



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

# **Inferência em Modelos de Força e Tensão baseados em Cópulas**

por

**Alex Rodrigues do Nascimento**

Brasília

2018

# **Inferência em Modelos de Força e Tensão baseados em Cópulas**

**por**

**Alex Rodrigues do Nascimento**

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Bernardo Borba de Andrade

Brasília

2018

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do grau de Mestre em Estatística.

Texto aprovado por:

Prof. Bernardo Borba de Andrade  
Orientador, EST/UnB

Prof<sup>a</sup>. Cira Etheowalda Guevara Otiniano  
EST/UnB

Prof<sup>a</sup>. Cátia Regina Gonçalves  
MAT/UnB

---

*We have not got the money, so we have got to think.*

(Ernest Rutherford)

---

Para os meus pais.



Meus sinceros agradecimentos aos professores do PPGEST/UnB. Em especial, ao professor Bernardo por toda a paciência e dedicação despendida para que esse trabalho se concretizasse. Ao professor Rathie que nos apresentou o tema e ajudou em diversas etapas.

Meus agradecimentos a toda minha família que sempre me apoiou e ajudou em momentos de dificuldade. Em especial, ao meu irmão Igor o qual me motivou a seguir o caminho da Estatística.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



# Resumo

Esta dissertação faz uma revisão de modelos de Força e Tensão baseados em cópulas ao passo que expõe uma metodologia para estimação intervalar da confiabilidade associada. A literatura atual concentra-se na estimativa pontual sem referência ao erro de amostragem. Com base na teoria apresentada por Joe (2005), desenvolvemos uma estrutura para estimação intervalar com base em uma aproximação normal para a confiabilidade, diretamente e após a transformação logit, e, também, usando uma *ad hoc* distribuição beta. Uma aplicação com dados da Pesquisa de Orçamentos Familiares (POF / IBGE) é usado para ilustrar a metodologia e se baseia em um modelo com marginais Dagum com quatro diferentes cópulas (Gaussian, Frank, Clayton e Gumbel). A confiabilidade estimada foi proposta na literatura aplicada como medida de fragilidade financeira.

**Palavras chaves:** Modelos de Força e Tensão, Confiabilidade, Cópulas, Máxima Verossimilhança, Inferência para Marginais, Dependência, Intervalo de Confiança.



# Abstract

This dissertation reviews copula-based Stress-Strength models while developing a framework for interval estimation of the associated reliability. Current literature focuses on point estimation without reference to sampling error. Based on the theory developed by Joe (2005), we developed a framework for interval estimation based on a normal approximation for the reliability, directly and after logit transformation, and also using an *ad hoc* Beta distribution. An application with data from the Household Budget Survey (POF/IBGE) is used to illustrate the methodology based on a model with Dagum margins under four different copulas (Gaussian, Frank, Clayton and Gumbel). The estimated reliability has been proposed in the applied literature as a measure of financial fragility.

**Keywords:** Stress-Strength models, Reliability, Copula, Maximum Likelihood, Inference Functions for Margins, Dependence, Interval Estimation.



# Sumário

<b>1 Modelos de Força e Tensão <math>P(X &lt; Y)</math></b>	<b>3</b>
1.1 Introdução . . . . .	3
1.2 Modelos de Força e Tensão . . . . .	5
1.2.1 Confiabilidade de Exponenciais Independentes . . . . .	6
1.2.2 Confiabilidade de Normais Independentes . . . . .	9
1.2.3 Expressões Fechadas para R . . . . .	13
1.2.4 Dependência em Modelos de Força e Tensão . . . . .	15
1.2.5 Procedimentos de Estimação intervalar . . . . .	16
<b>2 Modelos de Força e Tensão Baseados em Cópulas</b>	<b>19</b>
2.1 Conceitos Básicos em Cópulas . . . . .	19
2.1.1 Cópula Produto . . . . .	27
2.1.2 Cópula Gaussiana . . . . .	27
2.1.3 Cópula Frank . . . . .	29
2.1.4 Cópula Gumbel . . . . .	29
2.1.5 Cópula Clayton . . . . .	30
2.2 Modelos de Força e Tensão Baseados em Cópulas . . . . .	32
2.2.1 Marginais Normais . . . . .	32
2.2.2 Marginais Exponencias . . . . .	35
2.3 Estimação para Modelos de Força e Tensão baseados em Cópulas . . . . .	39

---

2.3.1	Máxima Verossimilhança . . . . .	40
2.3.2	Funções de Inferência para Marginais . . . . .	41
2.4	Intervalos de Confiança para Confiabilidade . . . . .	45
2.4.1	Intervalos de Confiança Assintóticos para Confiabilidade . . . . .	46
2.4.2	Intervalos de Confiança com a Transformação Logito . . . . .	47
2.4.3	Intervalos de Confiança com Distribuição Beta . . . . .	48
2.4.4	Intervalos de Confiança Bootstrap . . . . .	50
<b>3</b>	<b>Estudo de Simulação</b>	<b>51</b>
3.1	Introdução . . . . .	51
3.2	Simulações para Marginais Normais . . . . .	52
3.3	Simulações para Marginais Gamma . . . . .	55
<b>4</b>	<b>Modelos de Força e Tensão como Medida de Fragilidade Financeira</b>	<b>59</b>
4.1	Introdução . . . . .	59
4.2	Distribuições Marginais . . . . .	62
4.3	Dependência das Marginais . . . . .	69
4.4	Fragilidade Financeira das famílias do Sudeste do Brasil . . . . .	71
4.5	Conclusão . . . . .	78
<b>5</b>	<b>Considerações finais</b>	<b>81</b>

# Introdução

Tradicionalmente em estatística ao se comparar dois grupos, com mensuração contínua e aproximadamente normais, utiliza-se a média. Quando a suposição de normalidade não é adequada, uma alternativa é recorrer a métodos não paramétricos os quais também fazem uso de medidas de localização. Após trabalhos de Wolfe e Hogg (1971), muito se produziu no contexto de modelos probabilísticos que comparem duas variáveis aleatórias não apenas localmente e sem suposição de normalidade, com a utilização de modelos de Força e Tensão. Os modelos de Força e Tensão são baseados no cálculo da quantidade  $R = P(X < Y)$  no qual  $X$  e  $Y$  são variáveis aleatórias e  $R$  é denominado confiabilidade.

A utilização de modelos de Força e Tensão flexibiliza problemas que se tem interesse na comparação de duas variáveis aleatórias, pois pode-se utilizar diversas distribuições conhecidas e não apenas supor normalidade, além de permitir que se incorpore uma estrutura de dependência entre as variáveis, seja por distribuições bivariadas ou com a utilização de cópulas.

Este trabalho faz uma revisão de modelos de Força e Tensão e cópulas visando aplicação que se tem objetivo comparar duas variáveis aleatórias.

Intervalos de confiança para a confiabilidade levando em consideração um grau de dependência entre as variáveis é pouco encontrado na literatura, sendo mais visto para modelos normais. Nesse contexto, o objetivo deste trabalho é explorar métodos de estimação que incorporem teoria de cópulas para  $R$ , de tal forma, propor intervalos assintóticos utilizando propriedades dos estimadores de máxima verossimilhança.

O capítulo 1 descreve os princípios básicos dos modelos de Força e Tensão, com exemplos

e conceitos. O capítulo 2 aborda de maneira mais detalhada a utilização de cópulas em modelos de Força e Tensão. Traz uma breve revisão da teoria de cópulas, aborda alguns modelos de Força e Tensão com diferentes marginais e cópulas e apresenta estimadores intervalares assintóticos para confiabilidade.

O capítulo 3 apresenta resultados de exercícios de simulação realizados com base na teoria apresentada nos capítulos anteriores. O capítulo 4 apresenta um modelo de Força e Tensão com variáveis dependentes como medida de fragilidade financeira, utilizando dados de despesa e renda de família residentes nos estados do Sudeste do Brasil.

# Capítulo 1

## Modelos de Força e Tensão $P(X < Y)$

Este capítulo descreve os princípios básicos dos modelos de Força e Tensão, com exemplos e conceitos.

### 1.1 Introdução

Os modelos de Força e Tensão são baseados no cálculo da quantidade  $R = P(X < Y)$  no qual  $X$  e  $Y$  são variáveis aleatórias e  $R$  é denominado confiabilidade. Esse tipo de abordagem flexibiliza o problema de comparar duas variáveis aleatórias. Este trabalho traz alguns tópicos que expõem os conceitos da medida  $P(X < Y)$ .

Por definição, o cálculo de  $R$ , para variáveis absolutamente contínuas, é dado por

$$R(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{\boldsymbol{\theta}}(x, y) I(x < y) dx dy, \quad (1.1)$$

no qual  $h_{\boldsymbol{\theta}}(x, y)$  é a função de densidade conjunta de  $X$  e  $Y$ . Com a hipótese de independência entre as variáveis, a densidade conjunta pode ser fatorada

$$R(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\theta_1}(x) g_{\theta_2}(y) I(x < y) dx dy, \quad (1.2)$$

no qual  $f_{\theta_1}(x)$  e  $g_{\theta_2}(y)$  são as funções de densidade de  $X$  e  $Y$ , respectivamente. Como a expressão possui uma função indicadora dentro da integral, a equação (1.2) é simplificada em termos da função de distribuição acumulada de  $X$ :

$$R(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} F_{X;\theta_1}(y)g_{\theta_2}(y)dy. \quad (1.3)$$

$R$  é função do parâmetro  $\boldsymbol{\theta} = (\theta_1, \theta_2)$ , o qual é usualmente estimado por Máxima Verossimilhança (MV).

A propriedade de invariância<sup>1</sup> (veja Casella e Berger (2002)) nos permite obter o Estimador de Máxima Verossimilhança (EMV) de  $R$  a partir de estimativas de  $\boldsymbol{\theta}$ ,  $\hat{R} = f(\hat{\theta}_1, \hat{\theta}_2)$ .

Como exemplo, ao supor que  $X$  e  $Y$  seguem distribuições independentes  $\text{Exp}(\theta_1)$  e  $\text{Exp}(\theta_2)$ , respectivamente, a expressão para cálculo de  $R$  é dada por

$$\begin{aligned} R(\boldsymbol{\theta}) &= \int_0^{\infty} (1 - e^{-\theta_1 y})\theta_2 e^{-\theta_2 y} dy \\ &= \int_0^{\infty} \theta_2 e^{-\theta_2 y} dy - \int_0^{\infty} \theta_2 e^{-(\theta_1 + \theta_2)y} dy \\ &= \frac{\theta_1}{\theta_1 + \theta_2}. \end{aligned}$$

Dado  $X_1, X_2, X_3, \dots$  uma amostra aleatória simples (AAS) de tamanho  $n_1$  de  $\text{Exp}(\theta_1)$  e  $Y_1, Y_2, Y_3, \dots$  uma AAS de tamanho  $n_2$  de  $\text{Exp}(\theta_2)$  então o EMV para  $R$  é dado por

$$\hat{R}_{EMV} = \frac{\frac{1}{\bar{X}}}{\left(\frac{1}{\bar{X}} + \frac{1}{\bar{Y}}\right)},$$

nos quais  $\bar{X}$  e  $\bar{Y}$  são as médias amostrais. Ao analisar as expressões acima verifica-se que, para este exemplo, o cálculo de  $R$  não foi complexo e, como resultado, não se tem uma estatística que compare apenas as variáveis localmente. Uma ampla discussão sobre modelos de

<sup>1</sup>Se  $\hat{\theta}$  é estimador de máximo verossimilhança (EMV) de  $\theta$ , então, para qualquer função  $\phi(\theta)$ ,  $\phi(\hat{\theta})$  é EMV de  $\phi(\theta)$ .

Força e Tensão é encontrada em Kotz e Pensky (2003), a principal referência para modelos com variáveis independentes.

Um resultado conhecido reduz o problema do cálculo de confiabilidade para duas variáveis aleatórias normais.

**Teorema 1.1.1.** Seja  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$  um vetor aleatório multivariado que possui distribuição normal  $p$ -variada  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , então a combinação linear dos elementos  $\mathbf{a}'\mathbf{Z} = a_1Z_1 + a_2Z_2 + \dots + a_pZ_p$  tem distribuição  $N_p(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ .

Para mais detalhes veja Johnson e Wichern (2004). Reduzindo  $\mathbf{Z}$  a um vetor aleatório bivariado  $\mathbf{Z} = (X, Y)$  e  $\mathbf{a}$  ao vetor  $(-1, 1)$ , tem-se o *EMV* de R para  $X$  e  $Y$  normais. Se  $X$  e  $Y$  são independentes temos

$$P(X < Y) = P(Y - X > 0) = \Phi\left(\frac{(\mu_2 - \mu_1)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right),$$

nos quais  $\mu_1$ ,  $\mu_2$  e  $\sigma_1^2$ ,  $\sigma_2^2$  são as médias e variâncias de  $X$  e  $Y$ , respectivamente, e  $\Phi$  a função de distribuição acumulada de uma normal padrão. Pelo princípio de invariância o *EMV* de R é dado por

$$\hat{R}_{MV} = \Phi\left(\frac{(\bar{Y} - \bar{X})}{\sqrt{\frac{n_1-1}{n_1}s_X^2 + \frac{n_1-1}{n_1}s_Y^2}}\right), \quad (1.4)$$

nos quais  $\bar{X}$  e  $\bar{Y}$ ,  $s_X^2$  e  $s_Y^2$  são as médias e variâncias amostrais, respectivamente. Neste exemplo fica claro que uma possível relação de dependência entre as variáveis pode ser naturalmente incorporada no método. Aqui pela matriz de covariância  $\boldsymbol{\Sigma}$ .

## 1.2 Modelos de Força e Tensão

Os modelos de Força e Tensão recebem essa denominação devido as suas aplicações de origem terem sido na engenharia, no qual  $X$  representava a tensão que o componente, material ou equipamento irá ser submetido e  $Y$  a força ou resistência do mesmo. Dessa forma, se  $X > Y$

o componente irá falhar ou o material irá se quebrar. Ao assumir que  $X$  e  $Y$  são variáveis aleatórias, tem-se o emprego de modelos de Força e Tensão.

Com esse tipo de abordagem é possível construir modelos nos quais  $X$  e  $Y$  seguem distribuição normal (Weerahandi e Johnson, 1992), distribuição Gamma (Krishnamoorthy et al., 2008), distribuição Weibull (Krishnamoorthy e Lin, 2010), distribuição de Lévi (Masoom Ali e Woo, 2005), exponencial com dois parâmetros (Sengupta, 2011), distribuição Dagum (AHMED, 2015). Otiniano et al. (2013) apresenta a confiabilidade quando  $X$  e  $Y$  são distribuições Estáveis Simétricas com diferentes índices de estabilidade, diferentes parâmetros de escala e parâmetros de deslocamento não nulos. Uma recente contribuição foi dada por Rathie e de SM Ozelim (2017) ao obter a expressão exata de  $R$  para distribuição de Lévi.

Nesta seção serão apresentados alguns resultados conhecidos na literatura em relação a conceitos inferenciais para se estimar  $R$  com algumas distribuições conhecidas, além de relatar alguns tópicos especiais na área.

### 1.2.1 Confiabilidade de Exponenciais Independentes

Como já apresentado na primeira seção, a estimativa de máxima verossimilhança de  $R$  supondo duas distribuições exponenciais é dada por

$$\hat{R}_{EMV} = \frac{\bar{Y}}{\bar{X} + \bar{Y}}. \quad (1.5)$$

Em muitas aplicações apenas a estimativa pontual não é suficiente para fundamentar alguma tomada de decisão ou mesmo comprovar a eficácia de um novo tratamento para aplicações na saúde, por exemplo. Nesse contexto, estimadores intervalares se fazem altamente apropriados. Ao supor duas distribuições exponenciais independentes para um modelo de Força e Tensão, um intervalo de confiança (IC) exato pode ser construído.

Dados  $n_1$  e  $n_2$  os tamanhos amostrais de  $X$  e  $Y$ , respectivamente, tanto  $n_1\bar{X}$  quanto  $n_2\bar{Y}$  possui distribuição gamma, com parâmetros  $(n_1, \theta_1)$  e  $(n_2, \theta_2)$ , respectivamente. Ao definir a

variável  $\Upsilon$  tal que

$$\Upsilon = \frac{n_1\theta_1\bar{X}}{n_1\theta_1\bar{X} + n_2\theta_2\bar{Y}},$$

reparametrizando o problema, temos  $\epsilon = n_1\theta_1\bar{X}$ ,  $\eta = n_2\theta_2\bar{Y}$  com  $\epsilon$  e  $\eta$  seguindo distribuição gamma com parâmetros  $(1, n_1)$  e  $(1, n_2)$  respectivamente. Ao expressar  $\Upsilon$  em termos das novas variáveis, temos

$$\Upsilon = \frac{\epsilon}{\epsilon + \eta}.$$

Pelo Jacobiano da transformação  $J = (1 - \Upsilon)^{-2}\eta$ , pode-se encontrar a seguinte distribuição conjunta de  $\Upsilon$  e  $\eta$

$$f(\Upsilon, \eta) = e^{\frac{-\eta}{(1-\Upsilon)}} \frac{\eta^{n_1+n_2-1} \Upsilon^{n_1-1}}{\Gamma(n_1)\Gamma(n_2)(1-\Upsilon)^{n_1+1}}.$$

Integrando-se em relação a  $\eta$  tem-se a distribuição marginal de  $\Upsilon$

$$f(\Upsilon) = [B(n_1, n_2)]^{-1} \Upsilon^{n_1-1} (1 - \Upsilon)^{n_2-1}, \quad 0 < \Upsilon < 1,$$

i.e.  $\Upsilon$  possui distribuição Beta com os parâmetros conhecidos  $n_1$  e  $n_2$ . Para qualquer  $0 < a < b$ ,

$$P(a < \Upsilon < b) = I_b(n_1, n_2) - I_a(n_1, n_2),$$

no qual,

$$I_x(n_1, n_2) = [B(n_1, n_2)]^{-1} \int_0^x z^{n_1-1} (1 - z)^{n_2-1} dz,$$

(função beta incompleta). Para associar  $\Upsilon$  a  $\hat{R}$  temos

$$\Upsilon = \left[ 1 + \frac{n_2 \hat{R} (1 - \hat{R})}{n_1 \hat{R} (1 - \hat{R})} \right]^{-1}.$$

O lado direito da igualdade é denominada quantidade pivotal. Assim, por

$$I_b(n_1, n_2) - I_a(n_1, n_2) = 1 - \gamma, \tag{1.6}$$

temos,

$$P\left(a < \left[1 + \frac{n_2 \widehat{R}(1 - \widehat{R})}{n_1 \widehat{R}(1 - \widehat{R})}\right]^{-1} < b\right).$$

Resolvendo a desigualdade, obtemos

$$P\left(\frac{n_2 \widehat{R}a}{n_1(1 - \widehat{R})(1 - a) + n_2 \widehat{R}a} < R < \frac{n_2 \widehat{R}b}{n_1(1 - \widehat{R})(1 - b) + n_2 \widehat{R}b}\right).$$

Uma dificuldade da aplicação deste intervalo de confiança exato é a solução da equação dada em (1.6) para  $a$  e  $b$ , tal expressão possui infinitas soluções e se deve preferir uma tal que  $a - b$  seja mínimo ou simétrico ( $a = q_{\frac{\gamma}{2}}, a = q_{1-\frac{\gamma}{2}}$ ). A vantagem da utilização deste resultado é a possibilidade de aplicá-lo para qualquer tamanho de amostras e foi apresentado por Kotz e Pensky (2003).

Uma outra abordagem mais simples é a utilização de propriedades assintóticas do EMV's. Conforme equação (1.5), temos que  $\widehat{R}$  (EMV de  $R$ ) é função de  $\widehat{\theta}_1 = 1/\overline{X}$  e de  $\widehat{\theta}_2 = 1/\overline{Y}$ ,  $\widehat{R} = f(\overline{X}, \overline{Y})$ . Sob condições de regularidade (serão apresentadas no capítulo seguinte), sabe-se que  $\widehat{R}$  possui distribuição assintótica dada por

$$\sqrt{n}(\widehat{R} - R) \xrightarrow{D} N(0, \nu),$$

Tal que  $\nu = j' \mathcal{I}^{-1} j$ ,  $\mathcal{I}$  representa a matriz de informação de Fisher e  $j' = \begin{pmatrix} \frac{\partial g}{\partial \theta_1} & \frac{\partial g}{\partial \theta_2} \end{pmatrix}$ , o gradiente. Para o modelo em questão  $\nu$  é dada por

$$\nu = \begin{bmatrix} \frac{\theta_2}{(\theta_2 + \theta_1)^2} & -\frac{\theta_1}{(\theta_1 + \theta_2)^2} \end{bmatrix} \begin{bmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{bmatrix} \begin{bmatrix} \frac{\theta_2}{(\theta_2 + \theta_1)^2} \\ -\frac{\theta_1}{(\theta_1 + \theta_2)^2} \end{bmatrix}.$$

Dessa forma, os limites de um intervalo de confiança  $(1 - \gamma)$  assintótico de  $R$  são dados por

$$\left(\widehat{R} - z_{\frac{\gamma}{2}} \sqrt{\frac{\widehat{\nu}}{n}}, \widehat{R} + z_{\frac{\gamma}{2}} \sqrt{\frac{\widehat{\nu}}{n}}\right),$$

no qual  $\hat{\nu}$  é dado pela matriz de informação observada de Fisher ( $\hat{\mathcal{I}} = \text{var}((\bar{Y}, \bar{X}))$ ) e o gradiente com os parâmetros estimados por MV,  $z_{\frac{\gamma}{2}}$  representa o quantil  $\frac{\gamma}{2}$  de uma distribuição normal padrão.

### 1.2.2 Confiabilidade de Normais Independentes

Na literatura, existem alguns resultados conhecidos para o intervalo de confiança da confiabilidade ( $R$ ). Owen et al. (1964) apresenta um método geral não paramétrico para o limite superior do intervalo de confiança para  $P(X > Y)$ . Church e Harris (1970) apresentam um intervalo de confiança aproximado para  $P(X > Y)$  quando  $X$  e  $Y$  possuem distribuição normal independentes com a distribuição de  $Y$  conhecida, Enis e Geisser (1971) apresentam uma abordagem Bayesiana para o problema e Reiser e Guttman (1986) generalizam esses dois trabalhos. Weerahandi e Johnson (1992) derivam um intervalo de confiança baseado em duas normais independentes com todos os parâmetros desconhecidos utilizando uma abordagem de p-valor generalizada. Guo e Krishnamoorthy (2004) constroem um limite inferior para  $R$  baseado na diferença de duas variáveis aleatórias normais.

Em trabalhos mais recentes, Barbiero (2011), considerando normais independentes, propõe um *bootstrap* paramétrico para construir um intervalo de confiança para  $R$  e relata variações assintóticas para o mesmo problema. Nguimkeu et al. (2015) apresentam um intervalo de confiança baseado na teoria de razão de máxima verossimilhança do sinal modificada.

O método de *bootstrap* proposto por Barbiero (2011) possui o seguinte procedimento:

1. estimar os parâmetros  $\mu_x, \mu_y, \sigma_x, \sigma_y$  por máximo verossimilhança
2. calcular  $R$  com a expressão dada em (1.4) utilizando os parâmetros estimados em 1.  $(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y)$ .
3. gerar amostras aleatórias *bootstrap* de  $x_B$  e  $y_B$ . Dado que  $x_B \sim N(\hat{\mu}_x, \hat{\sigma}_x)$  e  $y_B \sim N(\hat{\mu}_y, \hat{\sigma}_y)$ .

4. estimar  $R_b$  da mesma forma dada em 2.
5. repetir os passos 3 e 4 B vezes para obter a distribuição *bootstrap* de  $R_B$
6. estimar intervalo de confiança  $(1 - \gamma)$  *bootstrap* de  $R$  baseado nos percentis da amostra  $R_B$ .

Barbiero investiga a eficiência dos métodos propostos utilizando simulações, os compara com resultados assintóticos conhecidos na literatura e, para amostras relativamente grandes ( $n > 50$ ), os métodos apresentaram desempenho semelhantes, em termos de cobertura.

Ao considerar  $X$  e  $Y$  independentes com variâncias conhecidas e amostras de tamanho  $n_1$  e  $n_2$ , não necessariamente iguais. Temos

$$\left( (\bar{Y} - \bar{X}) - (\mu_2 - \mu_1) \right) / \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \sim N(0, 1).$$

Ao denotar

$$M = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}, \quad (1.7)$$

um intervalo de confiança  $(1 - \gamma)$  é dado por

$$P\left( \Phi\left(\frac{(\bar{Y} - \bar{X})}{\sqrt{\sigma_1^2 + \sigma_2^2}} - z_{\frac{\gamma}{2}}/\sqrt{M}\right) < R < \Phi\left(\frac{(\bar{Y} - \bar{X})}{\sqrt{\sigma_1^2 + \sigma_2^2}} + z_{\frac{\gamma}{2}}/\sqrt{M}\right) \right) = 1 - \gamma.$$

Para variâncias desconhecidas, Reiser e Guttman (1986) construíram um intervalo baseado em uma distribuição T não central. Dado que  $\sigma_1^2$  e  $\sigma_2^2$  podem ser estimados por  $s_x^2$  e  $s_y^2$ , então:

$$\bar{Y} - \bar{X} \sim N(\mu_2 - \mu_1, \sigma_1^2/n_1 + \sigma_2^2/n_2),$$

$$(n_1 - 1)s_x^2/\sigma_1^2 \sim \chi_{n_1-1} \text{ e } (n_2 - 1)s_y^2/\sigma_2^2 \sim \chi_{n_2-1}.$$

Com as três variáveis aleatórias independentes, considerando  $\Psi$ :

$$\Psi = \left( \frac{(\mu_2 - \mu_1)}{\sigma} \right),$$

com  $M$  dado em (1.7) e  $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ , então  $\sqrt{M}\sigma^{-1}\bar{Y} - \bar{X} \sim N(\sqrt{M}\Psi, 1)$  e

$$s^2 = s_x^2 + s_y^2 \sim \sigma_1^2/(n_1 - 1)\chi_{n_1-1}^2 + \sigma_2^2/(n_2 - 1)\chi_{n_2-1}^2.$$

Ao manipular a equação acima, temos

$$\frac{\nu s^2}{\sigma_1^2 + \sigma_2^2} \sim \chi_\nu^2, \text{ com } \nu = \frac{(\sigma_1^2 + \sigma_2^2)^2}{\left(\frac{\sigma_1^4}{n_1-1} + \frac{\sigma_2^4}{n_2-1}\right)}.$$

Dessa forma,

$$\sqrt{M}\widehat{\Psi} \sim t_\nu(\sqrt{M}\Psi),$$

no qual  $t_\nu(\sqrt{M}\Psi)$  representa uma distribuição T não central com  $\nu$  graus de liberdade e o parâmetro de não centralidade  $\sqrt{M}\Psi$ . Ao supor  $n_1 = n_2 = n$  e  $\sigma_1 = \sigma_2$ , a equação acima é simplificada com  $M = n$  e  $\nu = 2(n - 1)$ . Sem tal suposição,  $M$  e  $\nu$  podem ser estimados:

$$\widehat{M} = \frac{s_x^2 + s_y^2}{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}, \nu = \frac{(s_x^2 + s_y^2)^2}{\frac{s_x^4}{n_1-1} + \frac{s_y^4}{n_2-1}}.$$

Com isso se tem a seguinte aproximação:

$$\sqrt{\widehat{M}}\widehat{\Psi} \sim t_{\widehat{\nu}}(\sqrt{\widehat{M}}\Psi). \quad (1.8)$$

Para se obter um intervalo de confiança com a expressão dada acima é necessário resolver uma equação numericamente. Entretanto, a expressão pode ser simplificada utilizando a aproximação

$$(t_{\nu^*}(x) - x)/[1 + (t_{\nu^*}(x))^2/(2\nu^*)]^{\frac{1}{2}} \sim N(0, 1). \quad (1.9)$$

Aplicando (1.8) em (1.9):

$$\sqrt{\widehat{M}}(\widehat{\Psi} - \Psi)/\sqrt{1 + \widehat{M}\widehat{\Psi}^2/(2\widehat{\nu})} \sim N(0, 1).$$

O resultado apresentado é assintótico. Os limites do intervalo  $(1 - \gamma)$  são dados por

$$\widehat{\Psi} \pm \left( \sqrt{\frac{1}{\widehat{M}} + \frac{\widehat{\Psi}^2}{(2\widehat{\nu})}} \right) z_{\frac{\alpha}{2}}. \quad (1.10)$$

Uma outra abordagem pode ser construída utilizando o passo a passo apresentado em Mukherjee e Sharan (1985) para estimar a variância assintótica de  $\widehat{R}$ , dada por:

$$v(\widehat{R}) = \frac{1}{2\pi s^2} e^{-(\bar{X}-\bar{Y})^2/s^2} \left[ s_x^2/n + s_y^2/m + 1/2 \frac{(\bar{X} - \bar{Y})^2}{s^4} (s_x^4 + s_y^4/m) \right].$$

Uma forma de simplificar o problema utilizando a variância assintótica do estimador é considerar a variância da variável  $h$ :

$$h = \frac{\bar{X} - \bar{Y}}{\widehat{\sigma}_x^2 + \widehat{\sigma}_y^2}, \quad (1.11)$$

dada por:

$$v(\widehat{h}) = \frac{1}{s^2} (s_x^2/n + s_y^2/m + \frac{1}{2} \frac{(\bar{X} - \bar{Y})^2}{s^4} (s_x^2/n + s_y^2/m)),$$

e IC  $(1 - \gamma)$ :

$$P(\widehat{h} - z_{\frac{\alpha}{2}} \sqrt{v(\widehat{h})} < h < \widehat{h} + z_{\frac{\alpha}{2}} \sqrt{v(\widehat{h})}) = 1 - \gamma.$$

Um intervalo de confiança aproximado para  $R$  é derivado da expressão acima:

$$P(\Phi(\widehat{h} - z_{\frac{\alpha}{2}} \sqrt{v(\widehat{h})}) < R < \Phi(\widehat{h} + z_{\frac{\alpha}{2}} \sqrt{v(\widehat{h})})) = 1 - \gamma.$$

Esses métodos dependem do tamanho da amostra e podem apresentar inconsistências em

casos em que  $R$  está perto de 1. Na literatura existem alguns trabalhos que aplicam uma transformação de normalização em  $R$  visando uma estabilização da variância do estimador, gerando IC mais confiáveis. Krishnamoorthy e Lin (2010), por exemplo, sugerem uma transformação logit em  $R$ ,  $g(R) = \log\left(\frac{R}{1-R}\right)$ . Levando em consideração o método delta, se têm uma outra aproximação para IC  $(1 - \gamma)$ :

$$(e^L(1 + e^L)^{-1}, e^U(1 + e^U)^{-1}),$$

no qual L e U são os limites inferior e superior do IC para o logito:

$$\left( \log\left(\frac{\widehat{R}}{1 - \widehat{R}}\right) \pm z_{\frac{\alpha}{2}} \frac{\sqrt{v\widehat{R}}}{\widehat{R}(1 - \widehat{R})} \right).$$

Barbiero (2011) realiza um estudo de simulação para verificar as propriedades estatísticas dos intervalos de confiança apresentados nesta sessão. O IC dado em (1.10) apresentou bons resultados em termos de cobertura e comprimento do intervalo e, ainda, se mostrou resiliente em relação ao tamanho da amostra.

### 1.2.3 Expressões Fechadas para R

A estimação de  $R$  envolve o cálculo de uma integral que em muitos casos não possui solução analítica e faz com que métodos numéricos sejam utilizados para calcular  $P(X < Y)$ . As estimativas de  $R$  apresentadas até aqui possuem formas fechadas para cálculo, ou seja, as funções das estimativas de máximo verossimilhança são expressões exatas. Rathie e de SM Ozelim (2017) obtém a expressão de  $R$  para distribuição de Lévy independentes em termos de funções especiais.

Kotz e Pensky (2003) explicitam a confiabilidade considerando várias distribuições conhecidas (dois resultados estão informadas na tabela 1.1). Como exemplo, um resultado apresentado por Kotz é a confiabilidade para Gamas independentes o qual apresentaremos aqui. A função

de densidade da Gama com  $y > 0$  é dada por:

$$f_{\alpha,\beta}(y) = \frac{y^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{y}{\beta}}, \quad \alpha, \beta > 0.$$

Para  $Y_1$  e  $Y_2$  variáveis aleatórias independentes as quais seguem distribuição gama,  $Y_1 \sim \text{Gama}(\alpha_1, \beta_1)$  e  $Y_2 \sim \text{Gama}(\alpha_2, \beta_2)$  observa-se

$$\begin{aligned} R &= P\left(\frac{Y_2}{Y_1} > 1\right) = P\left(\frac{Y_2/\beta_2}{Y_1/\beta_1} > \frac{\beta_1}{\beta_2}\right) \\ &= P\left(\frac{Y_1/\beta_1}{Y_2/\beta_2 + Y_1/\beta_1} < \frac{\beta_2}{\beta_1 + \beta_2}\right). \end{aligned}$$

Dado que a variável aleatória  $Z = \frac{Y_1/\beta_1}{Y_2/\beta_2 + Y_1/\beta_1}$  possui distribuição beta (Johnson et al., 1995), então

$$R = I_{\frac{\beta_2}{\beta_1 + \beta_2}}(\alpha_1, \alpha_2).$$

No qual  $I_y(a, b)$  é a função beta incompleta. O suporte gama ( $\mathbb{R}^+$ ) faz com que essa distribuição seja uma opção natural utilizada em modelos de Força e Tensão na engenharia, dado que considerar a força de um material ou a tensão negativas não tem sentido físico.

Muitas dessas expressões não são ditas triviais, mas possuem formas fechadas que dispensam a necessidade de utilização de métodos numéricos, a menos de funções especiais que possuem implementações conhecidas como a beta e gamma. Entretanto, muitos dos resultados supõem distribuição comum e independência para  $X$  e  $Y$ .

Em casos que existe dependência entre as variáveis, ou estas não possuem distribuição comum, a obtenção de  $R$  em uma forma exata tende a ser complicada. Domma e Giordano (2012), por exemplo, calculam  $R$  em um caso aplicado utilizando métodos numéricos, levando em consideração uma estrutura de dependência entre as variáveis incorporada pela cópula Frank, e, em contrapartida, os mesmos autores, Domma e Giordano (2013), demonstram uma expressão fechada para  $R$  inserindo, também, cópulas no problema (cópula fgm nesse caso).

**Tabela 1.1:**  $P(X < Y)$  para distribuições conhecidas.

X	Y	P(X<Y)
$\text{Exp}(\mu_1, \sigma_1)$	$\text{Exp}(\mu_2, \sigma_2)$	$\begin{cases} \frac{\sigma_1}{\sigma_1 + \sigma_2} e^{-\sigma_2(\mu_1 - \mu_2)}, & \text{se } \mu_1 > \mu_2 \\ 1 - \frac{\sigma_2}{\sigma_1 + \sigma_2} e^{-\sigma_1(\mu_2 - \mu_1)}, & \text{se } \mu_1 \leq \mu_2 \end{cases}$
$\text{Weibull}(\alpha_1^1, \sigma_1)$	$\text{Weibull}(\alpha_2^1, \sigma_2)$	$\begin{cases} 1 - \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{\sigma_2}{\sigma_1}\right)^{j\alpha_1} \Gamma\left(\frac{\alpha_1}{\alpha_2} j + 1\right), & \text{se } \alpha_1 < \alpha_2 \\ \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{\sigma_1}{\sigma_2}\right)^{j\alpha_2} \Gamma\left(\frac{\alpha_2}{\alpha_1} j + 1\right), & \text{se } \alpha_1 > \alpha_2 \end{cases}$

<sup>1</sup> $\alpha_1$  e  $\alpha_2$  conhecidos.

### 1.2.4 Dependência em Modelos de Força e Tensão

A suposição de independência entre as variáveis  $X$  e  $Y$  não são adequadas em muitos casos. Um tratamento usual é considerar distribuições bivariadas conhecidas e obter expressões para o cálculo de  $R$  que incorpore uma estrutura de dependência entre as variáveis do problema. Owen et al. (1964) apresenta, ainda, um intervalo aproximado para quando  $X$  e  $Y$  possuem distribuição normal bivariada com variâncias e covariâncias conhecidas com base em uma amostra observada  $(x_1, \dots, x_{n_1}$  e  $y_1, \dots, y_{n_2})$  com  $n_1 = n_2 = n$ .

Dado que  $\Phi(\cdot)$  é monotonicamente crescente, o cálculo de um intervalo de confiança para um problema abordado por Owen se resume em calcular um intervalo de confiança para a função  $\Psi$ :

$$\Psi = \left( \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} \right).$$

Ao denotar:

$$\hat{\Psi} = \left( \frac{(\bar{Y} - \bar{X})}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} \right).$$

Temos que  $\sqrt{n}(\hat{\Psi} - \Psi)$  segue distribuição normal padrão. Dessa forma, um intervalo de confiança  $(1 - \gamma)$  para  $\Psi$  é dado por:

$$P\left(-z_{\frac{\alpha}{2}}/\sqrt{n} < \hat{\Psi} - \Psi < z_{\frac{\alpha}{2}}/\sqrt{n}\right) = 1 - \gamma,$$

e, conseqüentemente, um intervalo de confiança  $(1 - \gamma)$  para  $R$  é dado por

$$P\left(\Phi(\hat{\Psi} - z_{\frac{\alpha}{2}}/\sqrt{n}) < R < \Phi(\hat{\Psi} + z_{\frac{\alpha}{2}}/\sqrt{n})\right) = 1 - \gamma.$$

Kotz e Pensky (2003) trazem a generalização do resultado acima para o caso da matriz de variâncias e covariâncias desconhecida. O procedimento adotado aqui é semelhante ao apresentando para o IC da função  $h$  dada em 1.11.

A dependência entre  $X$  e  $Y$  em modelos de Força e Tensão foi abordada em alguns estudos utilizando distribuições bivariadas: Normal bivariada Gupta e Subramanian (1998), Exponenciais bivariada (Nadarajah e Kotz, 2006) e Log-Normal bivariada (Gupta et al., 2012). Entretanto, esse tipo de abordagem tem como hipótese que as variáveis possuem alguma forma específica de dependência e que as distribuições marginais sejam do mesmo tipo. Deste modo, a utilização da teoria de cópulas permite maior flexibilidade e será apresentada de maneira mais detalhada no próximo capítulo.

### 1.2.5 Procedimentos de Estimação intervalar

Os intervalos de confiança para a confiabilidade apresentados até aqui são para distribuições bem difundidas na estatística. Kotz apresenta intervalos, exatos ou assintóticos, para outras diversas distribuições como gamma, exponencial com dois parâmetros, gamma generalizada, Burr X e exponencial truncada a esquerda. O autor comenta que a literatura da época não apresentava intervalos de confiança exatos além dos que ele mostrou em seu livro e sugere a utilização de transformações monótonas nas variáveis para tal finalidade.

Em trabalhos mais recentes, Krishnamoorthy e Lin (2010) demonstram um intervalo de confiança para  $R$  com modelos Weibull utilizando quantidade pivotal generalizada. Nguimkeu et al. (2015) mostram um intervalo de confiança mais preciso para  $R$  com  $X$  e  $Y$  normais independentes baseado na teoria de razão de máxima verossimilhança do sinal modificada.

Intervalos de confiança para a confiabilidade levando em consideração um grau de depen-

dência entre as variáveis é pouco encontrado na literatura, sendo mais visto para modelos normais. Nesse contexto, este trabalho explora métodos de estimação que incorporem teoria de cópulas para  $R$ , de tal forma, propõem intervalos assintóticos utilizando propriedades do estimadores de máxima verossimilhança.



## Capítulo 2

# Modelos de Força e Tensão Baseados em Cópulas

Este capítulo aborda de maneira mais detalhada a utilização de cópulas em modelos de Força e Tensão. Traz uma breve revisão da teoria de cópulas, aborda alguns modelos de Força e Tensão com diferentes marginais e cópulas e apresenta estimadores intervalares assintóticos para confiabilidade.

A partir deste capítulo, algumas notações básicas, que foram utilizadas nos itens anteriores com o propósito de preservar a intuição e história por trás da medida  $P(X < Y)$ , mudam com o objetivo de trazer mais dinamismo ao texto e auxiliar na exposição das equações de estimação presentes na teoria de máxima verossimilhança.

### 2.1 Conceitos Básicos em Cópulas

A teoria de cópulas possui sólida literatura, com artigos, monografias e livros bem estruturados que abordam o tema em diversos aspectos, com diferentes níveis de detalhes e aplicações. O teorema base do assunto é apresentado por Sklar (1959), uma ampla discussão sobre o tema é encontrado no livro de Nelsen (2007) e na monografia apresentada por Joe (1997), duas das

principais referências da área. Schweizer et al. (1981) explicitam a relação entre cópulas e a dependência de variáveis aleatórias, Nelsen (1991) mostra que conceitos de dependência podem ser interpretados como propriedades de cópulas. Cherubini et al. (2004) aborda a teoria com foco em finanças e McNeil et al. (2005) com foco em gestão de risco. Em um trabalho mais recente, Joe (2014) aborda de forma completa toda a teoria da área.

Tendo em vista toda a literatura já existente, esta seção será delimitada em uma breve revisão de conceitos básicos de cópulas, sem aprofundamento sobre a construção, implementação e propriedades, com foco na aplicação da teoria em modelos de Força e Tensão. Tal aplicação é, basicamente, ajustar uma distribuição bivariada para as variáveis de Força e Tensão.

Até os anos 80, a maioria das publicações relacionados a estimação da medida  $P(Y_2 < Y_1)$  tinham como hipótese a independência de  $(Y_1, Y_2)$  a qual não é válida para muitos casos reais. Os primeiros estudos que incorporam dependência entre as variáveis foram conduzidos por Awad et al. (1981) e Abu-Salih e Shamseldin (1988) com a utilização de distribuição exponencial bivariada. Depois, Hanagal (1997) trabalhou com a pareto bivariada e Nadarajah (2005a) apresenta expressões explícita para  $R$  considerando a beta bivariada como distribuição conjunta.

As distribuições bivariadas/multivariadas estão em um contexto que se tem interesse pela análise simultânea de variáveis aleatórias. Considerando  $\mathbf{y} = (Y_1, Y_2)$  um vetor aleatório bivariado, a função de distribuição (FD) conjunta das variáveis aleatórias  $Y_1$  e  $Y_2$  é definida como

$$F(\mathbf{y}) = F(y_1, y_2) = P(Y_1 < y_1, Y_2 < y_2).$$

As condições necessárias e suficientes para uma função contínua a direita  $F$  em  $\mathbb{R}^2$  ser função de distribuição são

1.  $\lim_{y_j \rightarrow \infty} F(y_1, y_2) = 0, j = 1, 2$

e

$$\lim_{y_j \rightarrow \infty \forall j} F(y_1, y_2) = 1$$

2.  $F(y_1, y_2)$  é não decrescente em cada uma das variáveis  $y_1$  e  $y_2$ .

3. Para todos  $(a_1, a_2)$  e  $(b_1, b_2)$  com  $a_1 < b_1$ ,  $a_2 < b_2$ ,

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

Existem diversos trabalhos que abordam tópicos sobre distribuições bivariadas e multivariadas, Mardia (1970) *Families of Bivariate Distributions*, por exemplo, apresenta a binomial bivariada, hipergeométrica, Poisson, Gaussiana, Cauchy, gamma, entre outras.

Dado  $\mathbf{y} = (Y_1, Y_2) \in \mathcal{Y}$ , em modelos de Força e Tensão a confiabilidade associada com a Tensão  $Y_1$  e Força  $Y_2$  é

$$R(\boldsymbol{\theta}) = \iint_{\mathcal{Y}} f_{\boldsymbol{\theta}}(\mathbf{y}) \mathbb{I}(y_1 < y_2) d\mathbf{y}, \quad (2.1)$$

no qual  $f_{\boldsymbol{\theta}}(\mathbf{y})$  é exatamente a distribuição do vetor aleatório  $\mathbf{y}$ . Se as marginais (distribuição de  $Y_1$  e  $Y_2$ ) possuem a mesma distribuição e existe uma bivariada resultante conhecida para  $\mathbf{y}$ , a confiabilidade será dada pela solução da integral em (2.1) (com os parâmetros conhecidos/estimados), o que pode ser realizado com métodos numéricos de integração.

A normal multivariada é a distribuição mais utilizada em aplicações da medida  $P(Y_1 < Y_2)$ , assim como em várias áreas científicas. Sua estrutura de dependência é resumida na matriz de correlação, a qual, em geral, tende a não ser a melhor medida de dependência. Por exemplo, a correlação é uma medida de dependência linear que pode não alcançar os limite de  $\pm 1$  quando a distribuição das variáveis não são Gaussianas. A utilização de distribuições bivariadas conhecidas para o cálculo de  $R(\boldsymbol{\theta})$  pressupõe que as marginais são do mesmo tipo e adotam estruturas específicas de dependência.

Existem diversos conceitos de dependência em problemas que envolvem a distribuição conjunta de duas variáveis aleatórias. Joe (2014) apresenta como conceitos básicos :

1. Quadrante positivo

Dado  $\mathbf{y} = (Y_1, Y_2)$  vetor aleatório bivariado com função de distribuição  $F_\theta(\mathbf{y})$ .  $\mathbf{y}$  possui dependência no quadrante positivo se:

$$P(Y_1 > a_1, Y_2 > a_2) \geq P(Y_1 > a_1)P(Y_2 > a_2) \quad \forall a_1, a_2 \in \mathcal{R}.$$

Esta desigualdade pode ser tratada como um conceito de dependência positiva porque, possivelmente,  $Y_1$  e  $Y_2$  são maiores juntos ou menores juntos comparados a  $Y'_1$  e  $Y'_2$ , no qual  $Y_1 \stackrel{d}{=} Y'_1$  e  $Y_2 \stackrel{d}{=} Y'_2$ , e  $Y'_1 \perp Y'_2$ .

## 2. Dependência estocástica positiva

Dado  $\mathbf{y} = (Y_1, Y_2)$  vetor aleatório bivariado com função de distribuição  $F_\theta(\mathbf{y})$ .  $Y_2$  é estocasticamente crescente em  $Y_1$  se:

$$P(Y_2 > y_2 | Y_1 = y_1) \text{ cresce a medida que } y_1 \text{ cresce, } \forall y_2.$$

Esse tipo de dependência implica em uma dependência no quadrante positivo.

## 3. Cauda à direita crescente

Dado  $\mathbf{y} = (Y_1, Y_2)$  vetor aleatório bivariado com função de distribuição  $F_\theta(\mathbf{y})$ .  $Y_2$  possui cauda a direita crescente em  $Y_1$  se

$$P(Y_2 > y_2 | Y_1 > y_1) \text{ cresce a medida que } y_1 \text{ cresce, } \forall y_2.$$

Se trata de um conceito de dependência positiva pois é mais provável que  $Y_2$  assuma valores maiores quando  $Y_1$  cresce.

## 4. Associação de Variáveis aleatórias

Dado  $\mathbf{y} = (Y_1, Y_2)$  vetor aleatório bivariado.  $\mathbf{y}$  é positivamente associado se :

$$E[g_1(\mathbf{y})g_2(\mathbf{y})] \geq E[g_1(\mathbf{y})]E[g_2(\mathbf{y})], \quad (2.2)$$

valer para quaisquer funções  $g_1$  e  $g_2$  reais e se as esperanças existirem. Esta desigualdade está ligada a uma dependência positiva devido a definição de covariância, a qual é positiva quando a (2.2) se configura.

### 5. Concordância

Um par de variáveis aleatórias é concordante se valores grandes de uma variável tendem a estar associados com valores grandes da outra variável e valores pequenos de uma variável com valores pequenos da outra (Nelsen, 2007).

Dado  $(y_{1i}, y_{2i})$  e  $(y_{1j}, y_{2j})$  duas observações independentes do vetor  $(Y_1, Y_2)$  de variáveis aleatórias contínuas, então  $(y_{1i}, y_{2i})$  e  $(y_{1j}, y_{2j})$  são concordantes se

$$(y_{1i} - y_{1j})(y_{2i} - y_{2j}) > 0,$$

e discordantes se

$$(y_{1i} - y_{1j})(y_{2i} - y_{2j}) < 0.$$

### 6. Dependência de Cauda

O conceito de dependência da cauda diz respeito à força de dependência na cauda do quadrante superior ou na cauda do quadrante inferior de uma distribuição bivariada.

Dado  $Y_1$  e  $Y_2$  variáveis aleatórias contínuas com função de distribuição  $F_1$  e  $F_2$ , respectivamente. Se define como parâmetro de dependência da cauda superior

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbf{P}[Y_2 > F_2^{-1}(t) | Y_1 > F_1^{-1}(t)] \quad (2.3)$$

e como parâmetro de dependência da cauda inferior

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbf{P}[Y_2 \leq F_2^{-1}(t) | Y_1 \leq F_1^{-1}(t)]. \quad (2.4)$$

Joe (2014) apresenta de maneira mais detalhada estruturas de dependência entre variáveis aleatórias.

Outro importante tema presente na teoria de cópulas são os limites Fréchet. Para uma distribuição bivariada  $F(F_1, F_2)$  com marginais  $(F_1, F_2)$  e  $(y_1, y_2) \in \mathbb{R}^2$ , os limites Fréchet  $(F^-, F^+)$  são:

$$F^- = \max\{0, F_1(y_1) + F_2(y_2) - 1\} \leq F(y_1, y_2) \leq \min\{F_1(y_1); F_2(y_2)\} = F^+ \quad (2.5)$$

$F^-$  e  $F^+$  são funções de distribuição bivariadas e, se contínuas, podem representar dependências perfeitas, negativa e positiva respectivamente. Dado  $Y_1$  variável aleatória (V.A) com função de distribuição (F.D.)  $F_1$  e  $Y_2$  V.A. com F.D.  $F_2$ ,  $Y_1$  é quase certamente uma função crescente de  $Y_2$  se, e somente se,  $F(F_1, F_2)$  seguir distribuição bivariada  $F^+$ , e  $Y_1$  é quase certamente uma função decrescente de  $Y_2$  se, e somente se,  $F(F_1, F_2)$  seguir distribuição bivariada  $F^-$  (Nelsen, 2007, pag 32).

Os conceitos citados são importantes para construção e aplicação de distribuições bivariadas. A concordância, por exemplo, é uma forma de comparar distribuições com  $d$ -dimensões e mesmas marginais, ao verificar se uma distribuição multivariada específica apresenta um grau de dependência maior que outra (Joe, 2014). Outro exemplo é a dependência na cauda, a qual influencia de maneira relevante a probabilidade conjunta nos primeiros ou últimos percentis. Cópula é um maneira natural de se estudar dependência entre variáveis aleatórias (Embrechts et al., 2001)

De forma literal, uma cópula é uma função que faz a junção de duas ou mais distribuições univariadas para construir um função multivariada que leve em consideração uma dada estrutura de dependência. Cópula é uma função de distribuição multivariada com marginais uniforme(0, 1). Nelsen (2007) define a função Cópula e apresentaremos aqui.

**Definição 2.1.1.** Uma sub-cópula bidimensional é uma função  $C'$  com as seguintes propriedades:

- Seu Domínio ( $Dom$ ) é  $C' = S_1 \times S_2$ , no qual  $S_1$  e  $S_2$  são um subconjunto de  $I = [0, 1]$ , contendo 0 e 1.
- Seja  $B = [a_1, a_2] \times [b_1, b_2]$  um retângulo cujos vértices estão em  $Dom C'$

$$C'(a_2, b_2) - C'(a_2, b_1) - C'(b_1, a_2) + C'(a_1, b_1) \geq 0$$

- Para todo  $u \in S_1$  e  $v \in S_2$ ,

$$C'(u, 1) = u \text{ e } C'(1, v) = v \forall (u, v) \in S_1 \times S_2.$$

**Definição 2.1.2.** Uma cópula bidimensional é uma subcópula com domínio  $I^2$ .

**Teorema 2.1.1** (Teorema de Sklar). (Nelsen, 2007) Dado  $F(y_1, y_2)$  distribuição conjunta com marginais  $F_1$  e  $F_2$ . Existe uma cópula  $C$  para todo  $y_1$  e  $y_2 \in \mathbb{R}$  tal que:

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)); \tag{2.6}$$

Se  $F_1$  e  $F_2$  são contínuas, então  $C$  é única. Por outro lado, se  $C$  é uma cópula e  $F_1$  e  $F_2$  são funções de distribuição, então  $F$  dado em (2.6) é função de distribuição conjunta com marginais  $F_1$  e  $F_2$ .

Quando  $Y_1$  e  $Y_2$  são V.A.'s absolutamente contínuas, a função de densidade conjunta de  $(Y_1, Y_2)$  é dada por:

$$f(y_1, y_2) = c(F_1(y_1), F_2(y_2))f_1(y_1)f_2(y_2), \tag{2.7}$$

no qual  $f_1(y_1)$  e  $f_2(y_2)$  são as funções densidade de  $Y_1$  e  $Y_2$ , respectivamente, e  $c$  é denominada a densidade da cópula. Existem alguns métodos para a construção de cópulas, muitos desses são baseados na transformação de Laplace ou de Williamson, ou em propriedades de

distribuições univariadas, este trabalho não foca na construção e se restringe na apresentação de algumas cópulas que são úteis para modelos de dependência, para mais detalhes, veja Joe (1997).

Outro exemplo da importância dos conceitos de dependência citados é a relação entre os parâmetro de dependência da cauda superior (inferior) com a dependência da cauda superior (inferior) de uma cópula. Dado  $Y_1$  e  $Y_2$  variáveis aleatórias contínuas com função de distribuição  $F_1$  e  $F_2$ , respectivamente,  $C$  a cópula de  $Y_1$  e  $Y_2$  e os parâmetro de dependência  $\lambda_U$  e  $\lambda_L$ , se os limites em 2.3 e 2.4 existirem, então

$$\lambda_U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t}$$

e

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}.$$

O resultado apresentado se trata de um teorema apresentado em (Nelsen, 2007, pag 214). Se  $\lambda_U \in (0, 1]$ , então a cópula possui dependência na cauda superior e se  $\lambda_U = 0$  a cópula não possui dependência na cauda superior (análise similar para  $\lambda_L$ ).

No caso de cópulas, os limites de Fréchet () satisfazem:

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

Em uma concepção prática, a utilização de cópulas exige a necessidade de se avaliar a dependência entre duas variáveis aleatórias, as medidas de associação monótonas estão nesse contexto e se referem ao conceito de que se uma variável crescer a outra tende a crescer. Scarsini (1984) apresenta axiomas que avaliam medidas de associação. A correlação linear ( $\rho_r$ ) é um exemplo de medida de associação monótona, mas não cumpre os axiomas apresentados por Scarsini e por isso não é considerada a melhor medida de associação monótona, sendo mais utilizada o *Tau de Kendall* ( $\tau$ ) e a *Correlação de Spearman* ( $\rho_s$ ).

Dado  $(Y_{11}, Y_{21})$  e  $(Y_{12}, Y_{22})$  vetores aleatórios independentes e identicamente distribuídos, cada um com função de distribuição conjunta  $F$ , a versão populacional do *Tau de Kendall* ( $\tau$ ) é dada por

$$\tau = \mathbf{P}[(Y_{11} - Y_{12})(Y_{21} - Y_{22}) > 0] - \mathbf{P}[(Y_{11} - Y_{12})(Y_{21} - Y_{22}) < 0],$$

*Tau de Kendall* é a probabilidade de concordância menos o probabilidade de discordância.

Dado  $(Y_{11}, Y_{21})$ ,  $(Y_{12}, Y_{22})$  e  $(Y_{13}, Y_{23})$  vetores aleatórios independentes com função de distribuição conjunta  $F$  comum, a versão populacional da *Correlação de Spearman* ( $\tau$ ) é dada por

$$\rho_s = 3(\mathbf{P}[(Y_{11} - Y_{12})(Y_{21} - Y_{23}) > 0] - \mathbf{P}[(Y_{11} - Y_{12})(Y_{21} - Y_{23}) < 0]).$$

Joe (2014) apresenta tais medidas e as suas propriedades, além de abordar outras.

Em geral, as família de cópulas dependem de um ou mais parâmetros denominado parâmetros de associação ( $\delta$ ), os quais se relacionam com o grau de dependência entre as marginais.

### 2.1.1 Cópula Produto

A cópula mais simples, cópula produto, possui a forma

$$C(u, v) = uv, \quad 0 < u, v < 1.$$

A cópula produto é importante porque corresponde a independência.

### 2.1.2 Cópula Gaussiana

A cópula Gaussiana é derivada da normal multivariada, após as distribuições individuais serem transformadas em normais univariadas. Para um caso bivariado, a cópula Gaussiana é parame-

trizada por  $\rho \in [-1, 1]$  e sua função de distribuição acumulada é dada por

$$C_\rho(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho), \quad 0 < u, v < 1.$$

Sua densidade é dada por:

$$c_\rho(u, v) = (1 - \rho^2)^{-1/2} \exp\left[-\frac{1}{2}(y_1^2 + y_2^2 - 2\rho y_1 y_2)(1 - \rho^2)\right] \exp\left[\frac{1}{2}(y_1^2 + y_2^2)\right],$$

no qual  $y_1 = \Phi^{-1}(u)$ ,  $y_2 = \Phi^{-1}(v)$ ,  $0 < u, v < 1$  e  $\Phi^{-1}$  a inversa da função de distribuição de uma normal padrão.

A cópula gaussiana tem como propriedades: ser crescente em concordância em  $\rho$ , ou seja, se  $\rho_1 < \rho_2$  então  $C_{\rho_1} < C_{\rho_2}$ ;  $\rho = -1$  implica dependência negativa perfeita,  $\rho = 0$  implica independência,  $\rho = 1$  implica dependência positiva perfeita; dependência no quadrante positivo com  $0 < \rho < 1$ ; dependência intermediária nas caudas para  $\rho > 0$  e simetria de reflexão<sup>1</sup>;  $\tau(\rho) = 2 \frac{\arcsin(\rho)}{\pi}$  (*Tau de Kendall*).

As cópulas arquimediana são uma classe amplamente utilizada em aplicações, pois se tem uma grande variedade de cópulas com diferentes estruturas de dependência e propriedades matemática desejáveis. Como propriedade matemática podemos citar que o *Tau de Kendall* ( $\tau$ ) pode ser expresso como uma integral unidimensional.

**Teorema 2.1.2.** (Embrechts et al., 2001) Dado  $Y_1$  e  $Y_2$  vetores aleatórios com uma cópula arquimediana gerada por  $\phi$ . O *Tau de Kendall* para  $Y_1$  e  $Y_2$  é dado por

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

---

<sup>1</sup>Dado  $(U_1, U_2) \sim C_1$  e  $(1 - U_1, 1 - U_2) \sim C_2$ ,  $C_1 \equiv C_2$

### 2.1.3 Cópula Frank

A cópula Frank pertence a família de cópulas arquimediana. Dado  $-\infty < \delta < \infty$ , a cópula Frank é dada por

$$C_\delta(u, v) = -\delta^{-1} \log \left( \frac{1 - e^{-\delta} - (1 - e^{\delta u})(1 - e^{\delta v})}{1 - e^{-\delta}} \right), 0 \leq u, v \leq 1,$$

e a densidade é

$$c_\delta(u, v) = \frac{\delta(1 - e^{-\delta})e^{-\delta(u+v)}}{[1 - e^{-\delta} - (1 - e^{-\delta u})(1 - e^{-\delta v})]^2}, 0 \leq u, v \leq 1.$$

Pode-se citar como propriedades da Frank: ser crescente em concordância;  $\delta \rightarrow -\infty$  implica dependência negativa perfeita,  $\delta \rightarrow 0^+$  implica independência,  $\delta \rightarrow \infty$  implica dependência positiva perfeita; dependência no quadrante positivo com  $\delta > 0$ ; simetria de reflexão (única da classe Arquimediana que possui essa propriedade); assimetria nas caudas; dependência fraca nas caudas;  $\tau(\delta) = \text{sgn}(\delta) \left( 1 + \frac{4}{|\delta|} (D(|\delta|) - 1) \right)$ .

### 2.1.4 Cópula Gumbel

A cópula Gumbel também pertence a classe arquimediana. Sua cópula é dada por

$$C_\delta(u, v) = \exp\{-([\log(u)]^\delta + [\log(v)]^\delta)^{1/\delta}\}, 0 \leq u, v \leq 1, 1 \leq \delta \leq \infty,$$

com densidade:

$$c_\delta(u, v) = \exp\{-[x^\delta + y^\delta]^{1/\delta}\} [(x^\delta + y^\delta)^{1/\delta} + \delta - 1] [x^\delta + y^\delta]^{1/\delta - 2} (xy)^{\delta - 1} (uv),$$

no qual  $x = -\log u$ ,  $y = -\log v$ .

Pode-se citar como propriedades da Gumbel: ser crescente em concordância;  $\delta = 1$  implica independência,  $\delta \rightarrow \infty$  implica dependência positiva perfeita; não apresenta dependência nega-

tiva; dependência no quadrante positivo; assimetria nas caudas; Dependência assimétrica, com forte dependência na cauda a direita ( $\lambda_U = 2 - 2^{1/\delta}$ ) e fraca a esquerda;  $\tau(\delta) = \frac{(\delta-1)}{\delta}$ .

### 2.1.5 Cópula Clayton

A cópula Clayton também pertence a classe arquimediana. Com o parâmetro de associação  $0 < \delta < \infty$ , sua cópula é dada por

$$C_\delta(u, v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}$$

Pode-se citar como propriedades da Clayton :  $\delta \rightarrow \infty$  implica dependência positiva perfeita; não apresenta dependência negativa; dependência assimétrica, com forte dependência na cauda a esquerda e fraca a direita;  $\tau(\delta) = \frac{(\delta)}{\delta+2}$ .

Para ajustar um modelo de dependência com cópulas, os passos básicos são: (i) identificar candidatos para os modelos univariados, (ii) determinar uma estrutura de dependência apropriada, (iii) avaliar comportamento da cauda conjunta e (iv) identificar modelos de cópulas candidatos que atendam aos requisitos (ii) e (iii) (Joe, 2014).

Com a utilização de teoria de cópulas existe a possibilidade de flexibilizar os modelos de Força e Tensão ao utilizar marginais de diferentes tipos e considerar diversas formas de dependência, o que não é possível ao usar a abordagem de distribuições bivariadas. Domma e Giordano (2013) apresentam uma expressão fechada para  $R$  considerando marginais Dagum e a cópula fgm para estrutura de dependência, Gao et al. (2016) utiliza uma mistura de cópulas para incorporar estrutura de dependência no problema.

Ao considerar um vetor bivariado  $\mathbf{y} = (Y_1, Y_2) \in \mathcal{Y}$ , com a estrutura de dependência modelada por uma cópula  $C$  parametrizada pelo escalar  $\delta \in D$  com marginais  $F_j$  parametrizadas por  $\alpha_j \in A_j \subseteq \mathbb{R}^{p_j}$ ,  $j = 1, 2$ , a expressão geral para calcular  $R$  utilizando cópulas é dada por:

$$R(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{-\infty}^{y_2} c_\delta(F_1(y_1), F_2(y_2)) f_1(y_1) f_2(y_2) dx dy, \quad (2.8)$$

no qual  $\boldsymbol{\theta} = (\theta_1, \theta_2)$ .

Avaliando a equação apresentada aqui e a expressão usual dada em (2.1), depreende-se que ao invés de se aplicar a densidade conjunta  $f_{\boldsymbol{\theta}}(\mathbf{y})$ , utiliza-se a densidade conjunta expressa em (2.7) composta pelas densidades marginais ( $f_1(y_1)$  e  $f_2(y_2)$ ) e a densidade da cópula ( $c$ )

## 2.2 Modelos de Força e Tensão Baseados em Cópulas

Essa seção apresenta alguns modelos de Força e Tensão com cópulas, considerando diferentes marginais paramétricas com diferentes estruturas de dependência e, com utilização de recursos gráficos, evidencia a importância de se considerar uma estrutura de dependência quando  $Y_1$  e  $Y_2$  não são independentes, com  $Y_1$  representando a variável Tensão e  $Y_2$  a variável Força.

Como abordagem inicial, será apresentado modelos de Força ( $Y_2$ ) e Tensão ( $Y_1$ ) com marginais seguindo distribuição normal.

### 2.2.1 Marginais Normais

Nesta etapa iremos considerar que  $Y_1 \sim N(0, 1)$  e  $Y_2 \sim N(1, 1)$ . Com a suposição de independência, a confiabilidade  $P(Y_1 < Y_2)$  possui valor de  $R = 0.76$  e, ao considerar que  $Y_1$  e  $Y_2$  seguem distribuição normal bivariada, a confiabilidade é dada pela expressão:

$$R = \Phi \left( \frac{\mu_{y2} - \mu_{y1}}{\sqrt{\sigma_{y1}^2 + \sigma_{y2}^2 - 2\sigma_{12}}} \right),$$

no qual  $\mu_{y2} = 1$ ,  $\mu_{y1} = 0$ ,  $\sigma_{y1} = 1$ ,  $\sigma_{y2} = 1$  e  $\sigma_{12}$  é a covariância de  $Y_1$  e  $Y_2$ . Dado que a função de distribuição conjunta é não decrescente em cada uma das variáveis  $y_1$  e  $y_2$ , verifica-se que se  $\mu_{y2} - \mu_{y1} > 0$  a confiabilidade é crescente em  $\sigma_{12}$  e  $\mu_{y2} - \mu_{y1} < 0$ , decrescente em  $\sigma_{12}$ , ou seja, se a média da variável que representa a Força do material for maior que média da variável que representa a Tensão, quanto mais correlacionada elas forem maior a probabilidade que o material não irá quebrar, mas se a média da variável Força for menor que a média da variável Tensão maior a probabilidade que o material irá quebrar quanto mais correlacionada elas forem. Nesse contexto, tem-se uma aplicação em finanças, ao supor que  $Y_1$  é um índice de referência (*Benchmarking*) a ser batido, salienta o interesse em escolher ativos que possuam esperança de retorno maior e são correlacionados positivamente com o *benchmarking*.

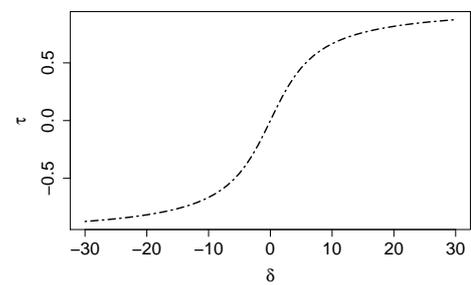
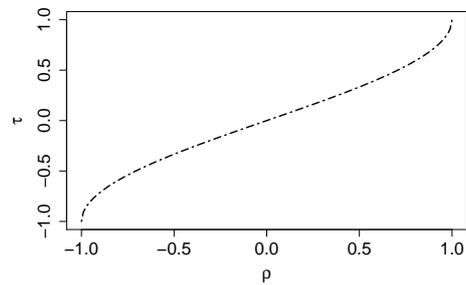
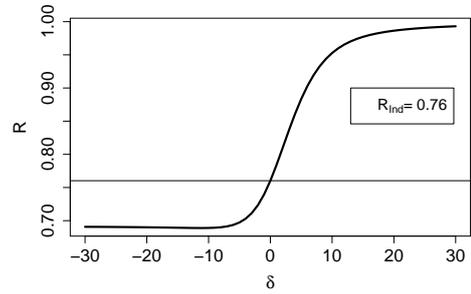
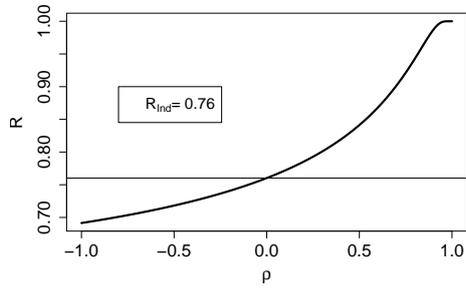
Ao avaliar a confiabilidade nos possíveis valores do parâmetro de dependência ( $-1 < \rho <$

1), identifica-se que  $R$  varia de 0.69 a 1.00 o que evidencia o impacto da dependência na medida ( $P(Y_1 < Y_2)$ ).

A figura (2.2.1) ilustra a confiabilidade para diferentes valores dos parâmetros de dependência da normal multivariada e cópulas Frank, Clayton e Gumbel com os parâmetros das marginais fixos ( $Y_1 \sim N(0, 1)$  e  $Y_2 \sim N(1, 1)$ ). Os resultados foram obtidos resolvendo a expressão (2.8) numericamente, com os parâmetros das marginais fixos e variando-se os parâmetros de dependência. Dessa forma, a figura se divide em quatro blocos, em cada bloco se tem dois gráficos, os quais associam a confiabilidade (gráfico acima) com o correspondente  $\tau$  de Kendall (gráfico abaixo) na região admissível do parâmetro de dependência (aqui expresso por  $\rho$  e  $\delta$ ) para cada distribuição bivariada considerada.

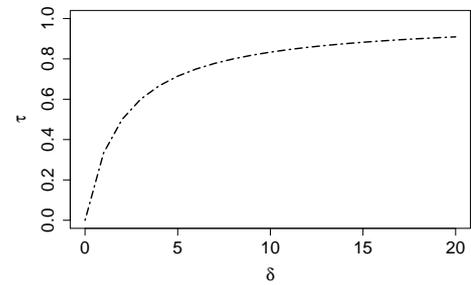
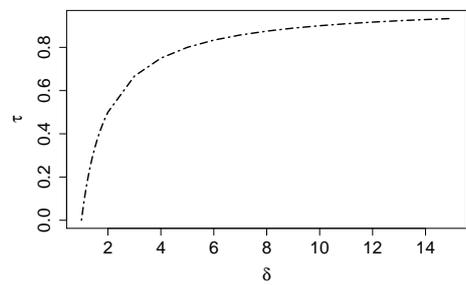
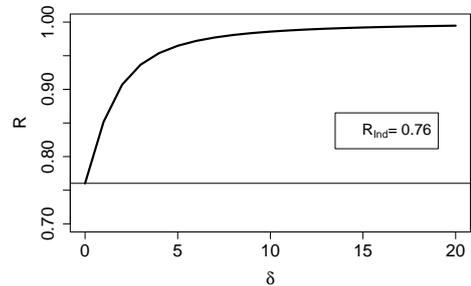
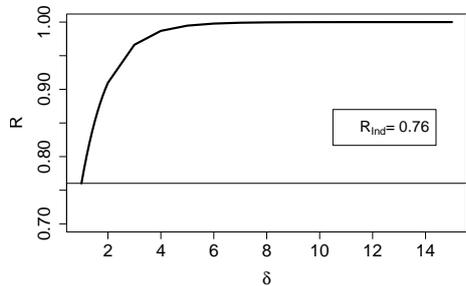
Depreende-se que a interpretação feita para a normal bivariada vale para as outras três cópulas apresentadas, aqui avaliando os parâmetros de associação, quando o parâmetro indica uma dependência mais forte (aumento de  $\tau$  de Kendall) a confiabilidade aumenta.

O comportamento da medida  $P(X < Y)$  para diferentes valores dos parâmetros de associação evidencia a influência da dependência. O erro de não considerar uma estrutura de dependência quando a mesma existe, tende a crescer a medida que  $\tau$  se aproxima de 1 ou -1.



(a) Normal Multivariada

(b) Cópula Frank



(c) Cópula Gumbel

(d) Cópula Clayton

**Figura 2.1:** (a)  $R$  para normal multivariada como função de  $\rho$  (b)  $R$  considerando cópula Frank como função de  $\delta$  (c)  $R$  considerando cópula Gumbel como função de  $\delta$  (d)  $R$  considerando cópula Clayton como função de  $\delta$ .

## 2.2.2 Marginais Exponencias

Nesta parte iremos considerar para o modelo de Força e Tensão marginais Exponenciais. Nadarajah e Kotz (2006) derivam uma expressão para  $R$  considerando a distribuição exponencial bivariada de Gumbel's a qual faremos uso e apresentaremos aqui. Kotz e Pensky (2003) também apresentam uma expressão para  $R$  considerando a distribuição exponencial bivariada de Gumbel's, mas com outra parametrização.

A função de densidade da exponencial bivariada de Gumbel's é dada por

$$f(y_1, y_2) = \{(1 - \omega)\alpha\beta + \omega\alpha^2\beta y_1 + \omega\alpha\beta^2 y_2 + \omega^2\alpha^2\beta^2 y_1 y_2\} e^{-(\alpha y_1 + \beta y_2 + \omega\alpha\beta y_1 y_2)},$$

para  $y_1 > 0$ ,  $y_2 > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$  e  $0 < \omega < 1$ . Esta foi uma das primeiras distribuições exponenciais bivariadas descobertas e possui aplicações em diversas áreas como tempo de falha, riscos competitivos, análise de precipitação regional (Nadarajah e Kotz, 2006). As distribuições marginais de  $Y_1$  e  $Y_2$  são Exponenciais com parâmetros  $\alpha$  e  $\beta$ , respectivamente, com  $E(Y_1) = \frac{1}{\alpha}$  e  $E(Y_2) = \frac{1}{\beta}$ . O coeficiente de correlação é dado por

$$\rho = 1 - \frac{1}{\omega} e^{\frac{1}{\omega}} \text{Ei}\left(\frac{1}{\omega}\right),$$

no qual  $\text{Ei}$  é a função exponencial integral definida como

$$\text{Ei}(x) = - \int_{-x}^{\infty} \frac{e^{-t}}{t} dt.$$

A confiabilidade para este caso é expresso por

$$\begin{aligned}
R &= \int_0^\infty \{(1-\omega)\alpha\beta + \omega\alpha^2\beta y_1\} \int_x^\infty e^{-(\alpha y_1 + \beta y_2 + \omega\alpha\beta y_1 y_2)} dy_2 dy_1 \\
&+ \int_0^\infty \{\omega\alpha\beta^2 y_2 + \omega^2\alpha^2\beta^2 y_1\} \int_x^\infty y_2 e^{-(\alpha y_1 + \beta y_2 + \omega\alpha\beta y_1 y_2)} dy_2 dy_1 \\
&= \int_0^\infty \frac{(1-\omega)\alpha + \omega\alpha^2 y_1}{1 + \omega\alpha y_1} e^{-(\alpha y_1 + \beta y_1 + \omega\alpha\beta y_1 y_1)} dy_1 + \int_0^\infty \frac{\omega\alpha + \omega^2\alpha^2 y_1}{(1 + \omega\alpha y_1)^2} e^{-(\alpha y_1 + \beta y_1 + \omega\alpha\beta y_1 y_1)} dy_1 \\
&= \alpha I_1 + \omega\alpha\beta I_2
\end{aligned}$$

no qual,

$$\begin{aligned}
I_1 &= \int_0^\infty e^{-(\alpha y_1 + \beta y_1 + \omega\alpha\beta y_1 y_1)} dy_1 \\
I_2 &= \int_0^\infty y_1 e^{-(\alpha y_1 + \beta y_1 + \omega\alpha\beta y_1 y_1)} dy_1
\end{aligned}$$

E o Lema A.1 de (Nadarajah e Kotz, 2006) mostra que

$$\begin{aligned}
I_1 &= \sqrt{\frac{\pi}{\omega\alpha\beta}} e^{\frac{(\alpha+\beta)^2}{4\omega\alpha\beta}} \left[ 1 - \Phi\left(\frac{\alpha + \beta}{\sqrt{2\omega\alpha\beta}}\right) \right], \\
I_2 &= \frac{1}{2\omega\alpha\beta} - \frac{\sqrt{\pi}(\alpha + \beta)}{2(\omega\alpha\beta)^{\frac{3}{2}}} e^{\frac{(\alpha+\beta)^2}{4\omega\alpha\beta}} \left[ 1 - \Phi\left(\frac{\alpha + \beta}{\sqrt{2\omega\alpha\beta}}\right) \right], \tag{2.9}
\end{aligned}$$

no qual  $\Phi(\cdot)$  denota a função de distribuição acumulada de uma normal padrão. Ao usar as expressões acima, temos

$$R = \frac{1}{2} + \frac{\sqrt{\pi}(\alpha + \beta)}{2\sqrt{\omega\alpha\beta}} e^{\frac{(\alpha+\beta)^2}{4\omega\alpha\beta}} \left[ 1 - \Phi\left(\frac{\alpha + \beta}{\sqrt{2\omega\alpha\beta}}\right) \right]. \tag{2.10}$$

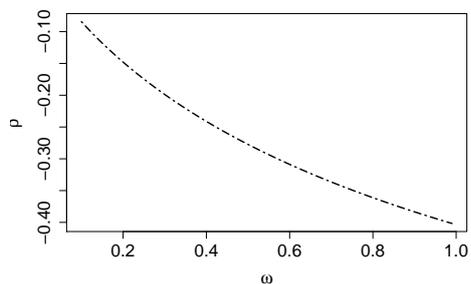
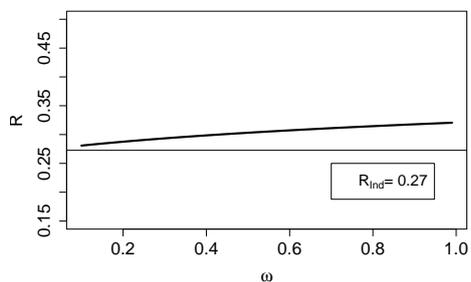
A figura (2.2) foi construída tendo como base o resultado em (2.10), calculando R com marginais Exponenciais  $Y_1 \sim Exp(3)$  e  $Y_2 \sim Exp(8)$  com diferentes valores para  $\omega$  e, ainda,

considerando modelos com essas mesmas marginais e diferentes cópulas, Frank, Clayton e Gumbel, com distintos valores para o parâmetro de dependência. Os resultados foram obtidos de maneira correspondente aos da etapa anterior. Esse exemplo evidencia, mais uma vez, a influência da dependência das variáveis na medida  $P(X < Y)$ .

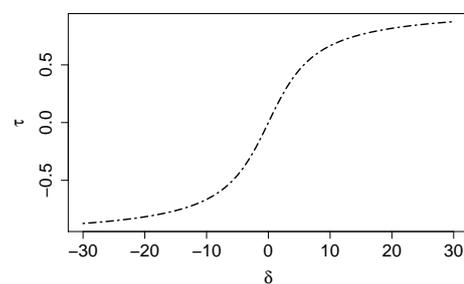
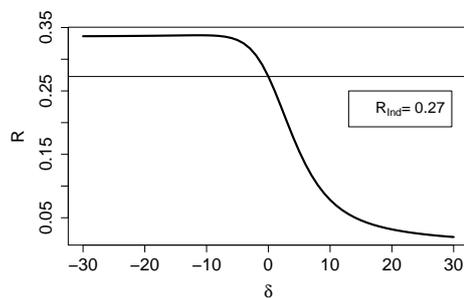
(Nadarajah, 2005b), o qual apresenta a confiabilidade para algumas distribuições Gama bivariadas, comete um equívoco ao calcular a confiabilidade para a distribuição gama bivariada McKay's (McKay, 1934) com função de densidade dada por:

$$f(y_1, y_2) = \frac{\alpha^{\beta_1 + \beta_2}}{\Gamma(\beta_1)\Gamma(\beta_2)} y_1^{\beta_1 - 1} (y_2 - y_1)^{\beta_2 - 1} e^{(-\alpha y_2)},$$

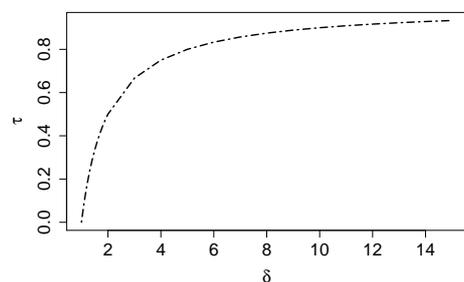
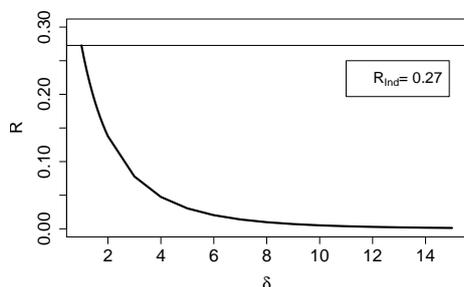
no qual  $y_2 > y_1 > 0$ ,  $\alpha > 0$ ,  $p > 0$ , e  $q > 0$ . Como  $y_2 > y_1 > 0$ , o suporte da distribuição, a medida  $P(Y_1 < Y_2)$  possui valor 1, a solução apresentada no trabalho possui erros nos cálculos.



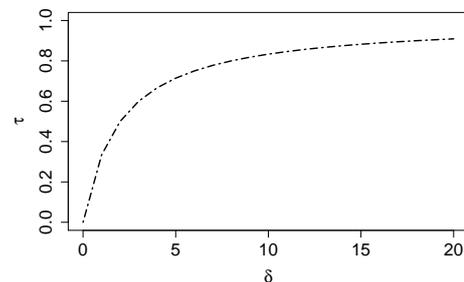
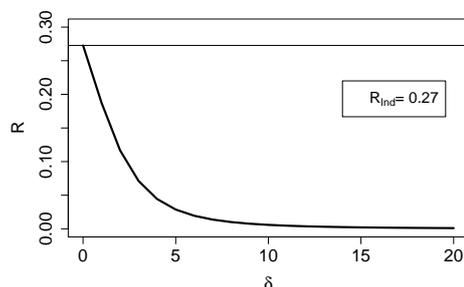
(a) Normal Multivariada



(b) Cópula Frank



(c) Cópula Gumbel



(d) Cópula Clayton

**Figura 2.2:** (a)  $R$  para Exponencial bivariada de Gumbel's como função de  $\omega$  (b)  $R$  considerando cópula Frank como função de  $\delta$  (c)  $R$  considerando cópula Gumbel como função de  $\delta$  (d)  $R$  considerando cópula Clayton como função de  $\delta$ .

### 2.3 Estimação para Modelos de Força e Tensão baseados em Cópulas

Nesta seção iremos, considerando modelos de Força e Tensão com cópulas, abordar métodos de estimação dos parâmetros das distribuições marginais, do parâmetro de dependência da cópula e da medida  $P(X < Y)$ , bem como apresentar métodos inferenciais.

Para esta seção vamos considerar um vetor bivariado  $\mathbf{y} = (Y_1, Y_2) \in \mathcal{Y}$ , com a estrutura de dependência modelada por uma cópula  $C$  parametrizada pelo escalar  $\delta \in D$ , com marginais  $F_j$  parametrizadas por  $\boldsymbol{\alpha}_j \in A_j \subseteq \mathbb{R}^{p_j}$ ,  $j = 1, 2$ . Vamos escrever  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$  com dimensão  $p = p_1 + p_2$  e  $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \delta)$  com dimensão  $d = p + 1$ . A distribuição conjunta e densidade de  $\mathbf{y}$  são

$$F_{\boldsymbol{\theta}}(\mathbf{y}) = C_{\delta}(F_1(y_1), F_2(y_2))$$

e

$$f_{\boldsymbol{\theta}}(\mathbf{y}) = c_{\delta}(F_1(y_1), F_2(y_2))f_1(y_1)f_2(y_2),$$

no qual  $dF_i = f_i dy_i$ . Iremos considerar a confiabilidade dada pela expressão (2.8) para a confiabilidade associada a Tensão  $Y_1$  e Força  $Y_2$ .

Uma amostra de  $n$  pares de observações de Tensão e Força será denotada pela matriz  $Y_{n \times 2} = [y_{ij}]$ . A verossimilhança marginal é dada por

$$\ell_j(\boldsymbol{\alpha}_j) = \sum_{i=1}^n \log f_j(y_{ij}), \quad j = 1, 2,$$

e a verossimilhança conjunta completa

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log f_{\boldsymbol{\theta}}(\mathbf{y}_i), \quad \mathbf{y}_i = (y_{i1}, y_{i2}).$$

Iremos considerar dois métodos paramétricos de estimação: máxima verossimilhança (MV) e um método de dois estágios de estimação conhecido como função de inferência para marginais (IFM).

### 2.3.1 Máxima Verossimilhança

Vamos considerar um modelo corretamente especificado com processo de geração de dados dado por  $f_{\theta^*}(\mathbf{y})$ , ou seja,  $\theta^*$  é o (verdadeiro) valor do parâmetro. Para o caso de uma verossimilhança regular com espaço paramétrico aberto, MV é equivalente a encontrar a raiz das equações escore o que é a versão amostral do fato que  $E[\Psi(\theta^*, \mathbf{y})] = \mathbf{0}$ , no qual  $\Psi$  é o (escore)  $d$ -dimensional do vetor de derivadas (com respeito a  $\theta$ ) da densidade conjunta,  $\Psi(\theta, \mathbf{y}) = \nabla_{\theta} \log f_{\theta}(\mathbf{y})^2$ .

No caso (correta especificação e verossimilhança bem comportada), nós podemos reduzir o problema de MV em termos das equações de estimação as quais são,

$$\sum_{i=1}^n \Psi(\theta, \mathbf{y}_i) \equiv \nabla_{\theta} \ell(\theta) = \begin{pmatrix} \nabla_{\alpha_1} \ell(\theta) \\ \nabla_{\alpha_2} \ell(\theta) \\ \frac{\partial \ell(\theta)}{\partial \delta} \end{pmatrix} = \mathbf{0}. \quad (2.11)$$

A solução (com respeito a  $\theta$ ) para (2.11) é o estimador de máxima verossimilhança (EMV),  $\hat{\theta}$ . Sob condições de regularidade, possui distribuição assintótica dada por

$$\sqrt{n}(\hat{\theta} - \theta^*) \rightarrow N(\mathbf{0}, H^{-1} I H^{-1}),$$

no qual

$$H(\theta)_{d \times d} = -E[\nabla \Psi(\theta, \mathbf{y})]$$

e

$$I(\theta)_{d \times d} = \text{Var}[\Psi(\theta, \mathbf{y})].$$

Usualmente,  $H = I$  e, para uma amostra suficientemente grande, a distribuição amostral do EMV é dado por  $N(\hat{\theta}, \hat{I}^{-1}/n)$ . Isso será assumido a partir de agora. As estimativas  $\hat{\theta}$  e  $\hat{I}$  são quase sempre obtidas utilizando otimização numérica para maximizar a verossimilhança por

<sup>2</sup>Dada uma função real  $g(\mathbf{t})$ ,  $\mathbf{t} \in \mathbb{R}^p$ ,  $\nabla_{\mathbf{t}} g$  denota o vetor  $p$ -dimensional de derivadas parciais  $(\frac{\partial g}{\partial t_1}, \dots, \frac{\partial g}{\partial t_p})$ .

meio de algoritmos tipo do Newton. Esse tipo de algoritmo gera, se obtiver sucesso, o máximo  $\hat{\theta}$  e uma aproximação para a Hessiana estimada como a iteração final  $n\hat{I} = -n\nabla_{\theta}\Psi(\theta, \mathbf{y})|_{\theta=\hat{\theta}}$ , conhecida como a informação observada de Fisher. A informação esperada de Fisher avaliada no máximo  $\hat{\theta}$  também pode ser usada para estimar a variância assintótica, mas raramente é fácil de calcular. Note que outros tipos de algoritmos como (simples) EM ou procura estocástica não geram diretamente uma aproximação para a Hessiana, nesses casos, somente  $\hat{\theta}$  é obtido.

### 2.3.2 Funções de Inferência para Marginais

O método de Funções de Inferência para Marginais é um procedimento de dois estágios (Joe, 2014), no qual os EMV's,

$$\tilde{\alpha}_j = \operatorname{argmax}_{\alpha_j \in A_j} \ell_j(\alpha_j), \quad j = 1, 2,$$

é inicialmente obtido, seguido pela maximização da verossimilhança completa com relação ao parâmetro da cópula,  $\delta$ , com o vetor de parâmetros marginais fixos  $\tilde{\alpha}$ ,

$$\tilde{\delta} = \operatorname{argmax}_{\delta \in D} \ell(\theta) |_{\alpha=\tilde{\alpha}}.$$

O estimador IFM é simplesmente  $\tilde{\theta} = (\tilde{\alpha}, \tilde{\delta})$ . O método de dois estágios é mais fácil de lidar computacionalmente do que estimação por máxima verossimilhança em modelos multivariados. MV pode se tornar computacionalmente inviável ou consumir muito tempo devido a um grande número de marginais (aqui temos que lidar apenas com 2), complexidade das marginais ou da cópula ou quando existem covariáveis. IFM é uma alternativa a MV e pode ser utilizado como boas estimativas iniciais para maximização numérica da verossimilhança completa. Também é conveniente para a comparação de diferentes cópulas com o mesmo conjunto de marginais. Sua eficiência, em relação à MV, foi examinada por Joe (2005), que adverte para a possibilidade de perda de eficiência com marginais contínuas e forte dependência.

IFM também pode ser construído em termos de equações de estimação, de maneira análoga a (2.11). Essas equações são baseada no vetor de  $d$  derivadas parciais,  $\Gamma = (\mathbf{g}_1, \mathbf{g}_2, g_3)$ , cuja a soma é zero em dois estágios,

$$\sum_{i=1}^n \Gamma(\boldsymbol{\theta}, \mathbf{y}_i) \equiv \begin{pmatrix} \sum_{i=1}^n \mathbf{g}_1(\boldsymbol{\alpha}_1, y_{i1}) \\ \sum_{i=1}^n \mathbf{g}_2(\boldsymbol{\alpha}_2, y_{i2}) \\ \sum_{i=1}^n g_3(\delta, \mathbf{y}_i) \end{pmatrix} = \mathbf{0}, \quad (2.12)$$

no qual

$$\mathbf{g}_j(\boldsymbol{\alpha}_j, y_{ij}) = \nabla_{\boldsymbol{\alpha}_j} \log f_j(y_{ij}), \quad j = 1, 2, \quad (2.13)$$

e

$$g_3(\delta, \mathbf{y}_i) = \frac{\partial}{\partial \delta} \log f_{\boldsymbol{\theta}}(\mathbf{y}_i). \quad (2.14)$$

Na definição de  $g_3$  poderíamos ter explicitado que  $f_{\boldsymbol{\theta}}(\mathbf{y})$  possui um componente  $\boldsymbol{\alpha}$  com  $\boldsymbol{\theta}$  fixo para  $\tilde{\boldsymbol{\alpha}}$ , mas tal fato está implícito porque a solução para (2.12) envolve três sistemas separados, então quando se soluciona  $g_3 = 0$ ,  $\boldsymbol{\alpha}$  já será conhecido.

A distribuição assintótica do estimador IFM foi estudado por Joe (2005) o qual estabeleceu, com base em em teoria de inferência de funções, que:

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \rightarrow N(\mathbf{0}, V),$$

no qual  $V = B^{-1}A(B^{-1})'$  com

$$A(\boldsymbol{\theta})_{d \times d} = \text{Var}[\Gamma(\boldsymbol{\theta}, \mathbf{y})]$$

e

$$B(\boldsymbol{\theta})_{d \times d} = -\text{E}[\nabla_{\boldsymbol{\theta}} \Gamma(\boldsymbol{\theta}, \mathbf{y})].$$

Joe (2005) mostrou que  $\text{Cov}(\mathbf{g}_j, g_3) = 0$ ,  $j = 1, 2$ , e assim, ao definir  $J_{11}(\boldsymbol{\theta})$  e  $J_{22}(\boldsymbol{\theta})$  como

a Informação esperada de Fisher das estimativas marginas e  $J_{12}(\boldsymbol{\theta}) = \text{Cov}(\mathbf{g}_1, \mathbf{g}_2)$ , temos

$$A = \begin{pmatrix} J_{11} & J_{12} & 0 \\ J'_{12} & J_{22} & 0 \\ 0 & 0 & I_{33} \end{pmatrix}$$

e

$$-B = \begin{pmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & 0 \\ I_{31} & I_{32} & I_{33} \end{pmatrix}.$$

O elemento  $I_{3k}$  é simplesmente a última linha de  $I$  definida na seção anterior 2.3.1.

A estimação de  $V$  é tipicamente feita utilizando o método de jackknife como proposto por Joe (Joe, 2005), sem referência explicitas a  $A$  e  $B$ . O estimador jackknife de  $V/n$  é dado por

$$\widehat{\text{Var}}_{\infty}(\sqrt{n}\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^n (\tilde{\boldsymbol{\theta}}^{(i)} - \tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}}^{(i)} - \tilde{\boldsymbol{\theta}})'$$

no qual  $\tilde{\boldsymbol{\theta}}^{(i)}$  é o estimador de  $\boldsymbol{\theta}$  com  $i$ -ésima observação retirada.

Aqui nós propomos utilizar um método de estimação para a variância sem utilizar um método computacionalmente intensivo.

As submatrizes  $\tilde{J}_{jj} = J_{jj}(\tilde{\boldsymbol{\theta}})$ ,  $j = 1, 2$ , são diretamente obtidas no primeiro estágio das maximizações marginais via algoritmo numérico do tipo quase Newton (BFGS em particular).

A covariância  $J_{12}(\boldsymbol{\theta})$  é estimada de forma simples por

$$\tilde{J}_{12} = \frac{1}{n} \sum_{i=1}^n (\tilde{\mathbf{g}}_{i1} - \tilde{\mathbf{g}}_1)(\tilde{\mathbf{g}}_{i2} - \tilde{\mathbf{g}}_2)',$$

no qual  $\tilde{\mathbf{g}}_{i1} = \mathbf{g}_j(\tilde{\boldsymbol{\alpha}}_j, y_{ij})$ .

A parte mais complexa de calcular  $A$  e  $B$  é estimar  $I$  (ou, mesmo, apenas a sua última linha) para o estimador IFM. Nós propomos usar o método do produto externo do gradiente *outer-*

*product of gradients (OPG)* relacionado ao algoritmo BHHH Berndt et al. (1974). É obtido diretamente pelas equações de estimação de MV com a estimativa do IFM,

$$\tilde{I} = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{\theta}, \mathbf{y}_i) \Psi(\tilde{\theta}, \mathbf{y}_i)'$$

A estimação acima requer o cálculo de algumas derivadas:

$$\mathbf{g}_j = \nabla_{\alpha_j} \log f_j \text{ e } \Psi = \nabla_{\theta} \log f_{\theta}.$$

Em alguns exemplos (próximo capítulo) nós usamos derivadas analíticas e numéricas e os resultados foram bastante similares. Dessa forma, nossa implementação computacional para aplicações em geral foi baseada em derivadas numéricas. O método de Bootstrap é uma forma alternativa para estimação de  $V$ .

## 2.4 Intervalos de Confiança para Confiabilidade

Estudos relacionados a intervalos de confiança para modelos de Força e Tensão com dependência não são tão numerosos, em muitas publicações se tem o objetivo de apresentar a expressão para o cálculo da confiabilidade com diversas particularidades, mas com pouca atenção voltada para questões inferenciais o que, para aspectos práticos, limita a aplicação de modelos de Força e Tensão, pois muitas vezes as pessoas estão mais interessadas em estimações intervalares sobre uma grandeza do que uma simples estimativa pontual.

Muitos dos intervalos de confiança apresentados possuem a hipótese de independência, como por exemplo, Bai et al. (2018a) que considera que as variáveis de Força e Tensão seguem uma distribuição de taxa de risco proporcional truncada independentes e apresenta uma IC baseado em quantidade pivotal. Nguimkeu et al. (2014) apresenta um intervalo de confiança considerando dependência, utilizando uma normal bivariada.

Os trabalhos Domma e Giordano (2012), Domma e Giordano (2013) e Gao et al. (2016) utilizam cópulas para incorporar estrutura de dependência em modelos de Força e Tensão, mas apresentam apenas estimativas pontuais.

Uma recente contribuição dada por Bai et al. (2018b), sem a hipótese de independência, foi apresentar um intervalo de confiança assintótico para o estimador de máxima verossimilhança da confiabilidade considerando que as variáveis do problema são dependentes, com a estrutura de dependência sendo tratada utilizando a cópula Gumbel.

O IC assintótico apresentado em Bai et al. (2018b) é baseado em propriedades dos estimadores de máxima verossimilhança, mas o autor não considera o parâmetro de associação da cópula para construção do intervalo. O mesmo estima o parâmetro de maneira semi-paramétrica e o considera fixo para elaborar o IC.

Nesse contexto, esta seção apresenta intervalos de confiança para modelos de Força e Tensão com cópulas considerando todos os parâmetros do modelo. Utiliza-se tanto o método de máxima verossimilhança completo quanto o método IFM, apresentados na seção anterior.

### 2.4.1 Intervalos de Confiança Assintóticos para Confiabilidade

Ao considerar os resultados apresentados na seção (2.3) e ao definir  $r(\theta) = \nabla_{\theta} R(\theta)$ , a variância para o estimador de  $R$  é dado de forma imediata pelo método delta,

$$\text{Var}(\hat{R}) = n^{-1} r(\hat{\theta})' I(\hat{\theta})^{-1} r(\hat{\theta}), \quad (2.15)$$

$$\text{Var}(\tilde{R}) = n^{-1} r(\tilde{\theta})' V(\tilde{\theta}) r(\tilde{\theta}). \quad (2.16)$$

No qual  $\hat{R}$  e  $\tilde{R}$  representam os estimadores de MV e IFM de  $R$ , respectivamente.

Aqui, novamente, a estimação requer o cálculo de derivadas e, para a implementação computacional, consideramos métodos numéricos.

Considerando as seguintes condições de regularidade para as densidades:

1. Observamos  $Y_1, \dots, Y_n$ , no qual  $Y_i \sim f_{\theta}(y)$  são iid.
2. Os parâmetros são identificáveis; se  $\theta \neq \theta'$ , então  $f_{\theta}(y) \neq f_{\theta'}(y)$ .
3. As densidades  $f_{\theta}(y)$  tem suporte comum, e  $f_{\theta}(y)$  é diferenciável em  $\theta$ .
4. o espaço paramétrico  $\Omega$  contém um conjunto aberto  $\omega$  do qual o valor do parâmetro verdadeiro  $\theta_0$  é um ponto interior.
5. Para cada  $y \in \mathcal{Y}$ , a densidade  $f_{\theta}(y)$  é três vezes diferenciável com relação a  $\theta$ , a terceira deriva é contínua em  $\theta$ , e  $f_{\theta}(y)dy$  pode ser diferenciado três vezes sob sinal integral
6. Para qualquer  $\theta_0 \in \Omega$ , existe um número positivo  $c$  e a função  $M(y)$  (ambos os quais podem depender de  $\theta_0$ ) de modo que

$$\left| \frac{\delta^3}{\delta \theta^3} \log f_{\theta}(y) \right| \leq M(y) \text{ para todo } y \in \mathcal{Y}, \theta_0 - c < \theta < \theta_0 + c,$$

com  $E_{\theta_0}[M(y)] < \infty$ .

Sejam  $Y_1, \dots, Y_n$  iid  $f_\theta(y)$ , ao supor um EMV  $\hat{\theta}$  e que  $\phi(\theta)$  seja uma função contínua de  $\theta$ , temos

$$\sqrt{n}[\phi(\hat{\theta}) - \phi(\theta)] \longrightarrow N(0, \nu(\hat{\theta}))$$

no qual  $\nu(\theta)$  é a variância de  $\phi(\theta)$ .

O resultado apresentado acima é uma propriedade assintótica dos estimadores de máxima verossimilhança e foi retirado de Casella e Berger (2002).

Como  $R$  é uma função contínua dos parâmetros, temos:

$$\sqrt{n}[\hat{R} - R] \longrightarrow N(0, \text{Var}(\hat{R})),$$

$$\sqrt{n}[\tilde{R} - R] \longrightarrow N(0, \text{Var}(\tilde{R})).$$

Assim o intervalo de confiança assintótico para  $R$  considerando o método de máxima verossimilhança completo é dado por:

$$\left[ \hat{R} - z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var} \hat{R}}{n}}, \hat{R} + z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var} \hat{R}}{n}} \right],$$

e para o método IFM:

$$\left[ \tilde{R} - z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var} \tilde{R}}{n}}, \tilde{R} + z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var} \tilde{R}}{n}} \right].$$

no qual  $\text{Var} \hat{R}$  e  $\text{Var} \tilde{R}$  são dados em (2.15) e (2.16), respectivamente,  $z_{\frac{\gamma}{2}}$  representa o quantil  $\frac{\gamma}{2}$  de uma distribuição normal padrão. No próximo capítulo iremos avaliar a performance do IC apresentado por intermédio de simulações.

### 2.4.2 Intervalos de Confiança com a Transformação Logito

Em muitas aplicações, (Barbiero, 2011), é comum a utilização de transformações na variável aleatória a fim de estabilizar a variância do estimador e, assim, obter intervalos de confiança que converjam para o coeficiente de confiança  $(1 - \gamma)$  mais rapidamente. Aqui consideraremos

uma transformação logito ( $g(R) = \log(\frac{R}{1-R})$ ) em  $R$ , e, segundo método delta, temos:

$$\text{Var}(\text{logito}(\hat{R})) = \frac{1}{(\hat{R}(1-\hat{R}))^2} \text{Var}(\hat{R}), \quad (2.17)$$

$$\text{Var}(\text{logito}(\tilde{R})) = \frac{1}{(\tilde{R}(1-\tilde{R}))^2} \text{Var}(\tilde{R}). \quad (2.18)$$

Dessa forma, levando em consideração propriedades assintóticas dos estimadores de máxima verossimilhança, temos o intervalo de confiança:

$$\left[ \frac{e^L}{e^L + 1}, \frac{e^U}{e^U + 1} \right],$$

no qual  $L$  é o limite inferior e  $U$  o limite superior do seguinte intervalo de confiança para o método MV

$$\log\left(\frac{\hat{R}}{1-\hat{R}}\right) \pm z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var}(\hat{R})}{n(\hat{R}(1-\hat{R}))^2}}$$

e para o método IFM

$$\log\left(\frac{\tilde{R}}{1-\tilde{R}}\right) \pm z_{\frac{\gamma}{2}} \sqrt{\frac{\text{Var}(\tilde{R})}{n(\tilde{R}(1-\tilde{R}))^2}}.$$

### 2.4.3 Intervalos de Confiança com Distribuição Beta

Supor que o estimador amostral de  $R$ , uma probabilidade ( $P(X < Y)$ ), segue distribuição assintótica normal talvez seja inapropriado e pode apresentar resultados em que os limites do intervalo de confiança sejam menores que 0 ou maiores que 1, dado que o suporte da distribuição normal é a reta real  $\mathbb{R}$ .

Considerar que a distribuição assintótica do estimador de  $R$  possui suporte limitado em 0 e 1 parece mais adequado e, dessa forma, iremos fazer um exercício considerando que o estimador de  $R$  segue distribuição Beta, tanto para o método de MV completo quanto para o IFM.

Suponha que o estimador  $\hat{R}$  tenha distribuição amostral Beta com média  $\hat{R}$  e variância

$\text{Var}(\widehat{R})$ . Distribuição Beta com densidade dada por:

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1}, \quad 0 \leq y \leq 1, \quad \alpha > 0, \quad \beta > 0,$$

com esperança

$$E[y] = \frac{\alpha}{\alpha + \beta}$$

e variância

$$\text{Var}[y] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Então, utilizando método de momentos (Casella e Berger, 2002), temos

$$\begin{aligned} \widehat{R} &= \frac{\alpha}{\alpha + \beta} \\ \frac{1}{\widehat{R}} &= 1 + \frac{\beta}{\alpha} \\ \frac{1 - \widehat{R}}{\widehat{R}} &= \frac{\beta}{\alpha} \\ c\alpha &= \beta \end{aligned}$$

tal que  $c = \frac{1 - \widehat{R}}{\widehat{R}}$  e

$$\text{Var}(\widehat{R}) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)},$$

ou ainda,

$$\text{Var}(\widehat{R}) = \frac{c\alpha^2}{(\alpha + c\alpha)^2(\alpha + c\alpha + 1)}. \quad (2.19)$$

Resolvendo a expressão (2.19) calcula-se  $\alpha$  (pode-se utilizar métodos numéricos) e, consequentemente,  $\beta$ . A construção é a mesma para o estimador IFM ( $\tilde{R}$ ).

Com  $\widehat{\alpha}$  e  $\widehat{\beta}$  estimados de  $\widehat{R}$ , consideramos os quantis  $\frac{\gamma}{2}$  e  $1 - \frac{\gamma}{2}$  de uma distribuição  $\text{Beta}(\widehat{\alpha}, \widehat{\beta})$

como os limites inferior e superior, respectivamente, de um intervalo de confiança  $(1 - \gamma)$  de  $\hat{R}$ . Resultado análogo para  $\tilde{R}$ .

#### 2.4.4 Intervalos de Confiança Bootstrap

O método de Bootstrap é uma alternativa para construir um intervalo de confiança para  $R$ . Com uma amostra aleatória de tamanho  $n$  de  $Y_1 (y_{11}, \dots, y_{1n})$  iid e  $Y_2 (y_{21}, \dots, y_{2n})$  iid, podemos implementar o método com os seguintes passos

1. estimar os parâmetros por máxima verossimilhança.
2. calcular  $R$  com a expressão utilizando os parâmetros estimados em 1.
3. gerar amostras aleatórias *bootstrap* de  $Y_1$  e  $Y_2$ , fazendo uso de reamostragem com reposição.
4. estimar os parâmetros por máxima verossimilhança das amostras geradas em 3.
5. estimar  $R_b$  (confiabilidade da amostra bootstrap  $b$ ) com os parâmetros de 4.
6. repetir os passos 3 a 4  $B$  vezes para obter a distribuição *bootstrap* de  $R_B$ .
7. estimar intervalo de confiança  $(1 - \gamma)$  *bootstrap* de  $R$  baseado nos percentis da amostra  $R_b$ .

# Capítulo 3

## Estudo de Simulação

Este capítulo apresenta resultados de exercícios de simulação realizados com base na teoria apresentada no capítulo anterior. O objetivo é estudar a eficiência assintótica dos métodos MV completo e IFM para estimação da confiabilidade, além dos parâmetros das marginais e de dependência, e avaliar a cobertura e tamanho dos intervalos de confiança apresentados para amostras grandes.

### 3.1 Introdução

O estudo de simulação foi delimitado para modelos de Força e Tensão com cópulas e envolve os seguintes passos:

1. Estabelecer os parâmetros das distribuições marginais e parâmetro de dependência.
2. Calcular  $R$ .
3. Simular  $n$  dados da cópula e marginais consideradas com os parâmetros expressos em 1.
4. Estimar, para os métodos considerados, os parâmetros das marginais e de dependência com a amostra de 3.
5. Calcular  $R$  com os parâmetros estimados em 4, para cada método considerado.

6. Estimar os intervalos de confiança apresentados no capítulo 2 para  $R$  obtido em 5.
7. Repetir os passos 3 a 6  $m$  vezes (número de simulações).

Consideramos  $n = 200$  (tamanho da amostra) e  $m = 500$  (número de simulações) para as simulações desse capítulo. A eficiência dos métodos será avaliada segundo erro quadrático médio (EQM) observado relativo. Ao denotar  $m$  como o número de simulações,  $\theta^{(0)}$  o valor verdadeiro do parâmetro,  $\theta^{(i)}$  a estimativa de  $\theta^{(0)}$  da  $i$ -ésima simulação, o EMQ observado é dada por  $EQM = \frac{\sum_{i=1}^m [\theta^{(i)} - \theta^{(0)}]^2}{m}$  e o EQM observado relativo ao método MV é dado por  $\frac{EQM \text{ observado para IFM}}{EQM \text{ observado para MV}}$ . Dado um intervalo de confiança da  $i$ -ésima simulação  $(L^{(i)}, U^{(i)})$ , o tamanho médio intervalo é dado por  $\frac{\sum_{i=1}^m U^{(i)} - L^{(i)}}{m}$  e a cobertura  $\frac{\sum_{i=1}^m I[\theta^{(0)} \in (L^{(i)}, U^{(i)})]}{m}$ .

### 3.2 Simulações para Marginais Normais

Para os resultados dessa etapa, foram considerados modelos de Força e Tensão com marginais normais,  $Y_1 \sim N(0, 1)$  e  $Y_2 \sim N(1, 1)$ , e as cópulas Frank, Gumbel e Clayton com diferentes valores para os parâmetros de dependência.

**Tabela 3.1:** Parâmetros originais para realização das simulações para marginais normais.

Parâmetros	Cópulas						
	Frank		Gumbel		Clayton		
$\delta$	10.00	2.00	-10.00	5.00	2.00	10.00	2.00
R	0.95	0.81	0.69	0.99	0.91	0.99	0.90
$\mu_{y_1} = 0, \sigma_{y_1} = 1, \mu_{y_2} = 1, \sigma_{y_2} = 1.$							

Os resultados, para EQM observado relativo, mostraram que o método de MV completo apresentou EQM menor para grande maioria das estimativas dos parâmetros marginais. Para o parâmetro de dependência, o método IFM apresentou EQM menor para a cópula Frank e para um caso da cópula Clayton, o que não se configurou para os outros casos do exercício. Em relação a confiabilidade, o método MV completo se mostrou mais eficiente em todos os casos testados, mas os resultados sugerem que os métodos possuem resultados similares quando o

parâmetro de dependência estabelece uma dependência mais fraca e, para cópula Clayton com dependência forte, o EQM do método IFM foi muito maior que o do método MV completo.

**Tabela 3.2:** Estimativas médias dos parâmetros para o método IFM.

Cópulas	$\mu_{y_1}$	$\sigma_{y_1}$	$\mu_{y_2}$	$\sigma_{y_2}$	$\delta$	R
	0.00	1.00	1.00	1.00	10.01	0.95
Frank	0.00	1.00	1.00	0.99	2.01	0.81
	0.00	1.00	1.00	0.99	-10.03	0.69
Gumbel	0.00	1.00	1.00	1.00	5.00	0.99
	0.00	1.00	1.00	1.00	2.01	0.91
Clayton	0.00	1.00	1.00	0.99	9.79	0.99
	0.00	1.00	1.00	1.00	2.00	0.91

**Tabela 3.3:** Estimativas médias dos parâmetros para o método MV.

Cópulas	$\mu_{y_1}$	$\sigma_{y_1}$	$\mu_{y_2}$	$\sigma_{y_2}$	$\delta$	R
	0.00	1.00	1.00	1.00	10.07	0.95
Frank	0.00	1.00	1.00	0.99	2.01	0.81
	0.00	1.00	1.00	0.99	-10.08	0.69
Gumbel	0.00	1.00	1.00	1.00	5.02	0.99
	0.00	0.99	1.00	1.00	2.01	0.91
Clayton	0.00	1.00	1.00	1.00	10.08	0.99
	0.00	1.00	1.00	1.00	2.02	0.91

Para a mesma cópula, a amplitude média dos intervalos tende a ficar menor a medida que a dependência aumenta e quase não apresentou diferença entre os intervalos de confiança considerados. Esse resultado pode ser influenciado pelo fato de R estar mais perto de 1 para dependências mais fortes com os parâmetros marginais considerados. Para todos os casos explorados, os intervalos de confiança para o método de estimação MV completo apresentaram amplitude média menor em relação ao método de estimação IFM. Para o mesmo método de estimação, os intervalos de confiança não mostram grande diferença, mas em geral o IC beta apresentou tamanho médio menor.

**Tabela 3.4:** Resultados das simulações para  $EQM_{IFM}/EQM_{MV}$ .

Cópuas	$\delta$	$\tau(\delta)$	$\mu_{y_1}$	$\sigma_{y_1}$	$\mu_{y_2}$	$\sigma_{y_2}$	$\delta$	R
Frank	10	0.66	1.37	1.09	1.37	1.14	0.99	1.15
	2	0.21	1.03	1.03	1.03	1.01	1	1.01
	-10	-0.66	1.28	1.12	1.29	1.07	0.99	1.28
Gumbel	5	0.8	1.01	1.08	1	1.05	1.01	1.14
	2	0.5	1.01	1.09	1.02	1.11	1.03	1.04
Clayton	10	0.83	1.10	1.32	1.07	1.33	1.03	2.57
	2	0.5	1.03	1.18	1	1.20	0.98	1.15

Depreende-se dos resultados da cobertura que todos os métodos considerados obtiveram valores perto do nível de confiança  $(1 - \gamma)$ . Para a mesma cópula, método de estimação e parâmetro de dependência, os IC's mostram diferença na terceira casa decimal. O método de estimação IFM apresentou IC's com maior cobertura e os resultados sugerem que as cópulas e grau de dependência alteram a cobertura do intervalo.

**Tabela 3.6:** Resultados das simulações para cobertura dos IC's considerados para marginais normais.

IC		Frank		Gumbel		Clayton		
$\delta$	10.00	2.00	-10.00	5.00	2.00	10.00	2.00	
$\tau(\delta)$	0.66	0.21	-0.66	0.80	0.50	0.83	0.50	
IFM	Normal	0.976	0.946	0.932	0.944	0.970	0.954	0.960
	Logit	0.988	0.954	0.934	0.968	0.982	0.944	0.954
	Beta	0.980	0.950	0.934	0.958	0.982	0.946	0.958
MV	Normal	0.952	0.942	0.932	0.916	0.956	0.952	0.944
	Logit	0.944	0.946	0.932	0.934	0.964	0.946	0.950
	Beta	0.942	0.944	0.934	0.926	0.962	0.948	0.946

**Tabela 3.5:** Resultados das simulações para amplitudes dos IC's considerados para marginais normais.

IC		Frank		Gumbel		Clayton		
$\delta$	10.00	2.00	-10.00	5.00	2.00	10.00	2.00	
$\tau(\delta)$	0.66	0.21	-0.66	0.80	0.50	0.83	0.50	
IFM	Normal	0.0383	0.0938	0.0973	0.0085	0.0677	0.0145	0.0637
	Logit	0.0391	0.0940	0.0971	0.0094	0.0687	0.0167	0.0644
	Beta	0.0381	0.0937	0.0972	0.0084	0.0675	0.0143	0.0635
MV	Normal	0.0316	0.0906	0.0870	0.0067	0.0591	0.0089	0.0561
	Logit	0.0321	0.0908	0.0869	0.0071	0.0598	0.0102	0.0566
	Beta	0.0315	0.0905	0.0870	0.0066	0.0590	0.0089	0.0560

### 3.3 Simulações para Marginais Gamma

Para os resultados dessa etapa, foram considerados modelos de Força e Tensão com marginais normais,  $Y_1 \sim Gamma(2, 1)$  e  $Y_2 \sim Gamma(1, 1)$ , e as cópulas Frank, Gumbel e Clayton com diferentes valores para os parâmetros de dependência.

**Tabela 3.7:** Parâmetros originais das simulações para marginais gamma.

Parâmetros	Cópulas						
	Frank		Gumbel		Clayton		
$\delta$	10.00	2.00	-10.00	5.00	2.00	10.00	2.00
R	0.054	0.201	0.320	0.004	0.096	0.024	0.110

$\alpha_{y_1} = 2.00, \beta_{y_1} = 1.00, \alpha_{y_2} = 1.00, \beta_{y_2} = 1.00.$

Assim como o caso de marginais normais, o método de MV completo apresentou EQM menor para grande maioria das estimativas dos parâmetros marginais com apenas dois valores maiores para cópula Frank. Para o parâmetro de dependência, o método IFM apresentou EQM menor para a cópula Frank. Em relação a confiabilidade, o método MV completo se mostrou mais eficiente em todos os casos testados, a menos de um caso para cópula Frank, e os resultados

sugerem, mais uma vez, que os métodos possuem resultados similares quando o parâmetro de dependência estabelece uma dependência mais fraca e, para cópula Clayton com dependência forte, também o EQM do método IFM foi muito maior que o do método MV completo. Os resultados do EQM relativo se mostraram parecidos para marginais normais e gamma.

**Tabela 3.8:** Estimativas médias dos parâmetros para o método IFM.

Cópulas	$\alpha_{y_1}$	$\beta_{y_1}$	$\alpha_{y_2}$	$\beta_{y_2}$	$\delta$	R
	2.03	1.01	1.01	1.01	10.00	0.05
Frank	2.04	1.02	1.02	1.02	1.99	0.20
	2.03	1.01	1.01	1.01	-10.06	0.32
Gumbel	2.04	1.02	1.02	1.03	4.96	0.00
	2.03	1.02	1.01	1.01	2.00	0.10
Clayton	2.04	1.02	1.02	1.02	9.76	0.02
	2.03	1.02	1.01	1.02	2.01	0.11

**Tabela 3.9:** Estimativas médias dos parâmetros para o método MV.

Cópulas	$\alpha_{y_1}$	$\beta_{y_1}$	$\alpha_{y_2}$	$\beta_{y_2}$	$\delta$	R
	2.03	1.01	1.01	1.01	10.06	0.05
Frank	2.04	1.02	1.02	1.02	1.99	0.20
	2.03	1.01	1.01	1.01	-10.12	0.32
Gumbel	2.04	1.02	1.02	1.03	4.98	0.00
	2.03	1.01	1.01	1.01	2.01	0.10
Clayton	2.04	1.02	1.02	1.02	10.01	0.02
	2.03	1.02	1.01	1.02	2.03	0.11

Os resultados para amplitude média foram similares aos apresentados para marginais normais. Para todos os casos explorados, os intervalos de confiança para o método de estimação MV completo apresentaram amplitude média menor em relação ao método de estimação IFM.

**Tabela 3.10:** Resultados das simulações para  $EQM_{IFM}/EQM_{MV}$ .

Cópuas	$\delta$	$\tau(\delta)$	$\alpha_{y_1}$	$\beta_{y_1}$	$\alpha_{y_2}$	$\beta_{y_2}$	$\delta$	R
	10	0.66	1.13	1.12	1.16	1.08	0.98	1.23
Frank	2	0.21	0.99	0.99	1.01	1.01	1	1
	-10	-0.66	1.09	1.06	1.08	1.09	0.96	1.19
Gumbel	5	0.8	1.02	1.03	1.06	1.12	1.03	1.12
	2	0.5	1.06	1.06	1.05	1.06	1.02	1.03
Clayton	10	0.83	1.18	1.35	1.12	1.31	1.05	2.15
	2	0.5	1.10	1.16	1.14	1.17	1	1.06

Assim como para o caso de marginais normais, os resultados da cobertura de todos os métodos considerados obtiveram valores em torno do nível de confiança  $(1 - \gamma)$ . Para a mesma cópula, método de estimação e parâmetro de dependência, os IC's mostram diferença na terceira casa decimal. O método de estimação IFM apresentou IC's com maior cobertura e os resultados sugerem que o as cópulas e grau de dependência alteram a cobertura do intervalo.

**Tabela 3.11:** Resultados das simulações para amplitudes dos IC's considerados para marginais gamma.

IC	Frank			Gumbel		Clayton		
$\delta$	10.00	2.00	-10.00	5.00	2.00	10.00	2.00	
$\tau(\delta)$	0.66	0.21	-0.66	0.80	0.50	0.83	0.50	
IFM	Normal	0.0440	0.0958	0.0958	0.0072	0.0718	0.0225	0.0678
	Logit	0.0449	0.0959	0.0955	0.0081	0.0728	0.0232	0.0684
	Beta	0.0438	0.0957	0.0957	0.0071	0.0716	0.0224	0.0677
MV	Normal	0.0363	0.0926	0.0859	0.0057	0.0613	0.0152	0.0625
	Logit	0.0369	0.0926	0.0857	0.0061	0.0619	0.0155	0.0629
	Beta	0.0362	0.0924	0.0858	0.0056	0.0611	0.0152	0.0623

**Tabela 3.12:** Resultados das simulações para cobertura dos IC's considerados para marginais gamma.

IC	Frank			Gumbel		Clayton		
$\delta$	10.000	2.000	-10.000	5.000	2.000	10.000	2.000	
$\tau(\delta)$	0.660	0.210	-0.66	0.800	0.500	0.830	0.500	
IFM	Normal	0.950	0.950	0.948	0.948	0.976	0.958	0.946
	Logit	0.962	0.952	0.952	0.974	0.982	0.954	0.960
	Beta	0.958	0.952	0.952	0.964	0.978	0.954	0.950
MV	Normal	0.934	0.944	0.954	0.910	0.950	0.926	0.930
	Logit	0.952	0.948	0.954	0.940	0.950	0.946	0.942
	Beta	0.938	0.946	0.956	0.926	0.958	0.940	0.926

## Capítulo 4

# Modelos de Força e Tensão como Medida de Fragilidade Financeira

### 4.1 Introdução

O trabalho presente neste capítulo se baseou no estudo de Domma e Giordano (2012), o qual utilizou um modelo de Força e Tensão com variáveis dependentes como medida de fragilidade financeira, utilizando dados de despesa e renda de família Italianas.

Após a crise financeira de 2009, aumentaram-se as preocupações em relação a manutenção da solidez do sistema financeiro. A fragilidade financeira é um conceito importante desse contexto tratado por Minsky et al. (1986), antes mesmo da crise ocorrer e com base em eventos econômicos passados. O autor analisa a propensão a crises financeiras avaliando alguns sinais, como por exemplo, a capacidade das entidades econômicas de honrar suas dívidas com seu caixa operacional. O presente capítulo tem como objetivo calcular uma medida de fragilidade financeira sugerida por Domma e Giordano (2012).

Minsky (1992) apresenta as hipóteses de instabilidade financeira e identifica que, após períodos prolongados de bons tempos econômicos com crescimentos estáveis, as economias capi-

talistas possuem uma tendência a migrar de uma estrutura financeira baseada em *hedge*<sup>1</sup> para uma estrutura baseada em especulação.

Eichengreen e Hausmann (1999) identificam três relações entre fragilidade financeira e taxas de câmbio: a relação entre risco moral<sup>2</sup> e a utilização de taxas de câmbio indexadas; a hipótese do pecado original<sup>3</sup> (*Original Sin Hypothesis*) e a necessidade de não possuir taxa de câmbio (dolarização); a ótica de problemas de compromisso<sup>4</sup> e a utilização de taxas de câmbio flexíveis. Goodhart et al. (2006) apresentam um modelo para analisar a fragilidade financeira e determinar requisitos e incentivos adequados para limitar a assunção de riscos por parte dos bancos. Esses estudos avaliam a fragilidade financeira em níveis macroeconomicos e, recentemente, se teve pesquisas mensurando a fragilidade financeira no contexto familiar, domicílios potencialmente mais vulneráveis a mudanças desfavoráveis no ambiente econômico.

Brown e Taylor (2008) identificam a probabilidade das famílias terem patrimônio negativo como medida de fragilidade financeira. Jappelli et al. (2008) avaliam a sensibilidade das dívidas familiares a choques adversos na economia, como desemprego e variações nas taxas de juros. Lusardi et al. (2011) medem a fragilidade financeira avaliando a capacidade da família saldar uma despesa inesperada de 2000 dólares em um mês, independente da fonte de recursos. A medida proposta por Domma e Giordano (2012) busca estimar a propensão dos gastos familiares excederem os recursos financeiros anuais. A ideia por trás da medida é identificar um modelo de Força e Tensão na gestão orçamentaria familiar.

O modelo identifica como variável aleatória de Força ( $Y_2$ ) a renda anual das famílias a qual é sujeita a uma variável de Tensão, a despesa anual ( $Y_1$ ). Se a Tensão exceder a Força, o componente irá falhar, ou seja, se os gastos anuais excederem a renda, a família passará por um problema financeiro que trará a necessidade de se endividar ou se desfazer de bens e, assim,

---

<sup>1</sup>Estrutura financeira de cobertura a qual os agentes econômicos podem cumprir todas as suas obrigações contratuais de pagamento com seus fluxos de caixa.

<sup>2</sup>Investidores assumem altos níveis de risco com base em garantias implícitas (p. ex. possibilidade de governos resgatarem participantes do mercado financeiro doméstico).

<sup>3</sup>Situação em que a moeda nacional não pode ser utilizada para empréstimos no exterior ou emprestar a longo prazo.

<sup>4</sup>Incapacidade das instituições financeiras lidarem com seus compromissos.

se tornar mais vulnerável. Vale ressaltar que nesse contexto estamos interessados na medida  $P(Y_1 > Y_2)$  e não em  $P(Y_1 < Y_2)$ , o que se caracteriza apenas como uma simples alteração de nomenclatura.

Os dados analisados nesta etapa tiveram como fonte a Pesquisa de Orçamentos Familiares - POF que visa principalmente mensurar as estruturas de despesa dos gastos, dos rendimentos e parte da variação patrimonial das famílias brasileiras. A POF é uma pesquisa realizada por amostragem, na qual são investigados os domicílios particulares permanentes. A pesquisa teve duração de 12 meses com início em 19 de maio de 2008 e término no dia 18 de maio de 2009. Apesar da POF ser de nível nacional, o presente trabalho analisou as famílias residentes nos estados da região sudeste do Brasil (São Paulo, Rio de Janeiro, Minas Gerais e Espírito Santo).

Considerou-se como observação a unidade de consumo que compreende um único morador ou conjunto de moradores que compartilham da mesma fonte de alimentação. Para variável renda, considerou-se o somatório dos rendimentos brutos monetários dos componentes das unidades de consumo, exclusive os empregados domésticos e seus parentes, acrescido do total dos rendimentos não monetários das unidades de consumo.

Considerou-se como rendimento não monetário a parcela equivalente às despesas não-monetárias que são aquelas efetuadas sem pagamento monetário, ou seja, aquisição obtida através de doação, retirada do negócio, troca, produção própria, pescado, caçado e coletado durante o período de referência da pesquisa, disponíveis para utilização.

Para a variável consumo/despesa, considerou-se gastos anualizados com: alimentação, habitação, vestuário, transporte, higiene e cuidados pessoais, assistência à saúde, educação, recreação e cultura, fumo, serviços pessoais e, ainda, despesas diversas como cerimônias e festas por exemplo.

Seguindo os passos expressos em Joe (2014) para ajustar um modelo de dependência com cópulas, como primeira etapa, identificamos candidatos para os modelos univariados.

## 4.2 Distribuições Marginais

Na economia muitas distribuições são consideradas como modelos para renda, entre elas pode-se citar distribuições conhecidas como a Lognormal, Pareto e Gamma. Na década de 70 foram apresentadas duas distribuições, Dagum (Dagum, 1977) e Singh-Maddala (Singh e Maddala, 1976), que ganharam grande relevância na área por capturarem tanto a cauda pesada observada em distribuições empíricas de renda, como a grande densidade presente em níveis mais baixos dessa variável.

McDonald (1984) identifica que essas duas distribuições amplamente utilizadas podem ser representadas como casos especiais da distribuição beta generalizada do segundo tipo (*GB2*). Kleiber (1996) faz um comparação dessas duas distribuições e expõe uma relação entre elas e Bandourian et al. (2002) fazem uma extensa comparação entre modelos paramétricos para renda.

Dagum (1977) teve como motivação para construção do seu modelo a observação empírica da elasticidade da renda sob sua função de distribuição acumulada (Kleiber, 2008) e, como resultado, permite uma interseção com a curva de Lorenz (Gastwirth, 1971), fenômeno observado empiricamente em distribuições de renda. Bandourian et al. (2002) mostra que a distribuição Dagum obteve bons ajustes, em comparação com outras distribuições, em 82 bases de dados de renda de vários países.

A função de densidade da Dagum é dada por

$$f_{\theta}(y) = \beta\lambda\alpha y^{-\alpha-1}(1 + \lambda y^{-\beta-1}) \quad (4.1)$$

e sua função de distribuição dada por

$$F_{\theta}(y) = (1 + \lambda y^{-\alpha})^{-\beta}, \quad (4.2)$$

no qual  $y \in \mathbb{R}^+$  e  $\theta = (\beta, \lambda, \alpha)$  positivo.  $\beta$  e  $\alpha$  são parâmetros de forma e  $\lambda$  de escala.

Dada a origem da distribuição Dagum, seus bons resultados em termos de qualidade de ajuste (em estudos com bases de dados de renda) e alta aplicabilidade, a consideramos como modelo para as distribuições marginais do modelo de Força e Tensão desse capítulo, ou seja, consideramos que tanto a renda como a despesa das famílias seguem distribuição Dagum, preservando a avaliação da qualidade do ajuste.

As estatísticas descritivas da renda e consumo das famílias, por estado residente, da amostra retirada da POF são exibidas na tabela (4.1) e indicam assimetria à direita, característica presente em dados de renda.

**Tabela 4.1:** Estatísticas descritivas da renda ( $\times 10^{-3}$ ) e consumo ( $\times 10^{-3}$ ) anual das famílias dos estados da região sudeste do Brasil (R\$), 2008-2009.

Estatísticas	São Paulo		Rio de Janeiro		Minas Gerais		Espírito Santo	
	Renda	Despesa	Renda	Despesa	Renda	Despesa	Renda	Despesa
n	3619	3619	1926	1926	5014	5014	3435	3435
Mínimo	1.7	0.5	0.6	0.4	0.9	0.3	0.0	0.3
Máximo	949.8	1171.3	550.6	485.3	1406.6	712.5	1143.7	1110.1
Mediana	25.8	18.0	20.6	14.4	18.7	13.6	17.3	12.1
Média	39.9	30.6	36.3	26.7	31.0	24.1	28.8	22.2
Desvio Padrão	48.3	44.7	53.7	41.1	53.2	41.7	44.8	40.3
Assimetria	5.5	8.5	4.9	5.1	11.7	7.4	9.5	12.7
Curtose	57.8	147.4	31.3	35.3	215.6	78.8	164.4	266.7

Ao considerarmos os conceitos apresentados no capítulo 2 deste trabalho, estimar os parâmetros das marginais por MV, para os dados do estudo, envolve a solução das equações de estimação dadas em (2.11). Kleiber (2008) apresenta essas equações para a distribuição Dagum e necessita a utilização de métodos numéricos para solução. Por questões de implementação computacional, consideramos a utilização de métodos numéricos para maximizar a log-verossimilhança dada por:

$$\ell_{\theta}(y) = n \log(\beta) + n \log(\lambda) + n \log(\alpha) - (\alpha + 1) \sum_{i=1}^n \log(y_i) - (\beta + 1) \sum_{i=1}^n \log(1 + \lambda y_i^{-\alpha}) \quad (4.3)$$

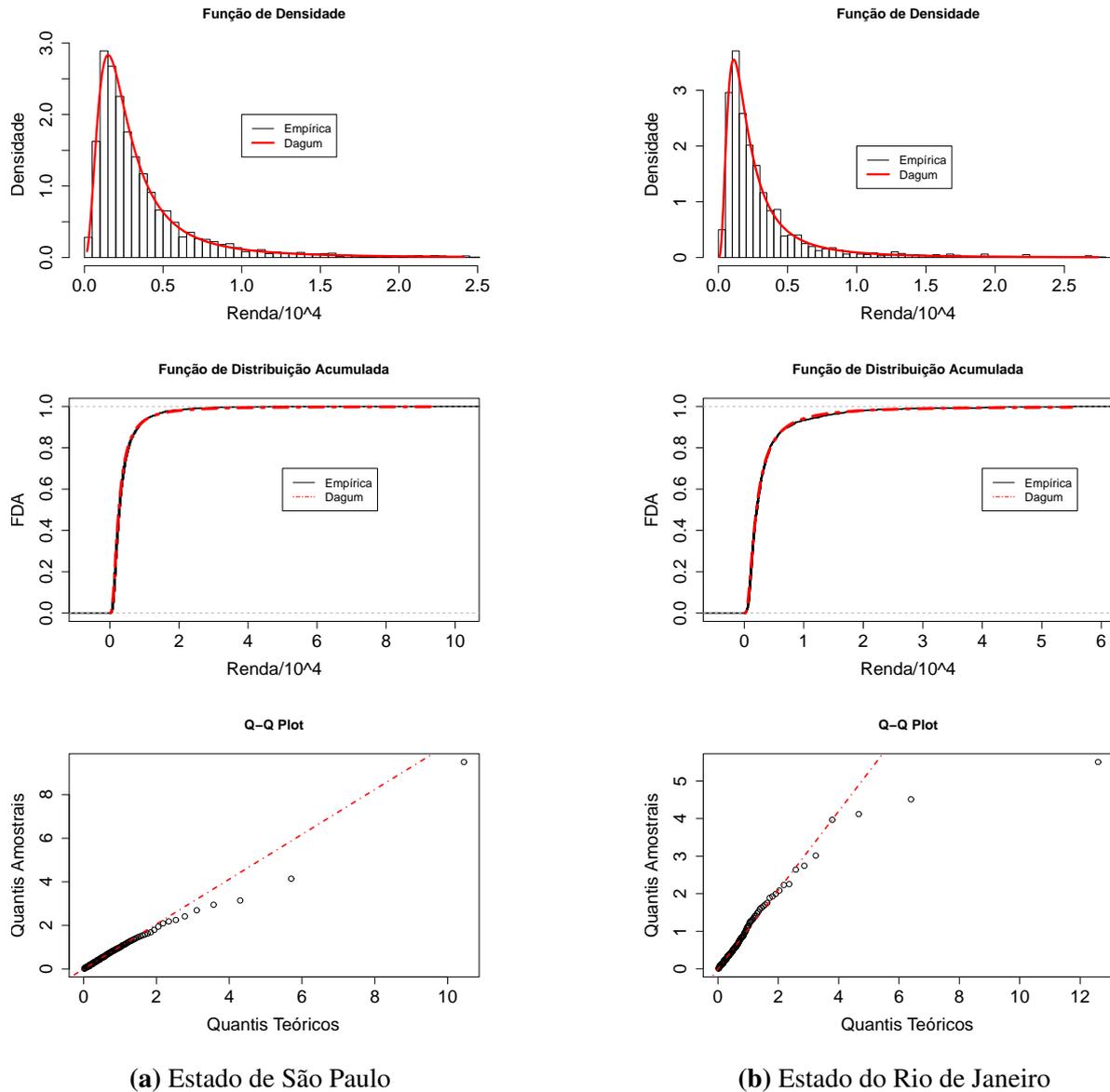
A equação (4.3) foi maximizada em relação a  $\theta = (\beta, \lambda, \alpha)$  via algoritmo numérico do tipo quase Newton (BFGS) para cada amostra de renda e consumo de cada estado considerado. Como valores iniciais, utilizou-se várias amostras arbitrárias e identificou-se convergência.

As estimativas ( $\hat{\theta} = (\hat{\beta}, \hat{\lambda}, \hat{\alpha})$ ), bem como seus erros padrões, estão expressos na tabela (4.2). O teste de Anderson-Darling, que possui como hipótese nula: a amostra tem distribuição Dagum, não rejeita a hipótese nula para nenhuma marginal considerando um nível de significância de 1% (dado o tamanho das amostras) indicando que a distribuição se ajustou bem aos dados. Vale ressaltar que para as fases de estimação as variáveis foram escalonadas ( $\times 10^{-4}$ ).

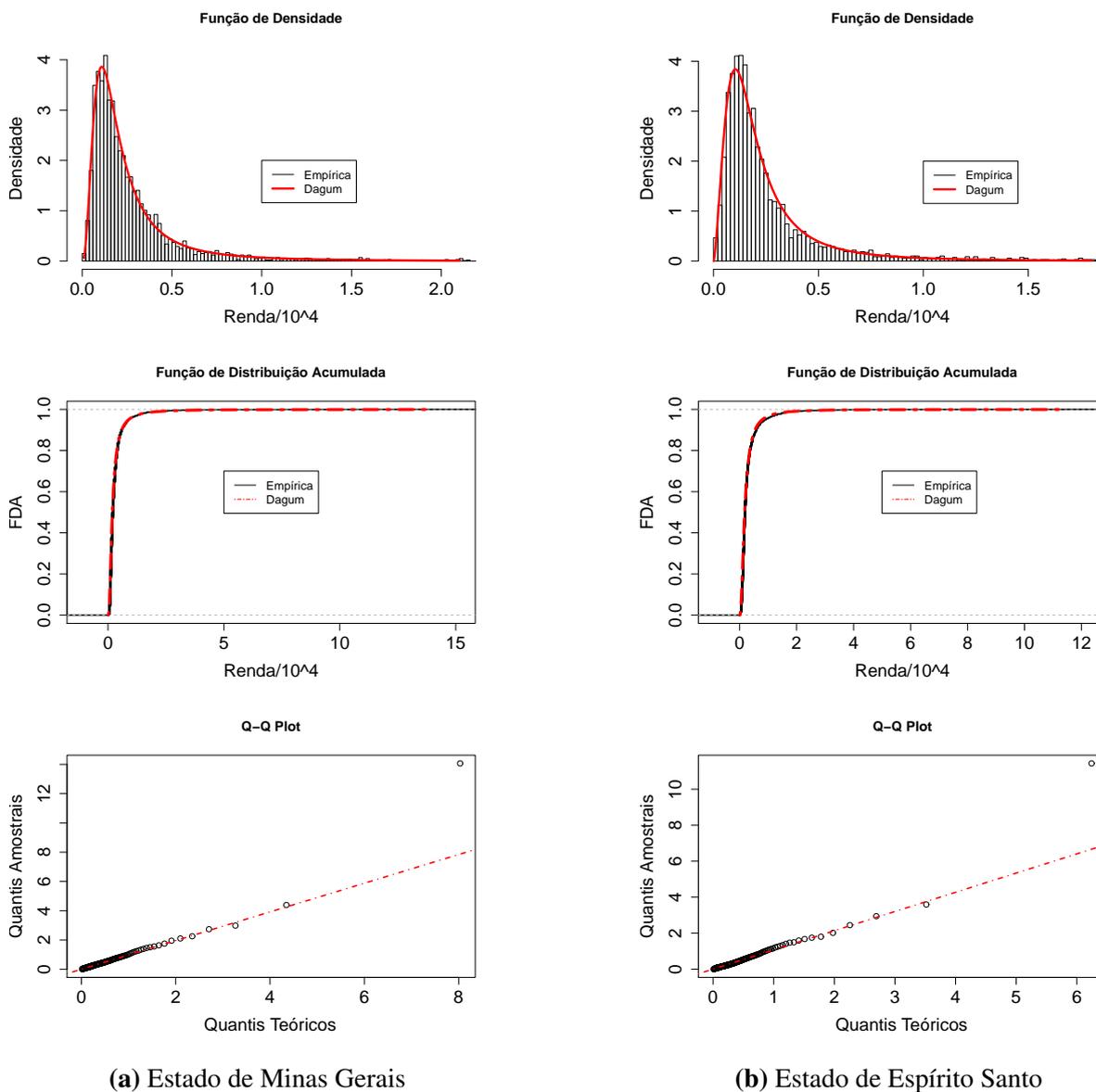
**Tabela 4.2:** Estimativas de Maxima Verossimilhança, Erros Padrão, Log Verossimilhança, AD teste(p-valor) da renda e consumo por estado.

$\hat{\theta}$	São Paulo	Rio de Janeiro	Minas Gerais	Espírito Santo
Renda				
$\hat{\beta}$ (EP)	1.86 (0.031)	2.37 (0.054)	1.89 (0.027)	1.27 (0.022)
$\hat{\lambda}$ (EP)	0.04 (0.001)	0.03 (0.001)	0.02 (0.0004)	0.03 (0.001)
$\hat{\alpha}$ (EP)	1.82 (0.013)	1.62 (0.014)	1.79 (0.01)	1.92 (0.013)
log-verossimilhança	29	407.9	1884.3	1421.9
AD-Teste (p-valor)	0.74 (0.526)	0.86 (0.439)	0.59 (0.66)	3.4 (0.017)
Despesa				
$\hat{\beta}$ (EP)	1.57 (0.026)	1.81 (0.041)	1.48 (0.021)	1.98 (0.034)
$\hat{\lambda}$ (EP)	0.03 (0.001)	0.02 (0.001)	0.02 (0.0004)	0.02 (0.0003)
$\hat{\alpha}$ (EP)	1.66 (0.011)	1.60 (0.014)	1.65 (0.009)	1.57 (0.009)
log-verossimilhança	1129.3	947.7	2944.9	2319.7
AD-Teste (p-valor)	1.53 (0.17)	0.53 (0.714)	0.58 (0.67)	1.21 (0.263)

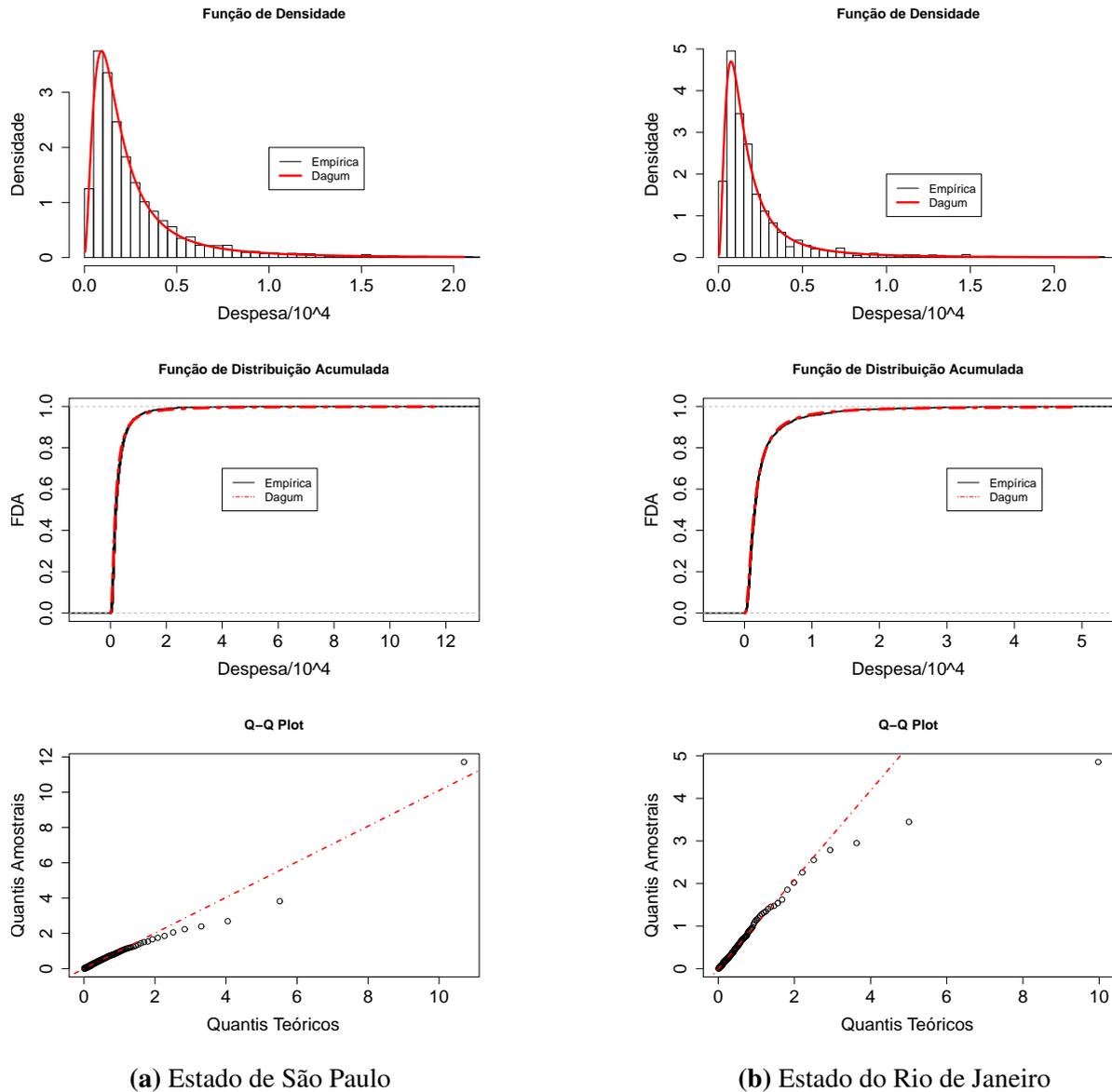
Para ilustrar a qualidade dos ajustes, construímos gráficos (figuras (4.1), (4.2), (4.3), (4.4)) que comparam a densidade, distribuição acumulada e quantis dos valores da distribuição Dagum (parâmetros estimados por MV) com valores empíricos. Os mesmos sugerem que a distribuição Dagum se ajustou bem aos dados do problema a menos de observações discrepantes, o que não foi tratado em um primeiro momento para preservar todas as informações da amostra, mas pode ser avaliado em estudos futuros.



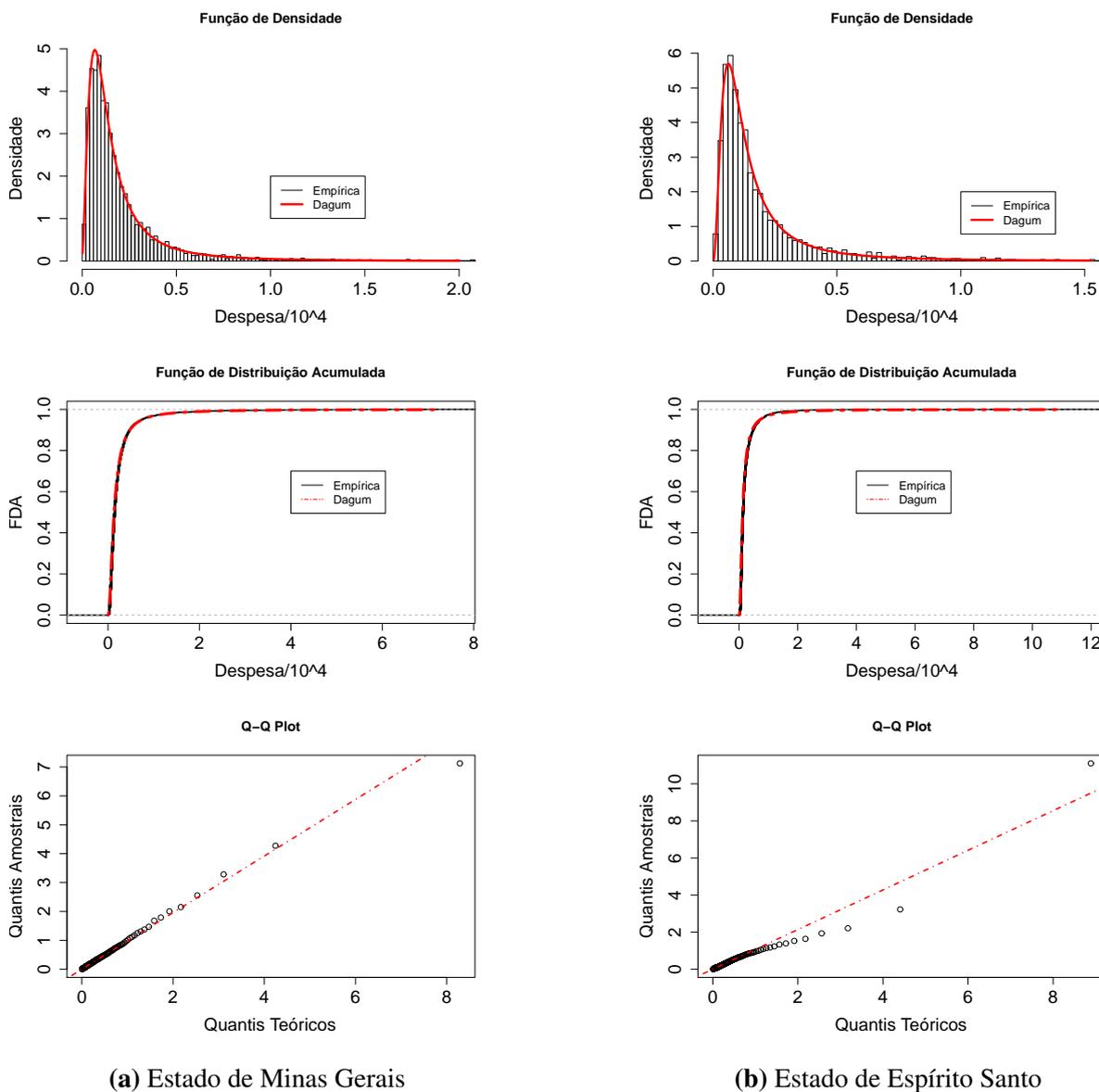
**Figura 4.1:** Gráficos de Qualidade de Ajuste para a renda dos estados de São Paulo e Rio de Janeiro (função de densidade empírica e ajustada, função de distribuição empírica e ajustada e QQ-plot).



**Figura 4.2:** Gráficos de Qualidade de Ajuste para a renda dos estados de Minas Gerais e Espírito Santo (função de densidade empírica e ajustada, função de distribuição empírica e ajustada e QQ-plot).



**Figura 4.3:** Gráficos de Qualidade de Ajuste para a Despesa dos estados de São Paulo e Rio de Janeiro (função de densidade empírica e ajustada, função de distribuição empírica e ajustada e QQ-plot).



**Figura 4.4:** Gráficos de Qualidade de Ajuste para a Despesa dos estados de Minas Gerais e Espírito Santo (função de densidade empírica e ajustada, função de distribuição empírica e ajustada e QQ-plot).

### 4.3 Dependência das Marginais

Seguindo os procedimentos para ajustar modelos de dependência com cópulas, a breve discussão presente nesta etapa pretende avaliar a estrutura de dependência e o comportamento das caudas das amostra bivariadas (renda e despesa) por estado.

Joe (2014) defende a utilização de gráficos scores normais para avaliar a estrutura de dependência das amostras. A aplicação dos gráficos de scores normal, para duas variáveis ao mesmo tempo, em conjunto com outras medidas pode fornecer informações se a estrutura de dependência é positiva (negativa), avaliar dependência na cauda, assimetria e existência de relação monótona.

Os scores normais são calculados com base nas estimativas de MV das marginais. Dado  $\Phi$  a função de distribuição de uma normal padrão e  $\Phi^{-1}$  sua inversa, o score normal é definidos como:

$$z_i = \Phi^{-1} \left[ F_{\hat{\theta}}(y_i) \right],$$

no qual  $F_{\hat{\theta}}$  é a função de distribuição da marginal (Dagum no nosso caso) com a estimativa de MV de  $\theta$  e  $y_i$  a  $i$ -ésima observação da amostra.

Um medida indicada por Joe (2014) para avaliar assimetria e dependência nas caudas é a semi-correlação superior e inferior, definidas, respectivamente, como:

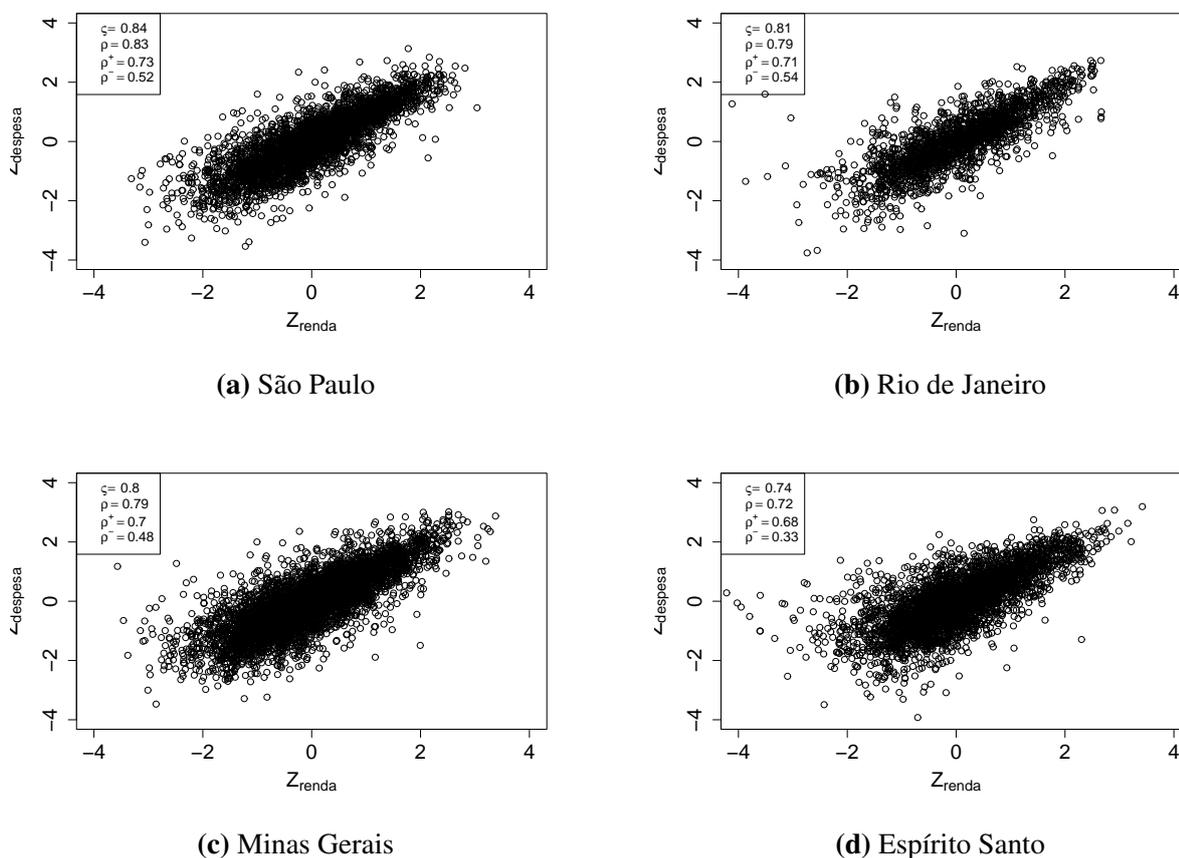
$$\rho^+ = Cor[Z_1, Z_2 | Z_1 > 0, Z_2 > 0],$$

$$\rho^- = Cor[Z_1, Z_2 | Z_1 < 0, Z_2 < 0],$$

no qual  $Z_1$  e  $Z_2$  são os scores normais de duas variáveis aleatórias e  $Cor$  medida de de semi-correlação. A versão amostral das medidas ( $\hat{\rho}^+$  e  $\hat{\rho}^-$ ) é basicamente a correlação no quadrante superior e inferior de duas variáveis após as mesmas terem sido transformadas em scores normais. Se  $\hat{\rho}^+ = \hat{\rho}^-$  é um indicativo que distribuição conjunta das amostras possuem simetria nas caudas.

Joe (2014) deriva a expressão exata para calcular a semi-correlação para uma normal bivariada como função do parâmetro de dependência ( $\rho$ ). Denotaremos a semi-correlação de uma normal bivariada como  $\zeta$ . Se a semi-correlação amostral inferior (superior) for maior que  $\zeta$ , calculado com  $\rho$  dos scores normais, é um indicativo da presença de dependência forte na cauda inferior (superior).

Com o objetivo de avaliar a estrutura de dependência de renda e despesa dos estados, as mesmas foram transformadas em scores normais e apresentadas simultaneamente em gráficos por estado, em conjunto com as medidas comentadas.



**Figura 4.5:** Gráficos de renda e despesa, por estado, após serem transformadas em scores normais (com as medidas de correlação amostral, semi-correlação positiva e negativa amostrais e semi-correlação de uma normal bivariada com correlação dada).

Os gráficos expressos em (4.5) sugerem que, para todos os casos avaliados, existe uma

dependência positiva entre as variáveis e relação monótona (característica imprescindível para ajustar cópulas bivariadas com um parâmetro de dependência). A forma dos gráfico com os cantos superiores mais afinados e a semi-correlação amostral superior maior que a inferior ( $\hat{\rho}^+ > \hat{\rho}^-$ ) sugerem assimetria nas caudas. Como as semi-correlações amostrais não apresentaram valores maiores que  $\varsigma$  temos um indicativo que não existe dependência forte nas caudas.

Na próxima etapa iremos estimar o parâmetro de dependência para algumas cópulas consideradas e calcular as medidas de confiabilidade objetivo do capítulo.

#### 4.4 Fragilidade Financeira das famílias do Sudeste do Brasil

Para modelar a dependência das variáveis renda e despesa consideramos as cópulas Gaussiana, Frank, Clayton e Gumbel. Os parâmetros de dependência foram estimados pelos método IFM, MV e, ainda, método não-paramétrico ( $\delta_{np}$ ). Se as estimativas desses parâmetros apresentam grande diferença (relativo ao erro padrão) é um indicativo que o modelo pode não ser adequado.

O método não paramétrico citado utiliza as funções de distribuição empíricas das marginais para estimar apenas o parâmetro de dependência por MV, para mais detalhes veja McNeil et al. (2005).

Como critério de comparação dos modelos, utilizamos o valor da log-verossimilhança nas estimativas e as medidas AIC e BIC. A log-verossimilhança tende a ser maior ou as medidas AIC/BIC tendem a ser menores para modelos de cópula em que há uma melhor aproximação da estrutura de dependência dos dados e uma melhor aproximação da dependência nas caudas (Joe, 2014).

Os resultados estão exibidos nas tabelas (4.3), (4.4), (4.6), (4.5). A primeira coluna (Ref.) expressa as estimativas de MV apresentadas na tabela (4.2), o  $\tau$  de Kendall empírico dos dados e a medida de fragilidade financeira observada, que definimos como:

$$R_{obs} = \frac{\text{número de casas que a despesa superou a renda}}{\text{número de casas da amostra}}.$$

As outras colunas expressam as estimativas dos parâmetros das marginais considerando o método MV com respectivos erros padrões, as estimativas dos parâmetros de dependência, estimativa de  $\tau$  de Kendall para o parâmetro de dependência estimado, a confiabilidade e intervalos de confiança para o método IFM e MV ( $\gamma = 0.05\%$ ), log-verossimilhança nas estimativas e AIC/BIC de cada cópula considerada, além da estimativa da confiabilidade considerando independência.

Pequenas diferenças entre o  $\tau$  de Kendall empírico e o tau estimado pelo parâmetro de dependência da cópula indicam boa aproximação da estrutura de dependência dos dados. As confiabilidades apresentadas foram calculadas utilizando a expressão (2.8).

Ao avaliar os resultados para o estado de São Paulo, depreende-se que os parâmetros das marginais estimados separadamente não apresentaram diferenças significativas aos estimados em conjunto para as cópulas Frank e Gumbel (indicativo de qualidade do ajuste), o que não se configurou para as cópulas Gaussiana e Clayton. Não identificou-se diferença significativa entre os parâmetros de dependência (estimados pelos três métodos considerados) para as cópulas Gaussiana, Frank e Gumbel.

Para esse estado, a cópula Frank foi a que melhor aproximou a estrutura de dependência dos dados, mas foi a cópula Gumbel que obteve os menores valores das medidas AIC e BIC. Como avaliado no item anterior deste capítulo, os dados possuem assimetria nas caudas com uma dependência maior no quadrante superior, característica da cópula Gumbel, e, ainda, esse modelo também aproximou bem a estrutura de dependência o que justifica o melhor ajuste. A cópula Clayton possui dependência forte na cauda inferior o que explica a falta de ajuste.

Avaliando a confiabilidade, o valor da fragilidade financeira observada não situa-se nos limites de nenhum intervalo de confiança construído com base no método IFM, mas está nos limites do intervalo construído com base no método de MV para a cópula Gumbel. O valor da confiabilidade considerando marginais independentes ( $R_{indep}$ ) evidencia a necessidade de incorporar a estrutura de dependência no problema.

Para o estado de Minas Gerais, os parâmetros das marginais estimados separadamente ou

em conjunto não apresentaram diferença significativa para as cópulas Frank e Gumbel. Os parâmetros de dependência não apresentaram diferença significativa segundo método de estimação para as cópulas Gaussiana, Frank e Gumbel. A cópula Gumbel também foi a que melhor se ajustou aos dados e a única que apresentou intervalo de confiança com a confiabilidade observada dentro dos limites considerando o método de MV, apesar da cópula Frank ter apresentado uma melhor aproximação a estrutura de dependência dos dados. Análise semelhante pode ser feita para o estado do Rio de Janeiro com a cópula Gumbel apresentando melhor ajuste

O estado de Espírito Santo apresentou resultados similares aos outros estados em relação aos parâmetros marginais, de dependência e estrutura de dependência. Entretanto, foi o único que a confiabilidade observada não situa-se nos limites de nenhum intervalo construído. Outras cópulas podem ser utilizadas para tentar melhorar o ajuste, o que será deixado para estudos posteriores.

**Tabela 4.3:** Resultados para o estado de São Paulo.

Parâmetros	Ref.	Gaussiana	Frank	Clayton	Gumbel
$\beta_{renda}$	1.86	1.22 (0.068)	1.77 (0.141)	0.65 (0.033)	2.21 (0.158)
$\lambda_{renda}$	0.04	0.05 (0.003)	0.05 (0.003)	0.12 (0.007)	0.03 (0.002)
$\alpha_{renda}$	1.82	2.02 (0.043)	1.77 (0.041)	2.32 (0.060)	1.74 (0.034)
$\beta_{despesa}$	1.57	1.09 (0.059)	1.45 (0.109)	0.56 (0.029)	1.73 (0.112)
$\lambda_{despesa}$	0.03	0.04 (0.002)	0.04 (0.003)	0.09 (0.005)	0.03 (0.002)
$\alpha_{despesa}$	1.66	1.84 (0.040)	1.62 (0.040)	2.17 (0.059)	1.62 (0.033)
$\delta_{np}$		0.83	9.38	1.81	2.637
$\delta_{ifm}$		0.83 (0.003)	9.19 (0.160)	1.83 (0.039)	2.7 (0.035)
$\delta_{MV}$		0.84 (0.005)	9.53 (0.193)	2.92 (0.096)	2.72 (0.048)
$\tau_{ifm}$	0.65	0.62	0.64	0.48	0.63
$\tau_{MV}$		0.63	0.65	0.59	0.63
$R_{ifm}$		0.239 (0.006)	0.225 (0.006)	0.269 (0.005)	0.224 (0.007)
$IC_{ifm}$	0.205	[0.228 - 0.251]	[0.213 - 0.237]	[0.26 - 0.278]	[0.21 - 0.238]
$R_{MV}$		0.238 (0.006)	0.229 (0.006)	0.245 (0.006)	0.216 (0.006)
$IC_{MV}$		[0.227 - 0.249]	[0.218 - 0.24]	[0.233 - 0.257]	[0.204 - 0.227]
log-vero		3559.2	3537.8	3096.5	3645.6
AIC		-7104.4	-7061.7	-6178.9	-7277.6
BIC		-7061.0	-7018.3	-6135.6	-7233.8
$R_{indep}$	0.38				

**Tabela 4.4:** Resultados para o estado de Minas Gerais.

Parâmetros	Ref.	Gaussiana	Frank	Clayton	Gumbel
$\beta_{renda}$	1.89	1.27 (0.062)	1.84 (0.123)	0.76 (0.034)	2.13 (0.126)
$\lambda_{renda}$	0.02	0.03 (0.001)	0.03 (0.002)	0.05 (0.003)	0.02 (0.001)
$\alpha_{renda}$	1.79	1.98 (0.0360)	1.73 (0.034)	2.17 (0.045)	1.73 (0.029)
$\beta_{despesa}$	1.48	1.07 (0.050)	1.43 (0.090)	0.64 (0.028)	1.63 (0.088)
$\lambda_{despesa}$	0.02	0.03 (0.001)	0.03 (0.001)	0.04 (0.002)	0.02 (0.001)
$\alpha_{despesa}$	1.65	1.81 (0.034)	1.60 (0.034)	2.03 (0.045)	1.59 (0.028)
$\delta_{np}$		0.79	8.06	1.53	2.408
$\delta_{ifm}$		0.79 (0.003)	8.01 (0.122)	1.53 (0.031)	2.42 (0.027)
$\delta_{MV}$		0.80 (0.005)	8.30 (0.148)	2.35 (0.070)	2.44 (0.036)
$\tau_{ifm}$	0.60	0.58	0.60	0.43	0.59
$\tau_{MV}$		0.59	0.61	0.54	0.59
$R_{ifm}$		0.269 (0.005)	0.253 (0.006)	0.292 (0.004)	0.256 (0.006)
$IC_{ifm}$	0.24	[0.259 - 0.28]	[0.242 - 0.264]	[0.283 - 0.3]	[0.244-0.268]
$R_{MV}$		0.269 (0.005)	0.254 (0.005)	0.271 (0.005)	0.247 (0.005)
$IC_{MV}$		[0.259 - 0.279]	[0.244 - 0.264]	[0.261 - 0.281]	[0.237 - 0.257]
log-vero		7311.7	7302.5	6695.4	7445.1
AIC		-14609.4	-14591.0	-13376.8	-14876.3
BIC		-14563.7	-14545.4	-13331.2	-14830.6
$R_{indep}$	0.39				

**Tabela 4.5:** Resultados para o estado do Rio de Janeiro.

Parâmetros	Ref.	Gaussiana	Frank	Clayton	Gumbel
$\beta_{renda}$	2.37	1.39 (0.107)	2.26 (0.251)	0.86 (0.065)	2.43 (0.236)
$\lambda_{renda}$	0.03	0.04 (0.003)	0.03 (0.003)	0.08 (0.006)	0.03 (0.002)
$\alpha_{renda}$	1.62	1.83 (0.050)	1.56 (0.045)	1.91 (0.062)	1.62 (0.041)
$\beta_{despesa}$	1.81	1.25 (0.092)	1.77 (0.176)	0.69 (0.046)	1.90 (0.158)
$\lambda_{despesa}$	0.02	0.03 (0.002)	0.03 (0.002)	0.05 (0.005)	0.02 (0.002)
$\alpha_{despesa}$	1.60	1.76(0.049)	1.54 (0.046)	1.95 (0.064)	1.57 (0.04)
$\delta_{np}$		0.799	8.67	1.62	2.54
$\delta_{ifm}$		0.79 (0.004)	8.68 (0.199)	1.60 (0.050)	2.55 (0.044)
$\delta_{MV}$		0.80 (0.008)	8.98 (0.252)	2.61 (0.122)	2.56 (0.061)
$\tau_{ifm}$	0.62	0.58	0.63	0.44	0.61
$\tau_{MV}$		0.59	0.64	0.57	0.61
$R_{ifm}$		0.271 (0.008)	0.246 (0.009)	0.293 (0.007)	0.244 (0.009)
$IC_{ifm}$	0.221	[0.256 - 0.287]	[0.229 - 0.263]	[0.28 - 0.306]	[0.226 - 0.262]
$R_{MV}$		0.268 (0.008)	0.244 (0.008)	0.264 (0.009)	0.237 (0.008)
$IC_{MV}$		[0.252 - 0.284]	[0.228 - 0.26]	[0.247 - 0.28]	[0.221 - 0.253]
log-vero		2343.149	2389.356	2109.696	2446.432
AIC		-4672.298	-4764.712	-4205.392	-4878.864
BIC		-4633.356	-4725.77	-4166.45	-4839.922
$R_{indep}$	0.39				

**Tabela 4.6:** Resultados para o estado de Espírito Santo.

Parâmetros	Ref.	Gaussiana	Frank	Clayton	Gumbel
$\beta_{renda}$	1.27	0.89 (0.041)	1.21 (0.071)	0.67 (0.033)	1.33 (0.070)
$\lambda_{renda}$	0.03	0.03 (0.002)	0.03 (0.002)	0.05 (0.003)	0.03 (0.002)
$\alpha_{renda}$	1.92	2.16 (0.045)	1.90 (0.043)	2.24 (0.053)	1.89 (0.037)
$\beta_{despesa}$	1.98	1.29 (0.082)	1.85 (0.151)	0.90 (0.055)	1.81 (0.126)
$\lambda_{despesa}$	0.02	0.02 (0.001)	0.02 (0.001)	0.03 (0.002)	0.02 (0.001)
$\alpha_{despesa}$	1.57	1.74 (0.039)	1.52 (0.036)	1.80 (0.045)	1.62 (0.034)
$\delta_{np}$		0.72	6.43	1.12	2.11
$\delta_{ifm}$		0.72 (0.006)	6.51 (0.138)	1.08 (0.034)	2.09 (0.03)
$\delta_{MV}$		0.73 (0.008)	6.74 (0.157)	1.70 (0.07)	2.09 (0.036)
$\tau_{ifm}$	0.54	0.51	0.54	0.35	0.52
$\tau_{MV}$		0.52	0.55	0.46	0.52
$R_{ifm}$		0.305 (0.006)	0.284 (0.007)	0.325 (0.006)	0.291 (0.007)
$IC_{ifm}$	0.26	[0.292 - 0.317]	[0.27 - 0.297]	[0.314 - 0.336]	[0.278 - 0.305]
$R_{MV}$		0.302 (0.006)	0.283 (0.006)	0.305 (0.007)	0.275 (0.006)
$IC_{MV}$		[0.29 - 0.315]	[0.27 - 0.296]	[0.292 - 0.318]	[0.262 - 0.288]
log-vero		5021.396	5038.251	4534.991	5156.707
AIC		-10028.792	-10062.502	-9055.982	-10299.414
BIC		-9985.8	-10019.51	-9012.99	-10256.422
$R_{indep}$	0.38				

## 4.5 Conclusão

Este capítulo abordou de maneira geral toda a teoria apresentada no trabalho. Conseguimos aplicar um modelo de Força e Tensão para avaliar a fragilidade financeira das famílias residentes na região sudeste do Brasil, não apresentando apenas estimativas pontuais como, também, intervalares, uma evolução em comparação ao estudo utilizado como base (Domma e Giordano, 2012).

Estimamos que a probabilidade da despesa anual das famílias de São Paulo ultrapassarem a renda anual é 21.6% ( $\pm 1.2\%$ ), para famílias do Rio de Janeiro é 23.7% ( $\pm 1.6\%$ ), as de Minas Gerais 24.7% ( $\pm 1\%$ ) e as do Espírito Santo 27.5% ( $\pm 1.2\%$ ).

Ao comparar os limites dos intervalos de confiança, identificamos que o estado de Espírito Santo possui domicílios potencialmente mais vulneráveis a mudanças desfavoráveis no ambiente econômico do que os domicílios do estado de São Paulo e Rio de Janeiro, assim como os domicílios de Minas Gerais estão mais vulneráveis que os de São Paulo.

Assumir independência entre a despesa e renda anual das família é inapropriado, a diferença entre os valores da confiabilidade apresentados na tabela (4.7) corroboram a afirmação. A tabela citada foi construída com base nos resultados da etapa anterior considerando a cópula Gumbel para estimação de  $R_{MV}$ , dado que foi a que apresentou melhor ajuste para todos os estados (AIC/BIC).

**Tabela 4.7:** Fragilidade Financeira das famílias residente nos estados do Sudeste do Brasil.

Parâmetros	São Paulo	Rio de Janeiro	Minas Gerais	Espírito Santo
$R_{obs}$	0.205	0.221	0.24	0.26
$R_{indep}$	0.38	0.39	0.39	0.38
$R_{MV}$	0.216	0.237	0.247	0.275
$IC_{MV}$	[0.204 - 0.227]	[0.221 - 0.253]	[0.237 - 0.257]	[0.262 - 0.288]

Domma e Giordano (2012) cita como possível evolução do estudo a avaliação temporal da

medida proposta, bem como utilização de covariáveis para explicar a fragilidade financeira, duas questões pertinentes também para nosso trabalho.

Utilizamos cópulas específicas para o exercício e, mesmo que tenham apresentado bons ajustes, a avaliação com outras cópulas pode ser considerada. Algumas cópulas que possuem dois parâmetros de dependência detêm características que podem melhorar os ajustes apresentados e toda a teoria apresentada no capítulo (2) pode ser generalizada para esse caso.

Aqui fizemos uma aplicação com os estados do sudeste do país, uma próxima etapa natural é estimar a medida proposta para outros estados e poder fazer uma comparação a nível nacional. A inclusão de possíveis variáveis explicativas pode auxiliar a avaliação do comportamento da medida.



# Capítulo 5

## Considerações finais

A extensão deste trabalho abordou em diversos aspectos os modelos Força e Tensão. Esses tipos de modelos possuem grande aplicabilidade e flexibilizam questões envolvidas na comparação de duas variáveis aleatórias. Já foram alvo de muitos estudos, os quais avaliaram o tema em diferentes momentos e detalhes. Acreditamos que conseguimos fazer uma breve contribuição para a área propondo métodos de estimação intervalares.

A medida  $R = P(X < Y)$  possibilita a comparação de duas variáveis aleatórias sem utilizar apenas medidas de localização. Em geral, o método de estimação utilizado é MV e alguns resultados conhecidos na literatura em relação a conceitos inferenciais para se estimar  $R$  com algumas distribuições conhecidas foram apresentados.

Apresentamos uma breve revisão sobre cópulas e conseguimos abordar de maneira geral a aplicação da teoria em modelos de Força e Tensão. Com base na teoria desenvolvida por Joe (2005), conseguimos apresentar métodos de estimação intervalares para confiabilidade utilizando tanto o método estimação FIM quanto MV.

A aplicação utilizou um modelo de Força e Tensão com variáveis dependentes para avaliar a fragilidade financeira das famílias residentes na região sudeste do Brasil, não apresentando apenas estimativas pontuais como, também, intervalares.

Os intervalos de confiança apresentados para modelos de Força e Tensão com cópulas são

todos assintóticos, esses os quais supõem amostras relativamente grandes. Avaliar as propriedades dos mesmos em amostras menores se faz necessária para utilização em aplicações dessa caso.

A construção de toda a teoria presente na parte 2.3 com covariáveis nos parece desafiadora, bem como a aplicação da mesma em modelos de Força e Tensão com multicomponente.

Uma outra contribuição dada por Kotz e Pensky (2003) é a fundamentação de modelos de Força e Tensão sob uma perspectiva Bayesiana, tema que nos parece pertinente avaliar em estudos futuros incorporando dependência.

De maneira geral, o presente trabalho explorou métodos de estimação que incorporam teoria de cópulas para  $R$ , de tal forma, propôs intervalos assintóticos utilizando propriedades do estimadores de máxima verossimilhança.

## Referências Bibliográficas

MS Abu-Salih e A Shamseldin. Bayesian estimation of  $p(x < y)$  for a bivariate exponential distribution. *Arab Gulf Journal of Scientific Research*, 6(1):17–26, 1988.

ALAA AHMED. Estimation of  $r = p(y < x)$  for the three parameter dagum distribution. *Journal of Statistics: Advances in Theory and Applications*, 13(2):167–178, 2015.

AM Awad, MM Azzam, e MA Hamdan. Some inference results on  $pr(x < y)$  in the bivariate exponential model. *Communications in Statistics-Theory and Methods*, 10(24):2515–2525, 1981.

Xuchao Bai, Yimin Shi, Yiming Liu, e Bin Liu. Reliability inference of stress-strength model for the truncated proportional hazard rate distribution under progressively type-ii censored samples. *Applied Mathematical Modelling*, 2018a.

Xuchao Bai, Yimin Shi, Yiming Liu, e Bin Liu. Reliability estimation of multicomponent stress–strength model based on copula function under progressively hybrid censoring. *Journal of Computational and Applied Mathematics*, 344:100–114, 2018b.

Ripsy Bandourian, James McDonald, e Robert Turley. A comparison of parametric models of income distribution across countries and over time. 2002.

Alessandro Barbiero. Confidence intervals for reliability of stress-strength models in the normal case. *Communications in Statistics-Simulation and Computation*, 40(6):907–925, 2011.

Ernst R Berndt, Bronwyn H Hall, Robert E Hall, e Jerry A Hausman. Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4*, pages 653–665. NBER, 1974.

Sarah Brown e Karl Taylor. Household debt and financial assets: evidence from germany, great britain and the usa. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3):615–643, 2008.

George Casella e Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Umberto Cherubini, Elisa Luciano, e Walter Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004.

James D Church e Bernard Harris. The estimation of reliability from stress-strength relationships. *Technometrics*, 12(1):49–54, 1970.

Camilo Dagum. New model of personal income-distribution-specification and estimation. *Economie appliquée*, 30(3):413–437, 1977.

Filippo Domma e Sabrina Giordano. A stress–strength model with dependent variables to measure household financial fragility. *Statistical Methods & Applications*, 21(3):375–389, 2012.

Filippo Domma e Sabrina Giordano. A copula-based approach to account for dependence in stress-strength models. *Statistical Papers*, 54(3):807–826, 2013.

Barry Eichengreen e Ricardo Hausmann. Exchange rates and financial fragility. Technical report, National bureau of economic research, 1999.

Paul Embrechts, Filip Lindskog, e Alexander McNeil. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, 2001.

- Peter Enis e Seymour Geisser. Estimation of the probability that  $y < x$ . *Journal of the American Statistical Association*, 66(333):162–168, 1971.
- Jianxiong Gao, Zongwen An, e Bo Liu. A dependent stress–strength interference model based on mixed copula function. *Journal of Mechanical Science and Technology*, 30(10):4443–4446, 2016.
- Joseph L Gastwirth. A general definition of the lorenz curve. *Econometrica: Journal of the Econometric Society*, pages 1037–1039, 1971.
- Charles AE Goodhart, Pojanart Sunirand, e Dimitrios P Tsomocos. A model to analyse financial fragility. *Economic Theory*, 27(1):107–142, 2006.
- Huizhen Guo e Kalimuthu Krishnamoorthy. New approximate inferential methods for the reliability parameter in a stress–strength model: The normal case. *Communications in Statistics-Theory and Methods*, 33(7):1715–1731, 2004.
- Ramesh C Gupta e Sundarraman Subramanian. Estimation of reliability in a bivariate normal distribution with equal coefficients of variation. *Communications in Statistics-Simulation and Computation*, 27(3):675–698, 1998.
- Ramesh C Gupta, ME Ghitany, e Dhaifalla K Al-Mutairi. Estimation of reliability in a parallel system with random sample size. *Mathematics and Computers in Simulation*, 83:44–55, 2012.
- David D Hanagal. Note on estimation of reliability under bivariate pareto stress-strength model. *Statistical papers*, 38(4):453, 1997.
- T Jappelli, M Pagano, e M Di Maggio. Households’ indebtedness and financial fragility’, csef working paper no. 208. 2008.

Harry Joe. *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC, 1997.

Harry Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419, 2005.

Harry Joe. *Dependence modeling with copulas*. Chapman and Hall/CRC, 2014.

Norman L Johnson, Samuel Kotz, e N Balakrishnan. Continuous univariate distributions, vol. 2 of wiley series in probability and mathematical statistics: applied probability and statistics, 1995.

Richard A Johnson e Dean W Wichern. Multivariate analysis. *Encyclopedia of Statistical Sciences*, 8, 2004.

Christian Kleiber. Dagum vs. singh-maddala income distributions. *Economics Letters*, 53(3): 265–268, 1996.

Christian Kleiber. A guide to the dagum distributions. In *Modeling Income Distributions and Lorenz Curves*, pages 97–117. Springer, 2008.

Samuel Kotz e Marianna Pensky. *The stress-strength model and its generalizations: theory and applications*. World Scientific, 2003.

K Krishnamoorthy e Yin Lin. Confidence limits for stress–strength reliability involving weibull models. *Journal of Statistical Planning and Inference*, 140(7):1754–1764, 2010.

K Krishnamoorthy, Thomas Mathew, e Shubhabrata Mukherjee. Normal-based methods for a gamma distribution: Prediction and tolerance intervals and stress-strength reliability. *Technometrics*, 50(1):69–78, 2008.

Annamaria Lusardi, Daniel J Schneider, e Peter Tufano. Financially fragile households: Evidence and implications. Technical report, National Bureau of Economic Research, 2011.

- Kantilal Varichand Mardia. *Families of bivariate distributions*, volume 27. Lubrecht & Cramer Ltd, 1970.
- M Masoom Ali e Jungsoo Woo. Inference on reliability  $p$  ( $y < x$ ) in the levy distribution. *Mathematical and Computer Modelling: An International Journal*, 41(8-9):965–971, 2005.
- James B McDonald. Some generalized functions for the size distribution of income. *Econometrica: journal of the Econometric Society*, pages 647–663, 1984.
- AT McKay. Sampling from batches. *Supplement to the Journal of the Royal Statistical Society*, 1(2):207–216, 1934.
- Alexander J McNeil, Rüdiger Frey, Paul Embrechts, et al. *Quantitative risk management: Concepts, techniques and tools*, volume 3. Princeton university press Princeton, 2005.
- Hyman P Minsky. The financial instability hypothesis. 1992.
- Hyman P Minsky et al. Stabilizing an unstable economy: The lessons for industry, finance and government. 1986.
- SP Mukherjee e LK Sharan. Estimation of failure probability from a bivariate normal stress-strength distribution. *Microelectronics Reliability*, 25(4):699–702, 1985.
- Saralees Nadarajah. Reliability for some bivariate beta distributions. *Mathematical Problems in Engineering*, 2005(1):101–111, 2005a.
- Saralees Nadarajah. Reliability for some bivariate gamma distributions. *Mathematical Problems in Engineering*, 2005(2):151–163, 2005b.
- Saralees Nadarajah e Samuel Kotz. Reliability models based on bivariate exponential distributions. *Probabilistic engineering mechanics*, 21(4):338–351, 2006.
- Roger B Nelsen. Copulas and association. In *Advances in probability distributions with given marginals*, pages 51–74. Springer, 1991.

- Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.
- Pierre Nguimkeu, Marie Rekkas, e Augustine Wong. Interval estimation for the stress-strength reliability with bivariate normal variables. *Open Journal of Statistics*, 4(08):630, 2014.
- Pierre Nguimkeu, Marie Rekkas, e Augustine Wong. Interval estimation of the stress-strength reliability with independent normal random variables. *Communications in Statistics-Theory and Methods*, 44(6):1210–1221, 2015.
- Cira E Guevara Otiniano, TR Sousa, e Pushpa N Rathie. Stable random variables: Convolution and reliability. *Journal of Computational and Applied Mathematics*, 242:1–11, 2013.
- Donald B Owen, KJ Craswell, e David Lee Hanson. Nonparametric upper confidence bounds for  $\text{pr}\{Y < X\}$  and confidence limits for  $\text{pr}\{Y < X\}$  when  $x$  and  $y$  are normal. *Journal of the American Statistical Association*, 59(307):906–924, 1964.
- Pushpa N Rathie e Luan Carlos de SM Ozelim. Exact and approximate expressions for the reliability of stable lévy random variables with applications to stock market modelling. *Journal of Computational and Applied Mathematics*, 321:314–322, 2017.
- Benjamin Reiser e Irwin Guttman. Statistical inference for  $\text{pr}(y < x)$ : the normal case. *Technometrics*, 28(3):253–257, 1986.
- Marco Scarsini. On measures of concordance. *Stochastica*, 8(3):201–218, 1984.
- Berthold Schweizer, Edward F Wolff, et al. On nonparametric measures of dependence for random variables. *The annals of statistics*, 9(4):879–885, 1981.
- S Sengupta. Unbiased estimation of  $p(x > y)$  for two-parameter exponential populations using order statistics. *Statistics*, 45(2):179–188, 2011.
- SK Singh e GS Maddala. A function for size distribution of incomes. *Econometrica: Journal of the Econometric Society*, pages 963–970, 1976.

M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

Samaradasa Weerahandi e Richard A Johnson. Testing reliability in a stress-strength model when x and y are normally distributed. *Technometrics*, 34(1):83–91, 1992.

Douglas A Wolfe e Robert V Hogg. On constructing statistics and reporting data. *The American Statistician*, 25(4):27–30, 1971.