phi(X_1, ..., X_a)) = \mathbb{E}_F(\phi^*(X_1, ..., X_a))$ e $\phi^*(X_1, ..., X_a)$ é simétrica em seus argumentos, temos que, na equação (4.1), pode-se assumir, sem perda de generalidade, que ϕ é simétrica em seus argumentos. Com isso, tem-se a seguinte definição:

Definição 4.1. *Para algum inteiro a≥1, seja $\phi : \mathbb{R}^a \to \mathbb{R}$ uma função simétrica em seus a argumentos denominada função núcleo. Sejam $X_1, ..., X_a$ variáveis aleatórias i.i.d. com f.d.a. F e seja, ainda, a esperança, $\mathbb{E}_F(\phi(X_1, ..., X_a))$, de $\phi(X_1, ..., X_a)$. A função $T(F) = \mathbb{E}_F(\phi(X_1, ..., X_a))$ é denominada funcional esperança.*

Suponha que $T(F)$ é um funcional esperança definido de acordo com a equação (4.1), para algum $a \geq 1$. Se temos uma amostra aleatória de tamanho n de F, um meio natural para se estimar $T(F)$ seria por meio do estimador "plug-in" de $T(F)$ dado por $T(\hat{F}_n)$, ou seja, como $\hat{F}_n$ atribui probabilidade $\frac{1}{n}$ para cada X_i, tem-se

$$V_n = T(\hat{F_n}) = \mathbb{E}_{\hat{F_n}}(\phi(X_1, ..., X_a)) = \frac{1}{n^a} \sum_{i_1=1}^{n} ... \sum_{i_a=1}^{n} \phi(X_{i_1}, ..., X_{i_a}) \tag{4.2}$$

O estimador supra-citado é denominado V-estimador ou V-estatística. Para o caso em que $a = 1$, $V_n = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i)$.

Para $a > 1$, o somatório, na equação (4.2), contém alguns termos nos quais $i_1, ..., i_a$ não são todos distintos. Com isso, a esperança de tais termos não é necessariamente igual a $\theta = T(F)$ e, assim, V_n é viciado para.

Sabendo-se que o vício de V_n é devido às duplicações de $i_1, ..., i_a$, um caminho natural para corrigir este vício seria a restrição dos somatórios, na equação (4.2), ao conjunto dos índices $i_1, ..., i_a$ que não contém duplicações. Daí, obtém-se a U-estatística de grau a [**?**, **?**], definida por

$$U_n = \frac{1}{\begin{pmatrix} n \\ a \end{pmatrix}} \sum_{1 \leq i_1 <} \sum_{i_2 <} ... \sum_{< i_a \leq n} \phi(X_{i_1}, ..., X_{i_a})$$

Note que o tamanho amostral deve ser no mínimo a para que a U-estatística esteja definida e que a V-estatística é obtida tomando-se $a = 1$ na definição da U-estatística.

Além da U-estatística ser não-viciada, prova-se que ela possui normalidade assintótica, isto é, se $\sigma_i^2 < \infty$ para $i = 1, ..., a$ então $\sqrt{n}(U_n - \theta) \xrightarrow{d} N(0, a^2\sigma_1^2)$ onde $\sigma_i = Var_F(\phi_i(X_1, ..., X_i))$ [**?**, **?**].

Se as observações consistem de duas amostras independentes $X_1, ..., X_m$ e $Y_1, ..., Y_n$, i.i.d. dentro de cada amostra, podendo ter diferentes distribuições. Seja $\phi(x_1, ..., x_r, y_1, ..., y_s)$ uma função conhecida que seja permutação simétrica em $x_1, ..., x_r$ e $y_1, ..., y_s$, separadamente. A U-estatística é definida para duas amostras com núcleo ϕ como sendo [**?**]:

$$U_n = \frac{1}{\begin{pmatrix} m \\ r \end{pmatrix}} \frac{1}{\begin{pmatrix} n \\ s \end{pmatrix}} \sum_{1 \leq i_1 <} \sum_{i_2 <} ... \sum_{< i_r \leq m} \sum_{1 \leq j_1 <} \sum_{j_2 <} ... \sum_{< j_s \leq n} \phi(X_{i_1}, ..., X_{i_r}, Y_{j_1}, ..., Y_{j_s})$$

A U-estatística supra-citada é um estimador não-viciado do parâmetro $\theta = \mathbb{E}(\phi(X_1, ..., X_r, Y_1, ..., Y_s))$ e possui normalidade assintótica [**?**].

A seguir, serão apresentados alguns exemplos de U-estatísticas [**?**]:

1. A U-estatística de grau $a = 1$ é a média $\frac{\sum_{i=1}^{n} \phi(X_i)}{n}$;

2. Para o núcleo de grau 2, $\phi(x_1, x_2) = \frac{(x_1 - x_2)^2}{2}$, o parâmetro $\theta = \mathbb{E}\phi(X_1, X_2) = Var(X_1)$ é a variância das observações. A correspondente U-estatística é dada por

$$U = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \sum \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2;$$

3. O núcleo de grau 1, para x, y e para o parâmetro $\theta = P(X \leq Y)$, é $h(x, y) = \mathbb{I}_{\{X \leq Y\}}$. A U-estatística correspondente é

$$U = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{I}_{\{X_i \leq Y_j\}}$$

A estatística mnU é conhecida como a Estatística de Mann-Whitney e é utilizada no teste de Mann-Whitney para duas amostras.

Método da Bissecção

O Método da Bissecção [?] é um método numérico para a resolução de equações.

Definição 4.2. *Dada uma função $f(x)$, um número ξ é um zero da função $f(x)$ ou raíz da equação $f(x) = 0$ se $f(\xi) = 0$.*

A idéia central dos métodos para a resolução de equações é a seguinte: a partir de uma aproximação inicial para a raíz, refina-se essa aproximação através de um processo iterativo. Por isso, esses métodos numéricos consistem de duas fases:

1. Localização ou isolamento das raízes: Consiste na obtenção de um intervalo que contém a raíz. Esta fase é realizada por meio da análise teórica e gráfica da função $f(x)$;

2. Refinamento: Consiste em, escolhidas aproximações iniciais no intervalo encontrado na fase anterior, melhorá-las sucessivamente até obter-se uma aproximação para a raíz dentro de uma precisão ϵ prefixada.

O Método da Bissecção se baseia no seguinte teorema.

Teorema 4.1. *Seja $f(x)$ uma função contínua num intervalo $[a, b]$. Se o produto$f(a)f(b)$ for negativo ($f(a)f(b) < 0$) então existe pelo menos um ponto $x = \xi$ entre a e b que é zero de $f(x)$.*

Demonstração: Para a demonstração consulte [**?**].

A fase 2. do Método da Bissecção consiste em: Seja uma função $f(x)$ contínua no intervalo $[a, b]$ tal que $f(a)f(b) < 0$. Por simplicidade, será assumido que o intervalo (a, b) contém uma única raíz da equação $f(x) = 0$. O objetivo do Método da Bissecção é reduzir a amplitude do intervalo que contém a raíz até se atingir a precisão requerida, $(b - a) < \epsilon$, usando-se, para isto, a sucessiva divisão de $[a, b]$ ao meio.

Prova-se que a sequência $\{x_k\}$, gerada pelo método, converge para a raíz da função, isto é, $lim_{n\to\infty} x_k = \xi$ com as suposições de que a função f é contínua em $[a, b]$ e de $f(a)f(b) < 0$ [**?**].

O Método da Bissecção está implementado no "software" R através do comando "uniroot".

Método de Otimização Irrestrita de Nelder-Mead

O Método de Nelder-Mead [**?**, **?**] é um método de busca que considera poliedros em $\mathbb{R}^n$ ("simplex"). Um "simplex" correspondente à duas e três variáveis é, respectivamente, um triângulo e um tetraedro.

Este método minimiza uma função $f(\mathbf{x})$ (função objetivo) de n variáveis independentes utilizando $n + 1$ vértices de um poliedro flexível em $\mathbb{R}^n$ sendo que cada vértice em $\mathbb{R}^n$ pode ser definido por um vetor $\mathbf{x}$. O termo poliedro flexível é devido ao fato de que o Método de Nelder-Mead permite alterar a forma de um "simplex".

O vértice do poliedro flexível em $\mathbb{R}^n$, que atribui à $f(\mathbf{x})$ o maior valor dentre os demais vértices, é escolhido e, projetando-o através do centróide dos vértices remanescentes, um novo ponto é obtido para substituí-lo. Valores menores da função objetivo são obtidos pela sucessiva troca do vértice, do "simplex", que atribui o maior valor à $f(\mathbf{x})$, por pontos melhores até que o mínimo de $f(\mathbf{x})$ seja alcançado.

A seguir será dado detalhes do Método de Nelder-Mead:

Seja $\mathbf{x_i^{(k)}} = \left[x_{i1}^{(k)}, ..., x_{ij}^{(k)}, ..., x_{in}^{(k)}\right]^T$, $i = 1, ..., n+1$, o i-ésimo vértice em $\mathbb{R}^n$ na k-ésima iteração da busca, $k = 0, 1, ...$, e seja $f\left(\mathbf{x_i^{(k)}}\right)$ o valor da função objetivo em $\mathbf{x_i^{(k)}}$. Além disso, sejam $f\left(\mathbf{x_h^{(k)}}\right) = max\left\{f\left(\mathbf{x_1^{(k)}}\right), ..., f\left(\mathbf{x_{n+1}^{(k)}}\right)\right\}$ com a correspondência $\mathbf{x_i^{(k)}} = \mathbf{x_h^{(k)}}$ e $f\left(\mathbf{x_l^{(k)}}\right) = min\left\{f\left(\mathbf{x_1^{(k)}}\right), ..., f\left(\mathbf{x_{n+1}^{(k)}}\right)\right\}$ com a correspondência $\mathbf{x_i^{(k)}} = \mathbf{x_l^{(k)}}$. Desde que o poliedro em $\mathbb{R}^n$ é constituído de $(n+1)$ vértices, $\mathbf{x_1}, ..., \mathbf{x_{n+1}}$, considere $\mathbf{x_{n+2}}$ como sendo o centróide de todos os vérices exceto o vértice $\mathbf{x_h}$. As coordenadas do centróide são dadas por

$$\mathbf{x_{n+2,j}^{(k)}} = \frac{1}{n}\left[\left(\sum_{i=1}^{n+1} \mathbf{x_{ij}^{(k)}}\right) - \mathbf{x_{hj}^{(k)}}\right]$$

com $j = 1, ..., n$ onde o índice j representa cada coordenada de de $\mathbf{x_i^{(k)}}$.

O procedimento para a obtenção de um novo vértice em $\mathbb{R}^n$, de modo que $f(\mathbf{x})$ tenha um valor menor, envolve as quatro operações seguintes:

1. Reflexão: Reflete $\mathbf{x_h^{(k)}}$ em relação ao centróide através da expressão $\mathbf{x_{n+3}^{(k)}} = \mathbf{x_{n+2}^{(k)}} + \alpha\left(\mathbf{x_{n+2}^{(k)}} - \mathbf{x_h^{(k)}}\right)$ onde $\alpha > 0$ é o coeficiente de reflexão, $\mathbf{x_{n+2}^{(k)}}$ é o centróide e $\mathbf{x_h^{(k)}}$ é o vértice que torna o valor de $f(\mathbf{x})$ o maior dentre os $(n+1)$ valores de $f(\mathbf{x})$ na k-ésima iteração;

2. Expansão: Se $f\left(\mathbf{x_{n+3}^{(k)}}\right) \leq f\left(\mathbf{x_l^{(k)}}\right)$, expande-se o vetor $\left(\mathbf{x_{n+3}^{(k)}} - \mathbf{x_{n+2}^{(k)}}\right)$ por meio do computo da expressão $\mathbf{x_{n+4}^{(k)}} = \mathbf{x_{n+2}^{(k)}} + \gamma\left(\mathbf{x_{n+3}^{(k)}} - \mathbf{x_{n+2}^{(k)}}\right)$ onde $\gamma > 1$ é o coeficiente de expansão.

   Se $f\left(\mathbf{x_{n+4}^{(k)}}\right) < f\left(\mathbf{x_l^{(k)}}\right)$, troca-se $\mathbf{x_h^{(k)}}$ por $\mathbf{x_{n+4}^{(k)}}$ e continua-se a partir da etapa 1. com $k = k+1$. Caso contrário, troca-se $\mathbf{x_h^{(k)}}$ por $\mathbf{x_{n+3}^{(k)}}$ e continua-se a partir da etapa 1. com $k = k+1$;

3. Contração: Se $f\left(\mathbf{x_{n+3}^{(k)}}\right) > f\left(\mathbf{x_i^{(k)}}\right)$, para todo $i \neq h$, contrai-se o vetor $\left(\mathbf{x_h^{(k)}} - \mathbf{x_{n+2}^{(k)}}\right)$ por meio do computo da expressão $\mathbf{x_{n+5}^{(k)}} = \mathbf{x_{n+2}^{(k)}} + \beta\left(\mathbf{x_h^{(k)}} - \mathbf{x_{n+2}^{(k)}}\right)$ onde $0 < \beta < 1$ é o coeficiente de contração. Troca-se $\mathbf{x_h^{(k)}}$ por $\mathbf{x_{n+5}^{(k)}}$ e retorna-se para a etapa 1. para o proseguimento da busca na $(k+1)$-ésima iteração;

4. Redução: Se $f\left(\mathbf{x_{n+3}^{(k)}}\right) > f\left(\mathbf{x_h^{(k)}}\right)$, reduz-se todos os vetores $\left(\mathbf{x_i^{(k)}} - \mathbf{x_l^{(k)}}\right)$, $i = 1, ..., n+1$, pela metade, a partir de $\mathbf{x_l^{(k)}}$, computando-se $\mathbf{x_i^{(k)}} = \mathbf{x_l^{(k)}} + 0,5\left(\mathbf{x_i^{(k)}} - \mathbf{x_l^{(k)}}\right)$, $i = 1, ..., n+1$, e retorna-se à etapa 1. para o proseguimento da busca na $(k+1)$-ésima iteração.

Nelder e Mead recomendaram os valores $\alpha = 1$, $\beta = 0,5$ e $\gamma = 2$ como valores, geralmente, satisfatórios para a minimização irrestrita.

O critério usado por Nelder e Mead para finalizar a busca é o seguinte:

$$\left\{ \frac{1}{n+1} \sum_{i=1}^{n+1} \left[f\left(\mathbf{x_i^{(k)}}\right) - f\left(\mathbf{x_{n+2}^{(k)}}\right) \right]^2 \right\}^{1/2} \leq \epsilon$$

onde ϵ é um número arbitrariamente pequeno e $f\left(\mathbf{x_{n+2}^{(k)}}\right)$ é o valor da função objetivo no centróide $\mathbf{x_{n+2}^{(k)}}$.

A vantagem deste método é a não utilização de derivadas de primeira e segunda ordem da função $f(\mathbf{x})$. No entanto, isso requer um número maior de operações pois o método tem que avaliar o valor da função objetivo em $n+1$ pontos a cada iteração e fazer as projeções sendo que, no caso de termos uma função com muitas variáveis, isto pode se tornar computacionalmente caro.

Vale observar que $max f(\mathbf{x}) = -min(-f(\mathbf{x}))$, ou seja, a minimização de uma função $f(\mathbf{x})$ corresponde a maximização da função $-f(\mathbf{x})$. Assim, pode-se aplicar o método de minimização de Nelder-Mead à $-f(\mathbf{x})$ a fim obter-se a maxização da função $f(\mathbf{x})$.

O Método de Nelder-Mead está implementado no "software" R através do comando "optim" escolhendo-se a opção "method="Nelder-Mead"". Este método é um método de otimização irrestrita sendo que a utilização conjunta das rotinas "optim" e "constrOptim" permite a otimização de uma função objetivo com a restrição do tipo $ux - c \geq 0$ onde x representa o vetor de parâmetros a ser otimizado e u e v são vetores. O comando "constrOptim" utiliza uma função barreira [**?**]. A função barreira é uma função das restrições que é incorporada à função objetivo, ou seja, considera-se a otimização de uma função $F(x) = f(x) - cB(x)$ tal que o valor da função $F(x)$ tende ao infinito quando o ponto x aproxima-se da vizinhança da região formada pelas restrições onde $f(x)$, c e $B(x)$ são, respectivamente, a função objetivo, uma constante positiva e a função barreira. Assim, os pontos ótimos são pontos do interior ou da fronteira da região formada pelas restrições sendo que os pontos iniciais devem ser pontos pertencentes ao interior das restrições. No "software" R, é utilizada uma barreira logarítmica.

A seguinte figura ilustra o "simplex" de duas e três variáveis. A figura correspondente ao "simplex" de duas variáveis ilustra a etapa de reflexão. O ponto A representa o vértice que faz

Figura 4.6: Figuras correspondentes aos "simplex" de duas e três variáveis

com que a função objetivo tenha o maior valor. Uma projeção é feita a partir do ponto A através do centróide (ponto C). O ponto A é desconsiderado e um novo "simplex" é formado composto pelos pontos antigos remanescentes e um novo ponto, B, locado ao longo da linha de projeção.

Bibliografia