



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

**Redução de Viés em Estimativas de Máxima  
Verossimilhança Modificadas para a Distribuição  
Weibull de Três Parâmetros**

por

**Adriana Pereira da Silva**

Brasília, março de 2025

# **Redução de Viés em Estimativas de Máxima Verossimilhança Modificadas para a Distribuição Weibull de Três Parâmetros**

**por**

**Adriana Pereira da Silva**

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

Orientador: Prof. Dr. Felipe S. Quintino

Coorientador: Prof. Dr. Frederico M. Almeida

Brasília, março de 2025

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 título de Mestre em Estatística.

Texto aprovado por:

Prof. Dr. Felipe Sousa Quintino  
Orientador, EST/UnB

Prof. Dr. Frederico M. Almeida  
Coorientador, EST/UnB

Prof. Dr. Jose Ailton Alencar Andrade  
Externo, UFC

Prof. Dr. Helton Saulo Bezerra Dos Santos  
Interno, EST/UnB

*O sucesso é a soma de pequenos esforços repetidos dia após dia..*

(Robert Collier)

À minha família pelo apoio incondicional e incentivo em todos os momentos. Vocês são minha maior fonte de inspiração e motivação para seguir sempre em frente.

# Agradecimentos

A jornada até a conclusão deste trabalho foi repleta de desafios, aprendizados e conquistas, e nada disso teria sido possível sem o apoio de pessoas especiais.

Agradeço, primeiramente, ao meu esposo, Elcio, pelo amor, paciência e incentivo inabalável. Sua compreensão nos momentos de ausência, seu apoio constante e sua fé em mim foram fundamentais para que eu chegasse até aqui.

Aos meus filhos, Davi e Daniel, minha maior fonte de inspiração. Cada olhar, cada abraço e cada palavra de carinho foram o combustível que me impulsionou a seguir em frente. Espero que este trabalho seja um exemplo de que dedicação e esforço.

À minha família e amigos, que sempre me apoiaram, oferecendo amor e suporte nos momentos mais difíceis.

Meus sinceros agradecimentos aos professores do PPGEST/UnB, em especial aos meus orientadores Felipe Quintino e Frederico Almeida, por compartilharem seu conhecimento e por me guiarem com paciência e dedicação ao longo deste percurso.

Aos colegas e amigos que, de alguma forma, contribuíram para esta caminhada, seja com palavras de incentivo, apoio nos momentos difíceis e estudos em grupo.

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

A todos vocês, minha mais profunda gratidão!

# Resumo

Neste trabalho, estudamos o problema de estimação de parâmetros em modelos Weibull de três parâmetros, para os quais estimativas não finitas podem ser obtidas para a função de verossimilhança em algumas regiões do espaço paramétrico. Baseados na penalização do logaritmo da função verossimilhança modificada, propomos uma nova classe de estimadores para tal modelo de distribuições. Além de obter estimativas finitas para os parâmetros do modelo, tal procedimento possibilita uma redução no viés dos estimadores modificados. O novo método foi comparado a outros métodos encontrados na literatura através de estudos de simulações de Monte Carlo. Os resultados dos estudos das simulações mostraram que o método de penalização do vetor escore modificado apresentou melhor desempenho que o método do logaritmo da função de verossimilhança modificada. Foram apresentadas duas aplicações em dados reais, a primeira relativa a resistência de fibras de carbono, bastante estudadas na literatura, e a segunda referente a investimentos estrangeiros que geram emprego ou renda no país para concessão de autorização de residência ao imigrante, analisada nesse contexto pela primeira vez.

**Palavras-chave:** Máxima Verossimilhança Modificada, Penalização, Simulações de Monte Carlo, Viés, Modelos Irregulares.

# Abstract

In this work, we investigate the parameter estimation problem based on the three-parameter Weibull models, for which non-finite estimates may be obtained for the log-likelihood function in some regions of the parametric space. Based on a penalization of the modified log-likelihood function, we propose a new class of estimators for this distribution. In addition to providing finite estimates for the model parameters, this procedure reduces the bias of the modified estimator. The new method is compared to others in the literature through a Monte Carlo simulation study. The simulation results showed that the modified score vector penalty method outperforms the modified log-likelihood function method. We present two applications using real data. The first relates to the resistance of carbon fibers, which has been extensively studied in the literature, and the second refers to foreign investments that generate employment or income in the country to grant a residence permit to immigrants, analyzed in this context for the first time.

**Keywords:** Modified Maximum Likelihood, Penalization, Monte Carlo simulations, Bias, Irregular Models.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Distribuição Weibull</b>	<b>5</b>
2.1	Introdução . . . . .	5
2.2	Contexto Histórico . . . . .	5
2.3	Construção da Distribuição Weibull via Mínimos Parciais . . . . .	7
2.4	Definições e Propriedades Probabilísticas . . . . .	11
2.4.1	Esperança . . . . .	15
2.4.2	Variância . . . . .	16
2.4.3	Momentos . . . . .	16
2.4.4	Função Quantil . . . . .	17
2.4.5	Função Geradora de Momentos . . . . .	18
2.4.6	Função Característica . . . . .	18
2.5	Métodos de Estimação dos Parâmetros . . . . .	19
2.5.1	EMVs Modificados . . . . .	20
2.5.2	EMVs Corrigidos . . . . .	25
<b>3</b>	<b>Redução de Viés nos EMVMs</b>	<b>27</b>
3.1	Introdução . . . . .	27
3.2	EMVs Duplamente Modificados . . . . .	27

<b>4</b>	<b>Estudos de Simulação</b>	<b>31</b>
<b>5</b>	<b>Aplicações a Dados Reais</b>	<b>38</b>
5.1	Resistência de Fibras de Carbono . . . . .	38
5.2	Investimentos Estrangeiros . . . . .	43
<b>6</b>	<b>Considerações Finais</b>	<b>50</b>
	<b>Apêndice</b>	<b>52</b>
	<b>Referências Bibliográficas</b>	<b>54</b>

# Lista de Tabelas

5.1	Parâmetros estimados, Máxima verossimilhança, AIC e BIC para os dados de resistência . . . . .	40
5.2	P-valor dos testes Kolmogorov-Smirnov (KS) e Cramér-von Mises (CVM) para os dados de resistência . . . . .	40
5.3	10 maiores valores de investimentos por País de origem . . . . .	46
5.4	Parâmetros estimados, Máxima verossimilhança, AIC e BIC para os dados de investimento . . . . .	47

# Lista de Figuras

2.1	Gráfico das classes de $L(x)$ para $\alpha=2$ . . . . .	11
2.2	Gráfico da FDP $f$ (esquerda) e FDA $F$ (direita) com variação do parâmetro de forma $0 < \alpha < 1$ . . . . .	12
2.3	Gráfico da FDP $f$ (esquerda) e FDA $F$ (direita) com variação do parâmetro de forma $\alpha$ . . . . .	13
2.4	Gráfico da FDP $f$ (esquerda) e FDA $F$ (direita) com variação do parâmetro de escala $\sigma$ . . . . .	13
2.5	Gráfico da FDP $f$ (esquerda) e FDA $F$ (direita) com variação do parâmetro de locação $\mu$ . . . . .	14
4.1	Viés (esquerda) e RMSE (direita) para as estimativas de $\alpha = 1.0$ (acima), $\sigma = 1.0$ (centro), and $\mu = 1.0$ (abaixo). . . . .	34
4.2	Viés (esquerda) e RMSE (direita) para as estimativas de $\alpha = 0.5$ (acima), $\sigma = 2.0$ (centro), and $\mu = 1.0$ (abaixo). . . . .	35
4.3	Viés (esquerda) e RMSE (direita) para as estimativas de $\alpha = 1.5$ (acima), $\sigma = 2.0$ (centro), and $\mu = 1.0$ (abaixo). . . . .	36
4.4	Probabilidade de cobertura das estimativas de $\alpha$ (esquerda) e $\sigma$ (direita). . . . .	37
5.1	Boxplot e histograma da resistência das fibras de carbono com calibre de 20mm com fit da distribuição Weibull de 2 parâmetros . . . . .	39

5.2	Ajuste do modelo Weibull pelos métodos de estimação via ECDF e Histograma da resistência das fibras de carbono com calibre de 20mm . . . . .	41
5.3	QQ-plot resíduos - EMVM . . . . .	42
5.4	QQ-plot resíduos - EMVC . . . . .	42
5.5	QQ-plot resíduos - EMVDM . . . . .	43
5.6	QQ-plot resíduos - EMV2p . . . . .	43
5.7	Valor do Investimento (x R\$ 100.000,00) por UF . . . . .	45
5.8	Boxplot e histograma dos valores do investimento . . . . .	45
5.9	Ajuste do modelo Weibull pelos métodos de estimação via ECDF e Histograma dos dados de investimento . . . . .	47
5.10	QQ-plot resíduos - EMVM . . . . .	48
5.11	QQ-plot resíduos - EMVC . . . . .	48
5.12	QQ-plot resíduos - EMVDM . . . . .	49
5.13	QQ-plot resíduos - EMV2p . . . . .	49



# Capítulo 1

## Introdução

Na inferência paramétrica os modelos não-regulares (veja Resnick (2008) e Smith (1985)) têm recebido ampla atenção na literatura. Tais modelos são aqueles cujas derivadas da função de verossimilhança, para determinados parâmetros, não existem ou não são contínuas, que a informação de Fisher é infinita ou não existe, ou ainda, que a função de verossimilhança apresenta singularidades. Este comportamento pode ocorrer em várias situações práticas como, por exemplo, nas distribuições cujo suporte depende de parâmetros desconhecidos.

Nos casos em que o suporte da função densidade de probabilidade (FDP) depende de parâmetros desconhecidos, os métodos convencionais de obtenção de estimativas dos parâmetros se tornam inviáveis. Isso se deve ao fato de que a função de verossimilhança pode exibir uma tendência monótona, fazendo com que o estimador de máxima verossimilhança não assuma valores finitos. Distribuições como a Raileigh, Weibull e Weibull inversa são típicos modelos de probabilidade não-regulares, pois o suporte da distribuição pode depender de um parâmetro de localização.

Para este estudo foi escolhido o conhecido modelo Weibull de três parâmetros, que é um exemplo típico de modelo probabilístico não-regular. Neste cenário, quando o parâmetro de localização se aproxima do menor valor na amostra pode-se observar funções de verossimilhança ilimitadas, dependendo do valor do parâmetro de forma. Para situações como esta, em que

a estimativa de máxima verossimilhança não existe, Smith (1985) e Cheng e Iles (1987) propuseram uma modificação na função de verossimilhança de modo a obter estimativas finitas e consistentes para os parâmetros. Além destes autores, outros propuseram métodos de estimação para os parâmetros do modelo Weibull. Por exemplo, Kundu e Raqab (2009) propuseram um método de estimação maximizando uma modificação da função de verossimilhança. Estes autores aplicaram o modelo Weibull para estimar a probabilidade de resistência-tensão.

Nagatsuka et al. (2013) introduziram um método de estimação consistente para a distribuição Weibull de três parâmetros baseada em uma transformação de dados que evita o problema da função verossimilhança ilimitada. A estimação de parâmetros baseada em uma amostra progressivamente censurada à direita do Tipo II foi estudada por Ng et al. (2012). Montoya et al. (2019) abordou o problema de estimação dos parâmetros de uma distribuição Weibull baseado na definição original de verossimilhança proposta por Fisher.

O método sugerido por Kundu e Raqab (2009) (veja também Ali et al. (2012), Kundu e Raqab (2013) e Kohansal e Rezakhah (2019)) consiste na exclusão da menor observação da amostra e é uma estratégia eficaz para garantir a estimação finita dos parâmetros da distribuição Weibull, assegurando a finitude da função de verossimilhança (modificada) e a existência dos estimadores. Já o método sugerido por Cheng e Iles (1987) propõe a correção da função de verossimilhança numa vizinhança da menor observação da amostra com base em um parâmetro de perturbação adicional.

Uma solução existente na literatura para evitar o problema da função de verossimilhança monótona é baseada na correção de Firth (1993), que envolve a introdução de uma penalização na função escore para garantir estimativas finitas. A abordagem de Firth (1993) foi originalmente proposta para reduzir o viés dos estimadores de máxima verossimilhança (EMVs), por meio de uma modificação adequada da função escore dos modelos lineares generalizados. Para distribuições regulares, a correção de Firth tem sido extensivamente empregada para abordar a inexistência de valores finitos para o EMV.

Vários autores empregaram o método de Firth em problemas envolvendo distribuições regu-



lares para resolver problemas de não existência de valores finitos para os EMV e, também, para redução do viés dos EMV para distribuições pertencentes a família exponencial. Por exemplo, Arrué et al. (2016) mostrou que o EMV modificado para o parâmetro de forma do modelo de distribuição skew-normal modificado permanece finito, mesmo quando o EMV não modificado seja infinito. Além disso, Almeida et al. (2022) investigou uma função score modificada para verossimilhança monótona no modelo semiparamétrico de cura por mistura. Com base no método de Firth, Arrué et al. (2023) desenvolveu uma inferência baseada na correção de viés da verossimilhança para a distribuição skew-t-normal modificada. Heinze e Schemper (2001) e Almeida et al. (2018) estudaram a ocorrência da função de verossimilhança monótona na presença de observações censuradas, e Almeida et al. (2021, 2022) propuseram uma extensão para contabilizar os sobreviventes de longa duração utilizando a função score modificada para a função de verossimilhança monótona.

O problema da função de verossimilhança ilimitada, conhecido como problema da função de verossimilhança monótona, também foi estudado em regressão logística binária e multinomial por Bull et al. (2002) e Heinze e Ploner (2003), na distribuição Weibull estendida regular modificada por Lima e Cribari-Neto (2019), em distribuições skew-normal e skew-t-normal modificadas por Arrué et al. (2016, 2023), dentre outros estudos.

O objetivo central do nosso trabalho é propor um novo procedimento de estimação dos estimadores da distribuição Weibull de três parâmetros baseado na penalização da função de verossimilhança modificada, proposto por Firth (1993), e comparar os resultados da estimação dos parâmetros com os valores obtidos pelos métodos da máxima verossimilhança corrigida e da máxima verossimilhança modificada propostos por Cheng e Iles (1987) e Kundu e Raqab (2009), respectivamente. Para validação dos novos estimadores, um estudo de simulações de Monte Carlo foi conduzido. Os novos estimadores foram aplicados na modelagem de dados reais. A primeira base de dados foi analisada por Kundu e Raqab (2009), dentre outros autores e a segunda base de dados refere-se à investimentos estrangeiros que geram emprego ou renda no país para concessão de autorização de residência ao imigrante, pela primeira vez analisada

nesse contexto.

Dessa forma, a presente dissertação está organizada em três capítulos. No Capítulo 2 apresentamos em detalhes a distribuição Weibull estudada e apresentamos o contexto histórico de sua construção e aplicações. Em seguida, apresentamos a construção matemática da distribuição, via convergência de mínimos parciais de variáveis aleatórias independentes e identicamente distribuídas. Em sequência, as principais propriedades probabilísticas foram obtidas. Então, os estimadores de máxima verossimilhança modificada de Kundu e Raqab (2009) e os estimadores de máxima verossimilhança corrigidos de Cheng e Iles (1987) são apresentados. No Capítulo 3, estudamos a construção dos chamados estimadores de máxima verossimilhança duplamente modificado (EMVDM), tema central desse trabalho. O capítulo está dividido em três seções. Na Seção 3.2, apresentamos a obtenção analítica dos EMVDMs. No Capítulo 4 apresentamos os resultados dos estudos de simulação, cujo intuito é de avaliar o desempenho do novo estimador. O Capítulo 5 apresenta as aplicações da distribuição Weibull de três parâmetros na modelagem de dados reais de duas bases de dados. Por fim, as considerações finais do trabalho são apresentadas no Capítulo 6.

# Capítulo 2

## Distribuição Weibull

### 2.1 Introdução

Neste capítulo, apresentamos conceitos que serão utilizados ao longo desta dissertação. Iniciamos a Seção 2.2 apresentando o contexto histórico da distribuição Weibull e suas aplicações propostas ao longo do tempo. A Seção 2.3 é dedicada a apresentação da construção matemática da distribuição, via convergência de mínimos parciais de variáveis aleatórias independentes e identicamente distribuídas. A Seção 2.4 apresenta as principais propriedades probabilísticas da distribuição Weibull de três parâmetros. Finalmente, destacamos que os conceitos abordados nesse capítulo servem como base para eventuais consultas durante a leitura do Capítulo 3. As principais referências utilizadas para a escrita deste capítulo foram os livros de Johnson e Kotz (1970), Galambos (1978) e Resnick (2008), e os artigos de Cheng e Iles (1987) e Kundu e Raqab (2009).

### 2.2 Contexto Histórico

Em 1951, o físico, engenheiro e matemático Waloddi Weibull, descreveu a distribuição Weibull detalhadamente, para representar a distribuição da resistência à ruptura de materiais e para uma ampla variedade de outras aplicações Johnson e Kotz (1970). É uma distribuição

continua amplamente utilizada em análises de confiabilidade, estudo de vida útil de produtos, e modelagem de dados que envolvam falhas e duração de eventos. As aplicações do modelo Weibull podem ser verificadas em diversas áreas da ciência como: medicina (MATSUSHITA et al., 1992), engenharia (KUNDU; RAQAB, 2009), risco em seguros (GEBIZLIOGLU et al., 2011), hidrologia (CLARKE, 2002), climatologia (PAPALEXIOU et al., 2018), e finanças (MITTNIK; RACHEV, 1993), dentre outras.

Na área de saúde, a distribuição Weibull pode ser utilizada na análise de dados de sobrevivência em ensaios clínicos e em incidência de doenças. Na área ambiental, Johnson e Kotz (1970), utilizou a distribuição Weibull para análise do diâmetro de árvores e momento adequado para coleta e para análise da velocidade do vento.

No artigo publicado por Matsushita et al. (1992), a distribuição Weibull foi aplicada à tábua de vida e aos padrões de idade das doenças no Japão. O objetivo era analisar a epidemiologia do envelhecimento humano e das doenças pela distribuição Weibull. A longevidade dos pacientes analisados, durante o século passado, manifestou-se com aumento nos parâmetros de escala em três períodos analisados e nos parâmetros de forma em um dos períodos com predominância feminina.

Em um estudo sobre retorno de ativos financeiros, Mittnik e Rachev (1993) comprovaram que a distribuição Weibull superou as outras distribuições estáveis consideradas pelos autores. A distribuição Weibull também vem sendo utilizada para detectar tendências ao longo do tempo de variáveis hidrológicas e climáticas. Clarke (2002) tratou das mudanças na frequência e intensidade de alguns fenômenos climáticos, particularmente as vazões mínimas anuais do Rio Paraguai por um período de 19 anos devido às precipitações, no qual o modelo Weibull mostrou tendências temporais nesses dados.

Na engenharia, a distribuição Weibull é utilizada para avaliar a vida útil de máquinas, produtos ou peças. O estudo de Kundu e Raqab (2009), estimou o parâmetro de tensão-resistência usando a distribuição Weibull de três parâmetros quando todos os parâmetros são desconhecidos, ou seja, se  $X$  é a resistência de um sistema que é submetido a uma tensão  $Y$ , então

$R = P(X < Y)$  é uma medida do desempenho do sistema. O sistema falha, se e somente se, a qualquer momento a tensão aplicada for maior que sua resistência. Supondo que as duas populações têm os mesmos parâmetros de forma e locação, mas parâmetros de escala diferentes, os autores obtiveram os estimadores de máxima verossimilhança modificados dos parâmetros desconhecidos excluindo a menor observação do conjunto de dados, para evitar problemas de função de verossimilhança monótona.

A distribuição Weibull é uma das mais utilizadas como modelo de probabilidade para gerenciamento de risco atuarial e financeiro. No estudo de Gebizlioglu et al. (2011) a distribuição Weibull e seus quantis foram considerados no contexto da estimativa de uma medida de risco, e através de simulações investigaram a eficiência de diversos estimadores.

Em climatologia, o estudo de Papalexiou et al. (2018) indicou a distribuição Weibull como um modelo mais robusto para mapeamento dos extremos de precipitação nos Estados Unidos. Este estudo destaca a importância de estudar intensidades de precipitação e suas probabilidades para melhoria de projeções de modelos climáticos que servem de apoio à decisão e operação na prevenção de inundações.

## 2.3 Construção da Distribuição Weibull via Mínimos Parciais

As definições e notações utilizadas nesta seção estão de acordo com as apresentadas por Galambos (1978). Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.) com função de distribuição acumulada (FDA) comum  $F(x)$ .

Defina o mínimo da amostra como

$$m_n = \min\{X_1, \dots, X_n\}.$$

Tem-se que, se a FDA  $F$  de  $X_i$  é conhecida, então, para cada valor de  $n \in \mathbb{N}$ , a FDA do

mínimo  $m_n$  pode ser obtida da seguinte forma:

$$F_{m_n}(x) := P(m_n \leq x) = 1 - (1 - F(x))^n.$$

De fato,

$$\begin{aligned} P(m_n > x) &= P(\min\{X_1, \dots, X_n\} > x) \\ &= P(X_1 > x, \dots, X_n > x). \end{aligned}$$

Como  $X_1, \dots, X_n$  são v.a.'s i.i.d., obtemos

$$\begin{aligned} P(m_n > x) &= P(X_1 > x) \cdots P(X_n > x) \\ &= (1 - F(x)) \cdots (1 - F(x)) \\ &= (1 - F(x))^n. \end{aligned}$$

Logo, a FDA de  $m_n$  é dada por

$$P(m_n \leq x) = 1 - (1 - F(x))^n.$$

Observe que

$$\lim_{n \rightarrow \infty} P(m_n \leq x) = 1 - \lim_{n \rightarrow \infty} (1 - F(x))^n.$$

Se  $x \in \mathbb{R}$  é tal que  $F(x) = 0$ , então

$$P(m_n \leq x) = 0, \quad \forall n \in \mathbb{N}.$$

Por outro lado, se  $x \in \mathbb{R}$  é tal que  $0 < F(x) \leq 1$ , então  $0 \leq 1 - F(x) < 1$ . Consequentemente,

$$\lim_{n \rightarrow \infty} P(m_n \leq x) = 1 - 0 = 1.$$

Segue que

$$\lim_{n \rightarrow \infty} P(m_n \leq x) = \begin{cases} 0, & x < \ell(F), \\ 1, & x \geq \ell(F), \end{cases}$$

em que o *ponto extremo inferior* é denotado por  $\ell(F) = \inf\{x \in \mathbb{R}; F(x) > 0\}$ . Logo,

$$m_n \xrightarrow{d} \ell(F), \text{ quando } n \rightarrow \infty.$$

Como  $\ell(F)$  é constante,

$$m_n \xrightarrow{p} \ell(F), \text{ quando } n \rightarrow \infty.$$

Em outras palavras,  $m_n$  é um estimador fracamente consistente para  $\ell(F)$ . Aqui  $\xrightarrow{d}$  e  $\xrightarrow{p}$  indicam convergência em distribuição e em probabilidade, respectivamente. Nesse caso, se desejarmos obter uma distribuição limite não-degenerada para o mínimo parcial, deveremos considerar uma normalização, ou seja, estudar o possível limite  $L(x)$  tal que

$$\lim_{n \rightarrow \infty} P\left(\frac{m_n - b_n}{a_n} \leq x\right) = L(x), \quad (2.1)$$

para sequências de números reais  $\{a_n\}$  e  $\{b_n\}$  escolhidas apropriadamente e sendo  $x$  ponto de continuidade de  $L(\cdot)$ .

**Observação 2.3.1.** A notação de mínimos parciais  $m_n$  adotada nesse capítulo segue as notações estabelecidas em Galambos (1978), onde foram apresentados em detalhes as provas dos teoremas limites envolvendo  $m_n$ . Além dessas provas, estabeleceu regras para a construção das sequências de constantes  $a_n, b_n > 0$ , além de critérios para  $F(x)$  em que (2.1) é válida. Essa definição também foi descrita por Resnick (2008).

O resultado a seguir é uma adaptação do Teorema de Fisher e Tippett (1928), verificado também por Resnick (2008), para o caso de máximos parciais.

**Proposição 2.3.1** (Adaptação do Teorema de Fisher e Tippett). *Suponha que existam sequências reais  $\{a_n\}$  e  $\{b_n\}$  com  $a_n > 0$  tais que,*

$$\lim_{n \rightarrow \infty} P\left(\frac{m_n - a_n}{b_n} \leq x\right) = \lim_{n \rightarrow \infty} [1 - (1 - F(a_n + b_n x))^n] = L(x), \quad (2.2)$$

*para todo  $x \in C(L)$ , ponto de continuidade de  $L(\cdot)$ , em que  $L$  é assumida não-degenerada. Então  $L$  é do mesmo tipo de uma das três classes a seguir:*

$$L_1(x; \alpha) = \begin{cases} 1 - \exp(-(-x)^{-\alpha}), & \text{se } x < 0, \\ 1, & \text{se } x \geq 0, \end{cases} \quad (2.3)$$

*para algum  $\alpha > 0$ .*

$$L_2(x; \alpha) = \begin{cases} 1 - \exp(-x^\alpha), & \text{se } x \geq 0, \\ 0, & \text{se } x < 0, \end{cases} \quad (2.4)$$

*para algum  $\alpha > 0$ .*

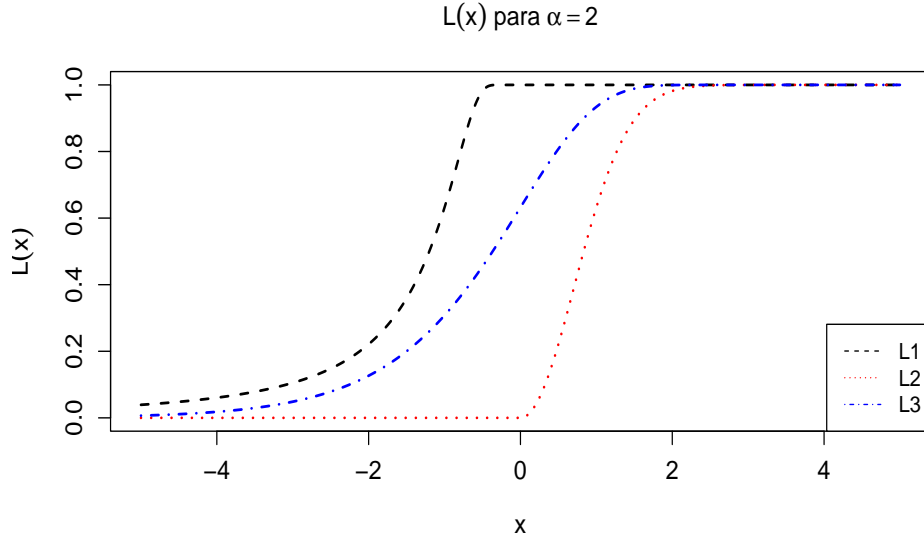
$$L_3(x) = 1 - \exp(-e^x), \quad x \in \mathbb{R}. \quad (2.5)$$

A Figura 2.1 é a representação gráfica da função de distribuição  $L(x)$ , considerando as três classes definidas na Proposição 2.3.1, considerando  $\alpha=2$  para (i) e (ii):

Neste trabalho, focaremos nossos estudos em distribuições do mesmo tipo de  $L_2(x; \alpha)$ , ou seja, as funções de distribuição  $G(\cdot)$  para as quais existem constantes  $\mu \in \mathbb{R}$  e  $\beta \in (0, \infty)$  tais que

$$G(x) = L_2\left(\frac{x - \mu}{\beta}; \alpha\right). \quad (2.6)$$





**Figura 2.1:** Gráfico das classes de  $L(x)$  para  $\alpha=2$

## 2.4 Definições e Propriedades Probabilísticas

Com base na distribuição em (2.6), dizemos que uma v.a.  $X$  tem distribuição Weibull, denotada por  $X \sim \text{WB}(\boldsymbol{\theta})$ , com  $\boldsymbol{\theta} = (\mu, \alpha, \beta)^\top$  se a FDA é definida da seguinte forma:

$$F(x; \boldsymbol{\theta}) = \begin{cases} 0, & x < \mu, \\ 1 - \exp\left(-\frac{1}{\beta}(x - \mu)^\alpha\right), & x \geq \mu. \end{cases} \quad (2.7)$$

A função densidade de probabilidade correspondente é dada por

$$f(x; \boldsymbol{\theta}) = \frac{\alpha}{\beta} \left(\frac{x - \mu}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x - \mu}{\beta}\right)^\alpha \mathbb{1}_{(\mu, \infty)}(x), \quad (2.8)$$

em que  $\mathbb{1}_A$  denota a função indicadora do conjunto  $A$ .

**Observação 2.4.1.** Podemos reparametrizar a FDP da distribuição Weibull considerando os parâmetros  $\boldsymbol{\theta} = (\mu, \alpha, \beta)^\top$ , onde  $\beta^\alpha = \sigma$ . Nesse caso, a função densidade de probabilidade

do modelo Weibull reparametrizada é apresentada em (2.9), notação que usaremos daqui em diante:

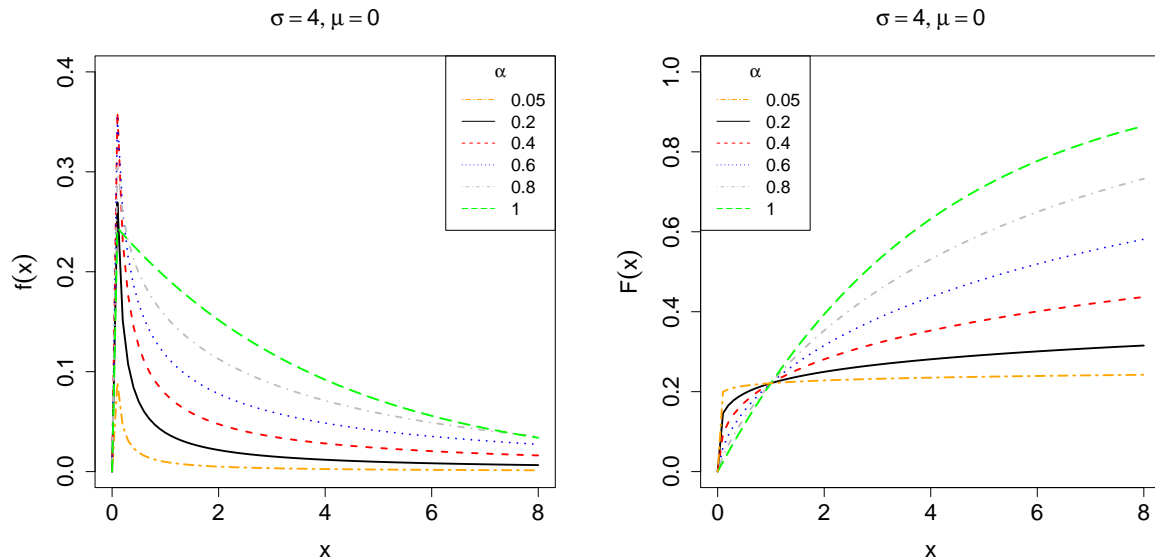
$$f(x; \boldsymbol{\theta}) = \frac{\alpha}{\sigma} (x - \mu)^{\alpha-1} e^{-\frac{1}{\sigma}(x-\mu)^\alpha} \mathbb{1}_{(\mu, \infty)}(x), \quad (2.9)$$

**Observação 2.4.2.** Além da construção da distribuição Weibull reparametrizada (2.9) apresentada na observação anterior, esta também poderia ser construída via distribuição exponencial, da seguinte forma: Seja  $X$  uma v.a. com distribuição  $WB(\boldsymbol{\theta})$  se

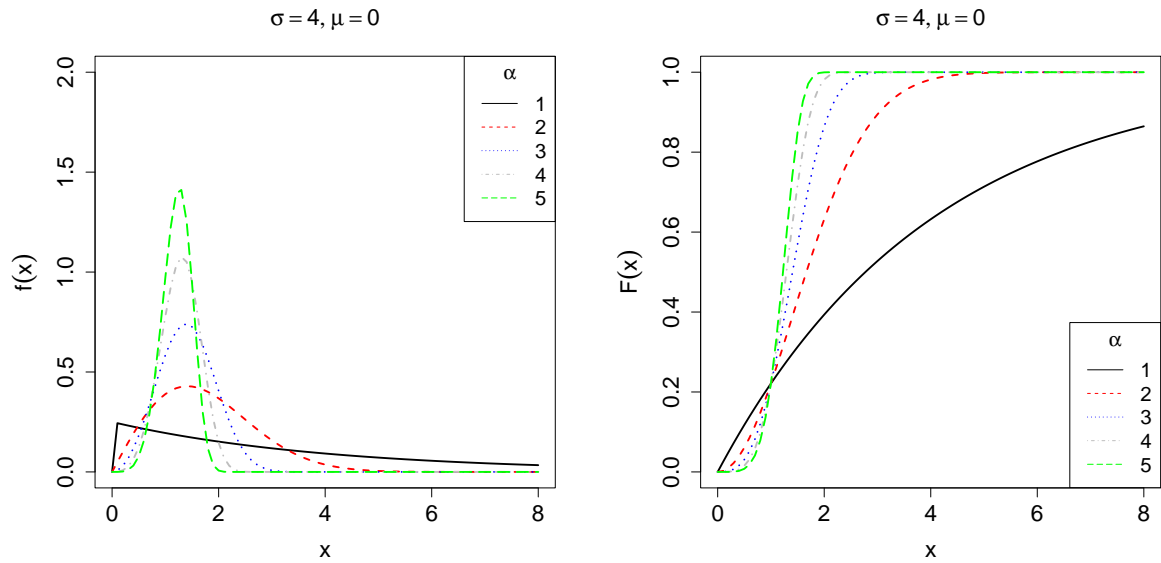
$$Y = \left( \frac{X - \mu}{\beta} \right)^\alpha,$$

onde  $Y$  tem distribuição exponencial padrão.

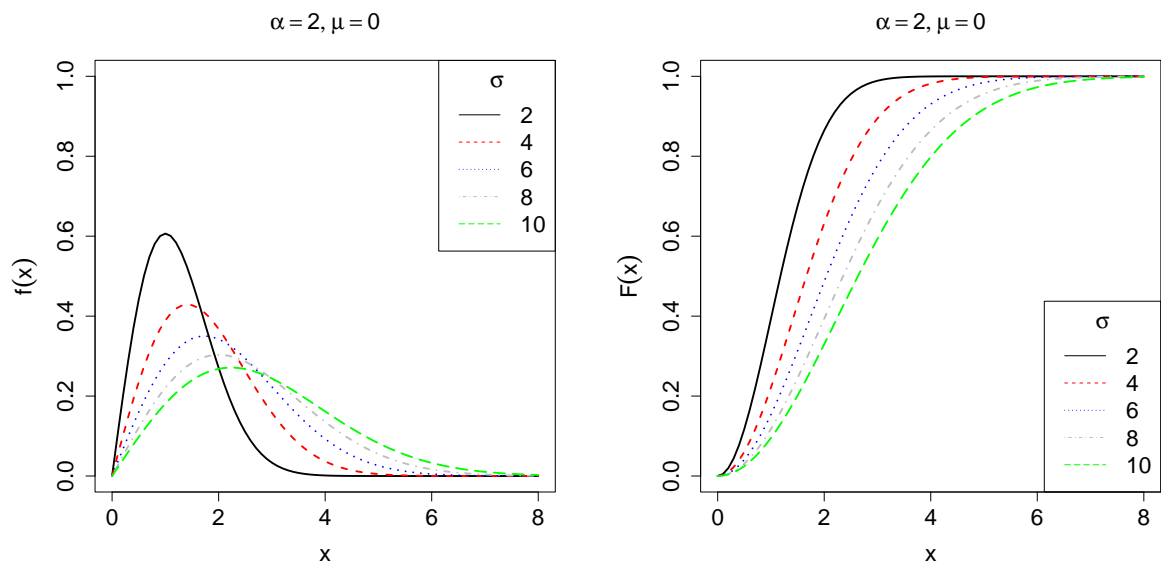
As Figuras 2.3-2.5 mostram o comportamento de  $f(x; \boldsymbol{\theta})$  e  $F(x; \boldsymbol{\theta})$  considerando diferentes valores para os parâmetros de interesse.



**Figura 2.2:** Gráfico da FDP  $f$  (esquerda) e FDA  $F$  (direita) com variação do parâmetro de forma  $0 < \alpha < 1$ .

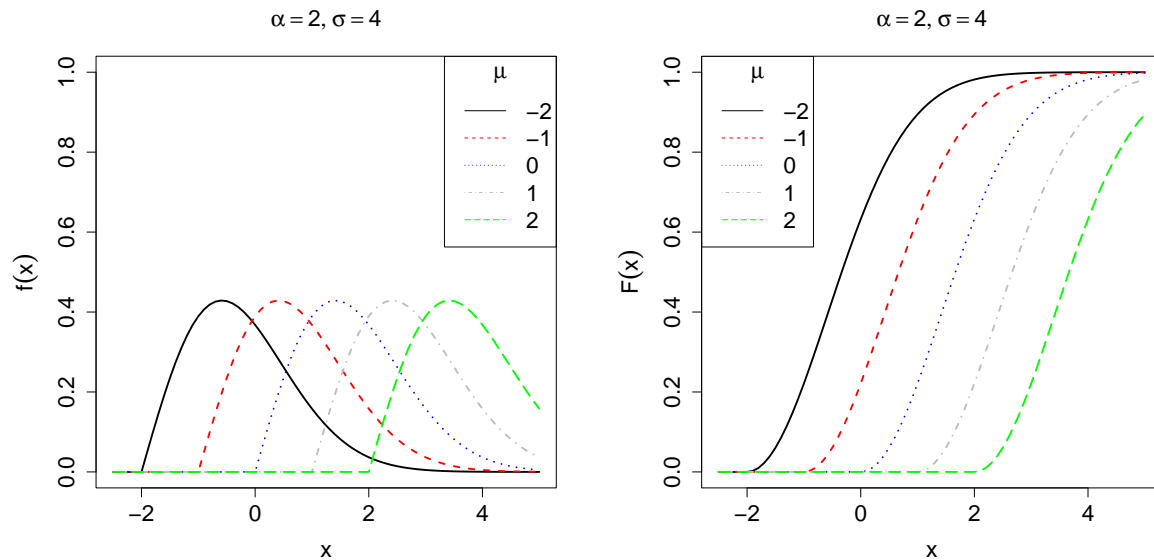


**Figura 2.3:** Gráfico da FDP  $f$  (esquerda) e FDA  $F$  (direita) com variação do parâmetro de forma  $\alpha$ .



**Figura 2.4:** Gráfico da FDP  $f$  (esquerda) e FDA  $F$  (direita) com variação do parâmetro de escala  $\sigma$ .

Conforme apresentado em Johnson e Kotz (1970) e ilustrado nas Figuras 2.3-2.5 os parâmetros  $\alpha$  (forma),  $\sigma$  (escala) e  $\mu$  (locação) alteram a distribuição do seguinte modo:



**Figura 2.5:** Gráfico da FDP  $f$  (esquerda) e FDA  $F$  (direita) com variação do parâmetro de locação  $\mu$ .

- **Forma ( $\alpha$ ):** define a forma da distribuição.
  - $\alpha < 1$ : torna a distribuição mais assimétrica e concentrada em valores pequenos, com uma cauda pesada à direita, indicando uma taxa de falha decrescente, muito usada para modelar sobrevivência.
  - $\alpha = 1$ : reduz-se à distribuição exponencial, indicando um decaimento exponencial da densidade com uma taxa de falha constante.
  - $\alpha > 1$ : a função de densidade tem um pico e depois decresce rapidamente, caracterizando cauda leve, o que significa que há um valor mais provável ao invés de apenas um decaimento exponencial, indicando uma taxa de falha crescente, muito usada para modelar "desgaste" de produtos.
- **Escala ( $\sigma$ ):** ajusta a dispersão dos dados ao longo do eixo  $x$ , quanto menor valor de  $\sigma$  maior assimetria e vice-versa. A unidade de medida da escala é a mesma da variável aleatória  $x$ .

- **Locação ( $\mu$ ):** define a posição da distribuição, ou seja, desloca a distribuição ao longo do eixo  $x$ .

**Observação 2.4.3.** Essas características também podem ser observadas por meio da função de risco.

A seguir apresentamos, nesta seção, as propriedades probabilísticas mais relevantes relacionadas à distribuição Weibull. As expressões encontradas para a esperança, variância, momentos, dentre outros, são detalhados em diversas referências. Para este estudo utilizamos como base o conteúdo exposto em Johnson e Kotz (1970).

### 2.4.1 Esperança

A esperança  $\mathbb{E}(X)$  de uma variável aleatória  $X$  com distribuição  $WB(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta}^\top = (\mu, \alpha, \sigma)$ , é dada por:

$$\mathbb{E}(X) = \int_{\mu}^{\infty} x f(x; \boldsymbol{\theta}) dx = \int_{\mu}^{\infty} x \frac{\alpha}{\sigma} (x - \mu)^{\alpha-1} e^{-\frac{1}{\sigma}(x-\mu)^{\alpha}} dx.$$

Fazendo a substituição  $u = \frac{1}{\sigma} (x - \mu)^{\alpha}$ , podemos reescrever a integral como:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} (\mu + \sigma^{\frac{1}{\alpha}} u^{\frac{1}{\alpha}}) e^{-u} du \\ &= \mu \int_0^{\infty} e^{-u} du + \sigma^{\frac{1}{\alpha}} \int_0^{\infty} u^{\frac{1}{\alpha}} e^{-u} du \\ &= \mu + \sigma^{\frac{1}{\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right), \end{aligned} \tag{2.10}$$

onde  $\Gamma(t) := \int_0^{\infty} u^{t-1} e^{-u} du$  denota a função Gama.

**Observação 2.4.4.** Note que o parâmetro de locação  $\mu$  é diferente da média da distribuição, ou seja,  $\mathbb{E}(X) \neq \mu$ .

### 2.4.2 Variância

A variância pode ser calculada por:

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2. \quad (2.11)$$

Como  $\mathbb{E}(X)$  foi obtida em (2.10), resta calcular  $\mathbb{E}(X^2)$ . Fazendo uma mudança de variável semelhante à usada para calcular  $\mathbb{E}(X)$ , obtem-se:

$$\mathbb{E}(X^2) = \mu^2 + 2\mu\sigma^{\frac{1}{\alpha}}\Gamma\left(1 + \frac{1}{\alpha}\right) + \sigma^{\frac{2}{\alpha}}\Gamma\left(1 + \frac{2}{\alpha}\right).$$

Substituindo  $\mathbb{E}(X)$  e  $\mathbb{E}(X^2)$  na Equação (2.11):

$$\begin{aligned} \text{Var}(X) &= \left[ \mu^2 + 2\mu\sigma^{\frac{1}{\alpha}}\Gamma\left(1 + \frac{1}{\alpha}\right) + \sigma^{\frac{2}{\alpha}}\Gamma\left(1 + \frac{2}{\alpha}\right) \right] - \left[ \mu + \sigma^{\frac{1}{\alpha}}\Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \\ &= \sigma^{\frac{2}{\alpha}} \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left( \Gamma\left(1 + \frac{1}{\alpha}\right) \right)^2 \right]. \end{aligned} \quad (2.12)$$

### 2.4.3 Momentos

Note que, para todo  $m \in \mathbb{N}$ ,

$$\mathbb{E}(X^m) = \int_{\mu}^{\infty} x^m \frac{\alpha}{\sigma} (x - \mu)^{\alpha-1} e^{-\frac{1}{\sigma}(x-\mu)^{\alpha}} dx.$$

Ao substituir  $u = \frac{1}{\sigma}(x - \mu)^{\alpha}$ , obtemos:

$$\mathbb{E}(X) = \int_0^{\infty} (\mu + \sigma^{\frac{1}{\alpha}} u^{\frac{1}{\alpha}})^m e^{-u} du.$$

Usando o binômio de Newton, obtemos:

$$\begin{aligned}\mathbb{E}(X^m) &= \sum_{k=0}^m \binom{m}{k} \int_0^\infty \mu^{m-k} \sigma^{\frac{k}{\alpha}} u^{\frac{k}{\alpha}} e^{-u} du \\ &= \sum_{k=0}^m \binom{m}{k} \mu^{m-k} \sigma^{\frac{k}{\alpha}} \int_0^\infty u^{\frac{k}{\alpha}+1-1} e^{-u} du \\ &= \sum_{k=0}^m \binom{m}{k} \mu^{m-k} \sigma^{\frac{k}{\alpha}} \Gamma\left(\frac{k}{\alpha} + 1\right).\end{aligned}$$

#### 2.4.4 Função Quantil

A função quantil,  $Q(p)$ , é a inversa da FDA (2.7). Dado um percentil  $p \in [0, 1)$ , ou seja,  $F(x_p) = p$ , a função quantil nos dá o valor de  $x_p$  correspondente a esse percentil. Temos:

$$p = 1 - \exp\left(-\frac{1}{\sigma}(x_p - \mu)^\alpha\right).$$

Ajustando a equação e aplicando logaritmo em ambos os lados, obtemos

$$\log(1 - p) = \left(-\frac{1}{\sigma}(x_p - \mu)^\alpha\right).$$

Multiplicando por -1 e elevando ambos os lados à  $\frac{1}{\alpha}$ :

$$(-\sigma \log(1 - p))^{1/\alpha} = x_p - \mu.$$

Finalmente, isolando  $x_p$ , temos:

$$Q(p) = x_p = \mu + \sigma^{\frac{1}{\alpha}} (-\log(1 - p))^{\frac{1}{\alpha}}, \quad (2.13)$$

com  $0 \leq p < 1$ .

### 2.4.5 Função Geradora de Momentos

Vamos obter a função geradora de momentos de  $X \sim \text{WB}(\theta)$ . Temos que, para  $t \in \mathbb{R}$ ,

$$M(t) = \mathbb{E}(e^{tX}) = \int_{\mu}^{\infty} e^{tx} \frac{\alpha}{\sigma} (x - \mu)^{\alpha-1} \exp\left(-\frac{1}{\sigma}(x - \mu)^{\alpha}\right) dx.$$

Fazendo a substituição  $u = \frac{(x-\mu)^{\alpha}}{\sigma}$ , obtemos:

$$M(t) = \int_0^{\infty} e^{t\mu + (\sigma u)^{\frac{1}{\alpha}}} e^{-u} du = e^{t\mu} \int_0^{\infty} e^{(t\sigma^{\frac{1}{\alpha}})u^{\frac{1}{\alpha}}} e^{-u} du.$$

Usando o fato de que  $e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}$ , obtemos

$$\begin{aligned} M(t) &= e^{t\mu} \int_0^{\infty} \left[ \sum_{k=0}^{\infty} \frac{(t\sigma^{\frac{1}{\alpha}})^k u^{\frac{k}{\alpha}}}{k!} \right] e^{-u} du \\ &= e^{t\mu} \sum_{k=0}^{\infty} \frac{t^k \sigma^{\frac{k}{\alpha}}}{k!} \int_0^{\infty} u^{(\frac{k}{\alpha}+1)-1} e^{-u} du \\ &= e^{t\mu} \sum_{k=0}^{\infty} \frac{t^k \sigma^{\frac{k}{\alpha}}}{k!} \Gamma\left(\frac{k}{\alpha} + 1\right). \end{aligned}$$

### 2.4.6 Função Característica

Considerando as respectivas alterações necessárias no domínio e no contradomínio da função geradora de momentos (2.14), obtemos que a função característica da distribuição  $WB(\theta)$  é

$$\varphi(t) = \mathbb{E}(e^{itX}) = e^{it\mu} \sum_{k=0}^{\infty} \frac{it^k \sigma^{\frac{k}{\alpha}}}{k!} \Gamma\left(\frac{k}{\alpha} + 1\right),$$

em que  $i^2 = -1$ .



## 2.5 Métodos de Estimação dos Parâmetros

Muitos autores vem discutindo o problema da maximização da função de verossimilhança comumente observada em distribuições com parâmetros de locação e estatísticas de ordem, como a Exponencial, Weibull, Gamma (e suas generalizações), dentre outras. Os estudos propostos por Kundu e Raqab (2009), Ali et al. (2012), Kundu e Raqab (2013) e Kohansal e Rezakhah (2019), por exemplo, excluem a menor observação do conjunto de dados como alternativa para garantir que a função de verossimilhança seja maximizada (finitude dos EMVs).

No entanto, a exclusão de observações pode, em parte, significar perda de informação. Neste ponto de vista, a abordagem proposta por Cheng e Iles (1987) e Cheng e Traylor (1995) pode, efetivamente, contornar as duas questões discutidas anteriormente.

Suponha que  $\{X_1, \dots, X_n\}$  são v.a.'s i.i.d. com distribuição  $WB(\theta)$ , cujo vetor de parâmetros é  $\theta$ . Sua função de verossimilhança correspondente é o produto das funções densidade de probabilidade (2.8) para todos os  $n$  valores observados da amostra:

$$\begin{aligned} \mathcal{L}(\theta; X_1, X_2, \dots, X_n) &= \prod_{i=1}^n \frac{\alpha}{\sigma} \left( \frac{X_i - \mu}{\sigma} \right)^{\alpha-1} \exp \left[ - \left( \frac{X_i - \mu}{\sigma} \right)^{\alpha} \right] \mathbb{1}_{(\mu, +\infty)}(X_i) \\ &= \left( \frac{\alpha}{\sigma} \right)^n \exp \left( - \frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu)^{\alpha} \right) \prod_{i=1}^n (X_i - \mu)^{\alpha-1} \mathbb{1}_{(\mu, +\infty)}(X_i). \end{aligned} \quad (2.14)$$

Pela função de verossimilhança dada em (2.14), obtemos que o EMV do parâmetro de locação  $\mu$  pode ser representado por

$$\hat{\mu} = \min \{X_1, X_2, \dots, X_n\} \stackrel{not.}{=} X_{(1)}.$$

Dessa forma, o parâmetro de localização  $\mu$  pode ser considerado um estimador natural e consistente do valor mínimo da amostra porque representa o ponto a partir do qual as observações começam na distribuição Weibull. O valor mínimo observado da amostra é uma estimativa

intuitiva para  $\mu$  e converge para o verdadeiro valor de  $\mu$  à medida que o tamanho da amostra aumenta.

Note que, se  $\alpha < 1$ , conforme  $\mu$  se aproxima de  $X_{(1)}$ , a função de verossimilhança em (2.14) aumenta gradualmente até o infinito. Consequentemente, os EMVs de  $\theta$  não existem.

Uma abordagem comum na literatura para se modelar dados com suporte positivo é considerar  $\mu = 0$ . Assim, a distribuição Weibull se torna um modelo regular e os EMVs de  $(\alpha, \beta)$  podem ser obtidos por métodos usuais.

Com o objetivo de modelagem de dados com o suporte (possivelmente) real, nas subseções a seguir apresentamos, respectivamente, os métodos de correção da função de verossimilhança propostos por Kundu e Raqab (2009) e Cheng e Iles (1987), sem a suposição de que o parâmetro de locação  $\mu$  seja nulo.

### 2.5.1 EMVs Modificados

Considerando  $\mu < X_{(1)}$ , e aplicando o logaritmo à função de verossimilhança apresentada em (2.14) e que o produto das funções indicadoras são iguais a 1, temos que:

$$\begin{aligned}
 \ell(\theta) &= \log \mathcal{L}(\theta; X_1, \dots, X_n) \\
 &= \log \left( \prod_{i=1}^n \frac{\alpha}{\sigma} \left( \frac{X_i - \mu}{\sigma} \right)^{\alpha-1} \exp \left[ - \left( \frac{X_i - \mu}{\sigma} \right)^\alpha \right] \right) \\
 &= \sum_{i=1}^n \log \left( \frac{\alpha}{\sigma} \left( \frac{X_i - \mu}{\sigma} \right)^{\alpha-1} \exp \left[ - \left( \frac{X_i - \mu}{\sigma} \right)^\alpha \right] \right) \\
 &= n \log \frac{\alpha}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu)^\alpha + (\alpha - 1) \sum_{i=1}^n \log (x_i - \mu). \quad (2.15)
 \end{aligned}$$

Por simplicidade de notações futuras, reescrevemos o vetor de parâmetros da seguinte forma:  $\theta^\top = (\mu, \phi^\top)$ , onde  $\phi^\top = (\alpha, \sigma)$ .

Vamos obter o *logaritmo da função de verossimilhança modificada* usando os mesmos pas-

sos de Kundu e Raqab (2009). Para garantir a existência e a finitude dos EMVs para o vetor  $\theta$ , o logaritmo da função de verossimilhança modificada é baseado em  $(n - 1)$  observações após ignorar a menor observação e substituir  $\mu$  pelo seu estimador natural  $\hat{\mu} = X_{(1)}$ . Fazendo as substituições em (2.15) temos:

$$\ell_M(\hat{\mu}, \phi) = (n-1) \log \alpha - (n-1) \log \sigma - \frac{1}{\sigma} \sum_{i=2}^n (X_i - \hat{\mu})^\alpha + (\alpha-1) \sum_{i=2}^n \log (X_i - \hat{\mu}). \quad (2.16)$$

Após a exclusão de  $X_1$ , as expressões que apresentamos a seguir para o vetor escore e a matriz de informação, dependem apenas da soma das v.a.'s restantes. Dessa forma, por simplicidade de notação, em (2.16) consideramos que  $X_1 = X_{(1)}$ .

Por conseguinte, o *vetor escore modificado* é

$$S_M(\phi) = (\partial_\alpha \ell_M(\hat{\mu}, \phi), \partial_\sigma \ell_M(\hat{\mu}, \phi))^\top, \quad (2.17)$$

onde  $\partial_\alpha \ell_M(\hat{\mu}, \phi)$  e  $\partial_\sigma \ell_M(\hat{\mu}, \phi)$  denotam as derivadas parciais de  $\ell_M(\hat{\mu}, \phi)$  em relação a  $\alpha$  e  $\sigma$ , respectivamente. Note que,

$$\partial_\alpha \ell_M(\hat{\mu}, \phi) = \frac{(n-1)}{\alpha} + \frac{1}{\alpha} \sum_{i=2}^n \log(X_i - \hat{\mu})^\alpha - \frac{1}{\sigma} \sum_{i=2}^n (X_i - \hat{\mu})^\alpha \log(X_i - \hat{\mu}) \quad (2.18)$$

e

$$\partial_\sigma \ell_M(\hat{\mu}, \phi) = \frac{-(n-1)}{\sigma} + \frac{1}{\sigma^2} \sum_{i=2}^n (X_i - \hat{\mu})^\alpha = 0. \quad (2.19)$$

Segue de (2.19) que

$$\frac{1}{\sigma^2} \sum_{i=2}^n (X_i - \hat{\mu})^\alpha = \frac{n-1}{\sigma},$$

logo

$$\hat{\sigma} = \hat{\sigma}(\alpha) = \sum_{i=2}^n \frac{(X_i - \hat{\mu})^\alpha}{n-1}.$$

Por outro lado, o EMV para  $\alpha$  não tem forma fechada. Dessa forma, métodos numéricos po-

dem ser usados para encontrar  $\hat{\alpha}$ . Por exemplo, o Método do Ponto Fixo é uma das possíveis alternativas, veja em Burden et al. (2015).

### Matriz de Informação Modificada

Uma vez concluídas as etapas iniciais, podemos obter a matriz de informação esperada para  $\phi$ , representada por  $\mathcal{I}(\phi)$ . Isso é possível calculando as derivadas de segunda ordem (ou cumulantes). Representaremos por  $k_{rs}$  a entrada genérica  $(r, s)$  da matriz  $\mathcal{I}(\phi)$ , de modo que,

$$k_{rs} = \mathbb{E} \left\{ -\frac{\partial^2 \ell_M(\phi)}{\partial \phi_r \partial \phi_s} \right\} = \mathbb{E} \left\{ \left( \frac{\partial \ell_M(\phi)}{\partial \phi_r} \right) \left( \frac{\partial \ell_M(\phi)}{\partial \phi_s} \right) \right\}, \text{ onde } r, s \in \{\alpha, \sigma\}.$$

Temos o seguinte resultado:

**Proposição 2.5.1.** *Seja  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $WB(\theta)$ . Suponha que o parâmetro de locação  $\mu$  é conhecido. Então, a matriz de informação de Fisher associada à função de verossimilhança modificada (2.16) é dada por*

$$\mathcal{I}(\phi) = \begin{pmatrix} k_{\alpha\alpha} & k_{\alpha\sigma} \\ k_{\sigma\alpha} & k_{\sigma\sigma} \end{pmatrix}, \quad (2.20)$$

em que

$$k_{\alpha\alpha} = \frac{n-1}{\alpha^2} \{ (1 + \Gamma''(2)) + 2 \log(\sigma) \Gamma'(2) + \log^2(\sigma) \}, \quad (2.21)$$

$$k_{\alpha\sigma} = k_{\sigma\alpha} = -\frac{1}{\alpha\sigma} (n-1) \Gamma'(2) \log(\sigma), \quad (2.22)$$

$$k_{\sigma\sigma} = \frac{n-1}{\sigma^2}. \quad (2.23)$$

**Prova:** Queremos provar (2.21)-(2.23). Primeiramente, considere  $Z$  uma v.a. com distribuição  $\text{Gamma}(a, 1)$ . Afirmamos que se  $a = 2$ , então

$$E(\log Z) = \Gamma'(2), \quad (2.24)$$

e

$$E(\log^2 Z) = \Gamma''(2). \quad (2.25)$$

De fato, lembre que  $\frac{\partial}{\partial a} z^{a-1} = z^{a-1} \log z$ . Então,

$$\begin{aligned} \mathbb{E}(\log Z) &= \int_0^\infty \log z \frac{1}{\Gamma(a)} z^{a-1} e^{-z} dz \\ &= \frac{1}{\Gamma(a)} \int_0^\infty \frac{\partial}{\partial a} z^{a-1} e^{-z} dz \\ &= \frac{1}{\Gamma(a)} \frac{\partial}{\partial a} \int_0^\infty z^{a-1} e^{-z} dz \\ &= \frac{\Gamma'(a)}{\Gamma(a)}. \end{aligned} \quad (2.26)$$

Em particular, se  $a = 2$ , então  $\Gamma(2) = 1$ , dessa forma (2.24) está provado. De maneira análoga, é possível mostrar a validade de (2.25). Agora, derivando (2.18) em relação à  $\alpha$ , obtemos

$$\begin{aligned} k_{\alpha\alpha} &= \mathbb{E} \left\{ -\frac{\partial^2 \ell_M(\phi)}{\partial \alpha^2} \right\} \\ &= \mathbb{E} \left\{ \frac{(n-1)}{\alpha^2} + \frac{1}{\sigma} \sum_{i=2}^n (X_i - \mu)^\alpha [\log(X_i - \mu)]^2 \right\} \\ &= \frac{n-1}{\alpha^2} + \frac{1}{\sigma} \sum_{i=2}^n \mathbb{E} \left\{ (X_i - \mu)^\alpha [\log(X_i - \mu)]^2 \right\}. \end{aligned} \quad (2.27)$$

Temos que

$$\begin{aligned} \mathbb{E} [(X_i - \mu)^\alpha \log^2(X_i - \mu)] &= \int_\mu^\infty (x - \mu)^\alpha \log^2(x - \mu) f(x; \boldsymbol{\theta}) dx \\ &= \int_\mu^\infty (x - \mu)^\alpha \log^2(x - \mu) \frac{\alpha}{\sigma} \exp\left(-\frac{1}{\sigma}(x - \mu)^\alpha\right) (x - \mu)^{\alpha-1} dx. \end{aligned}$$

Substituindo  $t = \frac{(x-\mu)^\alpha}{\sigma}$ , obtemos

$$\mathbb{E}[(X_i - \mu)^\alpha \log^2(X_i - \mu)] = \int_0^\infty t\sigma \frac{1}{\alpha^2} [\log(t\sigma)]^2 e^{-t} dt.$$

Após simplificações, pode-se mostrar que

$$\begin{aligned} \mathbb{E}[(X_i - \mu)^\alpha \log^2(X_i - \mu)] &= \frac{\sigma}{\alpha^2} \left\{ \log^2 \sigma + 2 \log(\sigma) \int_0^\infty \log(t) t e^{-t} dt + \int_0^\infty \log^2(t) t e^{-t} dt \right\} \\ &= \frac{\sigma}{\alpha^2} \left\{ \log^2 \sigma + 2 \log(\sigma) \mathbb{E}[\log(Z)] + \mathbb{E}[\log^2(Z)] \right\}, \end{aligned}$$

em que  $Z$  tem distribuição *Gama*(2, 1). Dessa forma, segue de (2.24)-(2.28) que,

$$\begin{aligned} k_{\alpha\alpha} &= \frac{n-1}{\alpha^2} + \frac{1}{\alpha^2} \sum_{i=2}^n \left\{ \log^2 \sigma + 2(\log \sigma) \Gamma'(2) + \Gamma''(2) \right\} \\ &= \frac{n-1}{\alpha^2} \left\{ (1 + \Gamma''(2)) + \log^2(\sigma) + 2(\log \sigma) \Gamma'(2) \right\}, \end{aligned} \quad (2.28)$$

o que prova (2.21).

Para calcular  $k_{\sigma\sigma}$  seguimos o mesmo raciocínio anterior

$$\begin{aligned} k_{\sigma\sigma} &= \mathbb{E} \left\{ -\frac{\partial^2 \ell_M(\phi)}{\partial \sigma^2} \right\} \\ &= \frac{2}{\sigma^3} \sum_{i=2}^n \mathbb{E} \{ (X_i - \mu)^\alpha \} - \frac{n-1}{\sigma^2}. \end{aligned}$$

Temos que

$$\mathbb{E} \{ (X_i - \mu)^\alpha \} = \int_\mu^\infty (x - \mu)^\alpha \frac{\alpha}{\sigma} (x - \mu)^{\alpha-1} \exp \left( -\frac{(x - \mu)^\alpha}{\sigma} \right) dx = \sigma \int_0^\infty t e^{-t} dt = \sigma,$$

onde na segunda igualdade utilizamos novamente a mudança de variáveis  $t = \frac{(x-\mu)^\alpha}{\sigma}$ . Logo,

$$k_{\sigma\sigma} = \frac{2}{\sigma^3} \sum_{i=2}^n (X_{(i)} - \hat{\mu})^\alpha - \frac{(n-1)}{\sigma^2} = \frac{2\sigma(n-1)}{\sigma^3} - \frac{(n-1)}{\sigma^2} = \frac{n-1}{\sigma^2},$$

o que prova (2.23). Por fim, segue de (2.18) e (2.19) que

$$\begin{aligned} \partial_\alpha \{ \partial_\sigma \ell_M(\phi) \} &= \partial_\sigma \{ \partial_\alpha \ell_M(\phi) \} \\ &= \frac{1}{\sigma^2} \sum_{i=2}^n (X_i - \mu)^\alpha \log(X_i - \mu), \end{aligned}$$

de modo que

$$\begin{aligned} k_{\alpha\sigma} = k_{\sigma\alpha} &= \mathbb{E} \left\{ -\frac{\partial^2 \ell_M(\phi)}{\partial \sigma \partial \alpha} \right\} \\ &= -\frac{1}{\sigma^2} \sum_{i=2}^n \mathbb{E} \{ (X_i - \mu)^\alpha \log(X_i - \mu) \} \\ &= -\frac{1}{\alpha\sigma} (n-1) \Gamma'(2) \log(\sigma), \end{aligned}$$

onde a terceira igualdade foi obtida seguindo os mesmos passos anteriores, com isso concluindo a proposição. ■

## 2.5.2 EMVs Corrigidos

Ainda visando os problemas das distribuições não-regulares, será apresentada mais uma técnica de correção da função de verossimilhança proposta por Cheng e Iles (1987). Essa correção ajusta a contribuição da menor observação, que pode ser particularmente problemática em distribuições com suporte dependente de parâmetros ou outras características não-regulares, como já discutido na Seção 2.5.1. A função de verossimilhança corrigida  $\tilde{\ell}(\theta, h)$  é definida como:

$$\tilde{\ell}(\boldsymbol{\theta}, h) = \log \int_{X_{(1)}}^{X_{(1)}+h} f(x; \boldsymbol{\theta}) dx + \sum_{i=2}^n \log f(X_i; \boldsymbol{\theta}), \quad (2.29)$$

em que  $X_{(1)}$  é a menor observação no conjunto de dados,  $h$  é um parâmetro de correção,  $f(x; \boldsymbol{\theta})$  é a função densidade de probabilidade, e  $\boldsymbol{\theta}$  é o vetor de parâmetros.

A correção proposta em 2.29 pode ser interpretada da seguinte forma:

- O primeiro termo,  $\log \int_{X_{(1)}}^{X_{(1)}+h} f(x; \boldsymbol{\theta}) dx$ , integra a densidade de probabilidade  $f(x; \boldsymbol{\theta})$  no intervalo  $[X_{(1)}, X_{(1)} + h]$ , ajustando a contribuição da menor observação  $X_{(1)}$ .
- O segundo termo,  $\sum_{i=2}^n \log f(X_i; \boldsymbol{\theta})$ , é a soma dos logaritmos da função de verossimilhança individuais das demais observações  $X_i$  para  $i = 2, \dots, n$ .

Ao integrar a densidade de probabilidade sobre um pequeno intervalo  $[X_{(1)}, X_{(1)}+h]$ , a correção reduz o impacto de qualquer singularidade ou anomalia associada à menor observação.

Alguns dos benefícios desta abordagem, destacados por Cheng e Iles (1987), são:

- Redução do viés dos EMVs;
- Maior robustez dos estimadores em problemas não-regulares;
- Melhor consistência dos estimadores em situações onde a função de verossimilhança padrão pode falhar.

Além disso, esta correção ajusta a função de verossimilhança de modo a obter estimadores mais robustos e menos viesados. Cheng e Iles (1987) descreveram que a técnica de correção do log da função de verossimilhança pode ser estendida para outras distribuições como a Gama, a Beta e a Log-normal, mantendo as propriedades assintóticas esperadas e estimativas de máxima verossimilhança consistentes. Este método também foi apresentado nos estudos de Smith (1990), Jones e Smith (1991) e Brown e Zhao (1996).



# Capítulo 3

## Redução de Viés nos EMVMs

### 3.1 Introdução

Neste capítulo, propomos um novo procedimento de estimação para os parâmetros de forma e de escala, baseados em uma amostra extraída de uma distribuição Weibull. No Capítulo 2, apresentamos os EMVs modificados (EMVMs) propostos por Kundu e Raqab (2009). Para isso, foram apresentados o logaritmo da função de verossimilhança modificada  $\ell_M(\phi)$  (2.16), o vetor *escore* modificado  $S_M(\phi)$  (2.17) e a matriz de informação de Fisher (modificada)  $\mathcal{I}(\phi)$  (2.20). Na Seção 3.2, apresentamos o método de penalização da função de verossimilhança proposto por Firth (1993). Tal método é a base para a construção da nova classe de estimadores proposta. No Capítulo 4, um estudo de simulações é apresentado para avaliar a qualidade dos novos estimadores. Os resultados obtidos via simulações de Monte Carlo são comparados com estimações feitas via EMVMs (Seção 2.5.1) e EMVCs (Seção 2.5.2).

### 3.2 EMVs Duplamente Modificados

O propósito dessa seção é propor um estimador baseado na penalização do vetor *escore* modificado dado em (2.17). Para isso, usaremos o método de Firth (1993) que remove o viés de primeira ordem dos EMVs por uma adequada penalização da função *escore*.

Seja  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $WB(\theta)$ ,  $\theta^\top = (\mu, \phi^\top)$  com  $\phi^\top = (\alpha, \sigma)$ . Tome o estimador natural de  $\mu$  como sendo o menor valor da amostra  $\hat{\mu} = X_{(1)}$ .

Considere o vetor *escore* modificado  $S_M(\phi)$  (2.17). Definimos o vetor *escore duplamente modificado* da seguinte forma:

$$S_M^*(\phi_j) := S_M(\phi_j) + A(\phi_j) \quad \text{para } \phi_j \in \{\alpha, \sigma\}, j = 1, 2, \quad (3.1)$$

onde  $A(\phi_j)$  é a penalidade de  $\phi_j$  e será apresentada adiante. Então, os *EMVs duplamente modificados* (EMVDMs),  $\hat{\phi}_j^*$ , podem ser obtidos resolvendo a seguinte equação:

$$S_M^*(\phi_j) = 0, \quad j = 1, 2. \quad (3.2)$$

**Observação 3.2.1.**  $S_M^*(\phi_j)$  representa a  $j$ -ésima entrada de  $S_M^*(\phi)$ , e  $A(\phi)$  é o vetor das penalidades, sendo da ordem  $\mathcal{O}(1)$  quando  $n \rightarrow \infty$ .

Na proposta de Firth (1993), os elementos de  $A(\phi_j)$  são obtidos derivando as entradas genéricas da matriz de informação de Fisher do seguinte modo:

$$A(\phi_j) = \frac{1}{2} \sum_{r,s \in \{\alpha, \sigma\}} k^{rs} k_{r,s,j} = \frac{1}{2} \text{tr} \left\{ \mathcal{I}^{-1}(\phi) \left( \frac{\partial \mathcal{I}(\phi)}{\partial \phi_j} \right) \right\}, \quad (3.3)$$

onde  $k^{rs}$  representa as entradas genéricas  $(r, s)$  de  $\mathcal{I}^{-1}(\phi)$ ,  $k_{r,s,j} = \partial_{\phi_j} k_{rs}$ , para  $k_{rs}$  representando a entrada  $(r, s)$  da matriz  $\mathcal{I}(\phi)$  e  $\text{tr}(M)$  denotando o traço da matriz  $M$ .

Observe que no caso da distribuição Weibull modificada, utilizaremos o logaritmo da função de verossimilhança modificada (2.16) ao invés do logaritmo da função de verossimilhança (2.15).

**Observação 3.2.2.** Os EMVDMs têm as mesmas propriedades teóricas que os EMVs modificados  $\hat{\phi}$ . Além disso, a matriz de covariância de primeira ordem para  $\hat{\phi}^*$  é a mesma obtida para

$\hat{\phi}$ , com  $\hat{\phi}^*$  substituindo  $\hat{\phi}$ . Consequentemente,  $\hat{\phi}^* \sim \mathcal{N} \left( \phi, \mathcal{I}^{-1} \left( \hat{\phi}^* \right) \right)$  quando  $n \rightarrow \infty$ .

**Observação 3.2.3.** Computacionalmente, podemos obter os EMVDMs via resolução do sistema de equações (3.2) ou via maximização do *logaritmo da função de verossimilhança duplamente modificada*

$$\ell_M^*(\phi) := \ell_M(\hat{\mu}, \phi) + \frac{1}{2} \log \det \mathcal{I}(\phi),$$

em que  $\mathcal{I}(\phi)$ , definida em (2.20), é a matriz de informação, cujo determinante é igual a

$$\det \mathcal{I}(\phi) = k_{\alpha\alpha}k_{\sigma\sigma} - k_{\alpha\sigma}k_{\sigma\alpha}.$$

A seguir, para estimar os parâmetros que resolvem (3.2), são apresentadas as derivadas parciais de terceira ordem da matriz de informação  $\mathcal{I}(\phi)$  com base nos cumulantes apresentados na Subseção 2.5.1. Para isso os elementos  $k_{r,s,j}$ , que compõe (3.3), são obtidos via derivação parcial (com respeito à  $\alpha$  e  $\sigma$ ) dos termos da matriz de informação  $\mathcal{I}(\phi)$  (2.20). Os cálculos são apresentados a seguir:

$$\begin{aligned} k_{\alpha\alpha\alpha} = \partial_\alpha k_{\alpha\alpha} &= \partial_\alpha \left\{ \frac{n-1}{\alpha^2} \left[ (1 + \Gamma''(2)) + 2 \log(\sigma) \Gamma'(2) + \log^2(\sigma) \right] \right\} \\ &= -\frac{2(n-1)}{\alpha^3} \left\{ (1 + \Gamma''(2)) + 2 \log(\sigma) \Gamma'(2) + \log^2(\sigma) \right\}. \end{aligned}$$

$$\begin{aligned} k_{\alpha\sigma\alpha} = k_{\sigma\alpha\alpha} = \partial_\alpha k_{\sigma\alpha} &= \partial_\alpha \left\{ -\frac{(n-1) (\Gamma'(2) \log(\sigma))}{\alpha\sigma} \right\} \\ &= \frac{(n-1) (\Gamma'(2) \log(\sigma))}{\alpha^2\sigma}. \end{aligned}$$

$$k_{\sigma\sigma\alpha} = \partial_\alpha k_{\sigma\sigma} = \partial_\alpha \left\{ \frac{(n-1)}{\sigma^2} \right\} = 0.$$

$$\begin{aligned}
k_{\alpha\alpha\sigma} = \partial_{\sigma} k_{\alpha\alpha} &= \partial_{\sigma} \left\{ \frac{n-1}{\alpha^2} [(1 + \Gamma''(2)) + 2 \log(\sigma) \Gamma'(2) + \log^2(\sigma)] \right\} \\
&= \frac{n-1}{\alpha^2} \left( \frac{2\Gamma'(2)}{\sigma} + \frac{2 \log(\sigma)}{\sigma} \right) \\
&= \frac{2(n-1)}{\alpha^2 \sigma} (\Gamma'(2) + \log(\sigma)) \\
&= \frac{2(n-1)}{\alpha^2 \sigma} (\Gamma'(2) + \log(\sigma)) .
\end{aligned}$$

$$\begin{aligned}
k_{\alpha\sigma\sigma} = k_{\sigma\alpha\sigma} = \partial_{\sigma} k_{\sigma\alpha} &= \partial_{\sigma} \left\{ -\frac{(n-1) (\Gamma'(2) \log(\sigma))}{\alpha \sigma} \right\} \\
&= \frac{(n-1)}{\alpha \sigma^2} (\Gamma'(2) + \log(\sigma)) + \left\{ -\frac{(n-1)}{\alpha \sigma} \frac{1}{\sigma} \right\} \\
&= \frac{(n-1)}{\alpha \sigma^2} (\Gamma'(2) + \log(\sigma)) - \left\{ \frac{(n-1)}{\alpha \sigma^2} \right\} \\
&= \frac{n-1}{\alpha \sigma^2} (\Gamma'(2) + \log(\sigma) - 1) .
\end{aligned}$$

Finalmente,

$$k_{\sigma\sigma\sigma} = \partial_{\sigma} k_{\sigma\sigma} = -\frac{2(n-1)\sigma}{\sigma^4} = -\frac{2(n-1)}{\sigma^3}.$$

Nas próximas seções, avaliaremos a aplicabilidade dos novos estimadores em dados simulados e em dados reais.

# Capítulo 4

## Estudos de Simulação

Os estudos de simulação apresentados nesse capítulo, foram programados no software R Core Team (2024). Avaliamos a estimação dos parâmetros da distribuição Weibull, via simulações de Monte Carlo. As EMVMs, EMVCs e EMVDMs, foram obtidas considerando os seguintes passos no processo de geração dos dados:

- Fixamos um vetor de três parâmetros  $\boldsymbol{\theta}^\top = (\mu, \alpha, \sigma) \in \{(1, 1, 1), (1, 0.5, 2), (1, 1.5, 2)\}$ , onde cada vetor representa um cenário.
- Fixamos o tamanho de amostra  $n \in \{50, 100, 500, 1000\}$ .
- Para cada configuração de  $\boldsymbol{\theta}$  e  $n$  escolhidos, foram geradas  $M = 1000$  replicações de Monte Carlo. Em outras palavras, para  $j = 1, \dots, M$ , geramos uma amostra  $\mathbf{X}^{(j)} = \{X_1^{(j)}, \dots, X_n^{(j)}\}$  da distribuição  $WB(\boldsymbol{\theta})$ . Assim, a  $j$ -ésima estimativa do parâmetro  $\boldsymbol{\theta}$  é obtida com base na amostra  $\mathbf{X}^{(j)}$ , digamos  $\hat{\boldsymbol{\theta}}_j = \hat{\boldsymbol{\theta}}_j(\mathbf{X}^{(j)})$ ;
- O desempenho do modelo em termos de estimação foi avaliado usando as seguintes medidas: Viés Relativo (RB), Raiz do Erro Quadrático Médio (RMSE) e a Probabilidade de Cobertura (CP), que mede a proporção de vezes em que os intervalos de confiança vão conter o verdadeiro valor do parâmetro que está sendo estimado, veja Casella e Berger (2002). Para cada cenário, foi considerado o nível de significância de 5%.

O primeiro cenário  $\theta^\top = (\mu, \alpha, \sigma) = (1, 1, 1)$ , a Weibull se reduz a uma distribuição exponencial, indicando um decaimento exponencial da densidade, devido a  $\alpha < 1$ , com alta frequência de valores pequenos e uma cauda pesada. O segundo cenário  $\theta^\top = (\mu, \alpha, \sigma) = (1, 0.5, 2)$ , em que  $\alpha = 0.5$ , gera valores ainda mais concentrados próximos ao mínimo permitido ( $\mu$ ) e com uma cauda mais longa (pesada). Tomamos por base os valores estimados na análise dos dados de investimento (ver Tabela 5.4). O terceiro cenário  $\theta^\top = (\mu, \alpha, \sigma) = (1, 1.5, 2)$ , em que  $\alpha = 1.5$ , tem um pico na densidade, indicando um valor modal mais definido, e uma cauda mais leve atingindo o zero de forma mais rápida que nos outros cenários.

Os resultados obtidos foram utilizados para comparar os métodos de estimação EMVDM, EMVM e EMVC, propostos por Firth (1993), Kundu e Raqab (2009) e Cheng e Iles (1987), respectivamente.

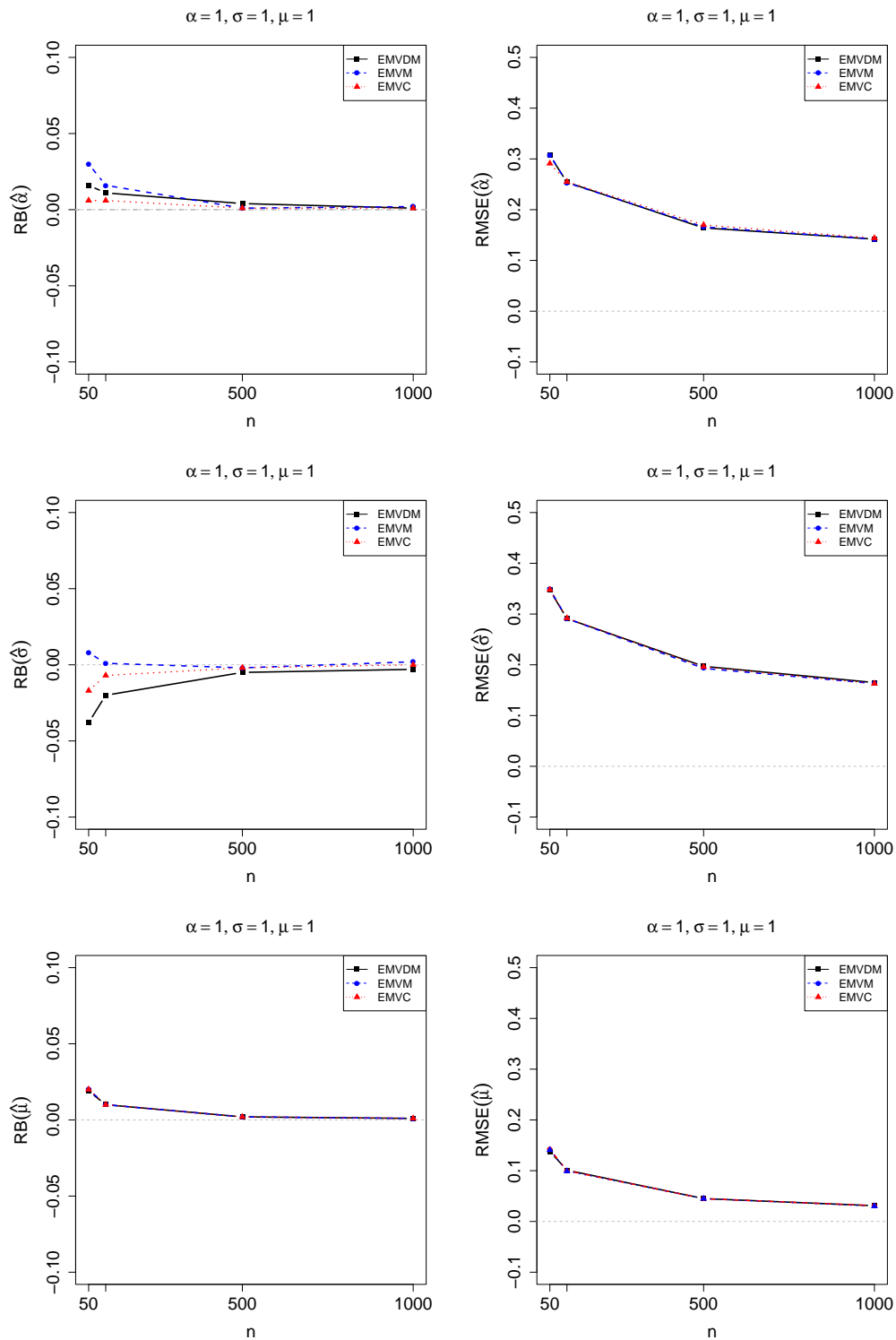
Desta forma, nas Figuras 4.1-4.3 apresentamos os resultados de RB e RMSE. Observe que à medida que o tamanho  $n$  da amostra aumenta, tanto o RB quanto o RMSE se aproximam de zero, indicando que os parâmetros estimados estão mais próximos do real valor dos parâmetros.

Podemos verificar diferenças entre os métodos de estimação para tamanhos de amostra menores. Para amostras maiores que 500, essa diferença no RB e no RMSE é quase nula. Os valores RMSE dos três estimadores são semelhantes.

O EMVDM tende a reduzir o viés em comparação ao EMVM para a configuração de parâmetros apresentada na Figura 4.2, segundo cenário, demonstrando que para distribuições de cauda pesada ( $\alpha < 1$ , o EMVDM se mostra melhor que os demais indicando influência da cauda na estimação dos parâmetros. Já, a comparação do RB dos EMVDMs contra EVMCs, mostrou que, em geral, os estimadores corrigidos apresentaram um viés menor. Nesse caso, pode-se pensar em fazer uma penalização da função log-verossimilhança corrigida (2.29) utilizando novamente o método de Firth, tentando assim, melhorar os resultados na simulação.

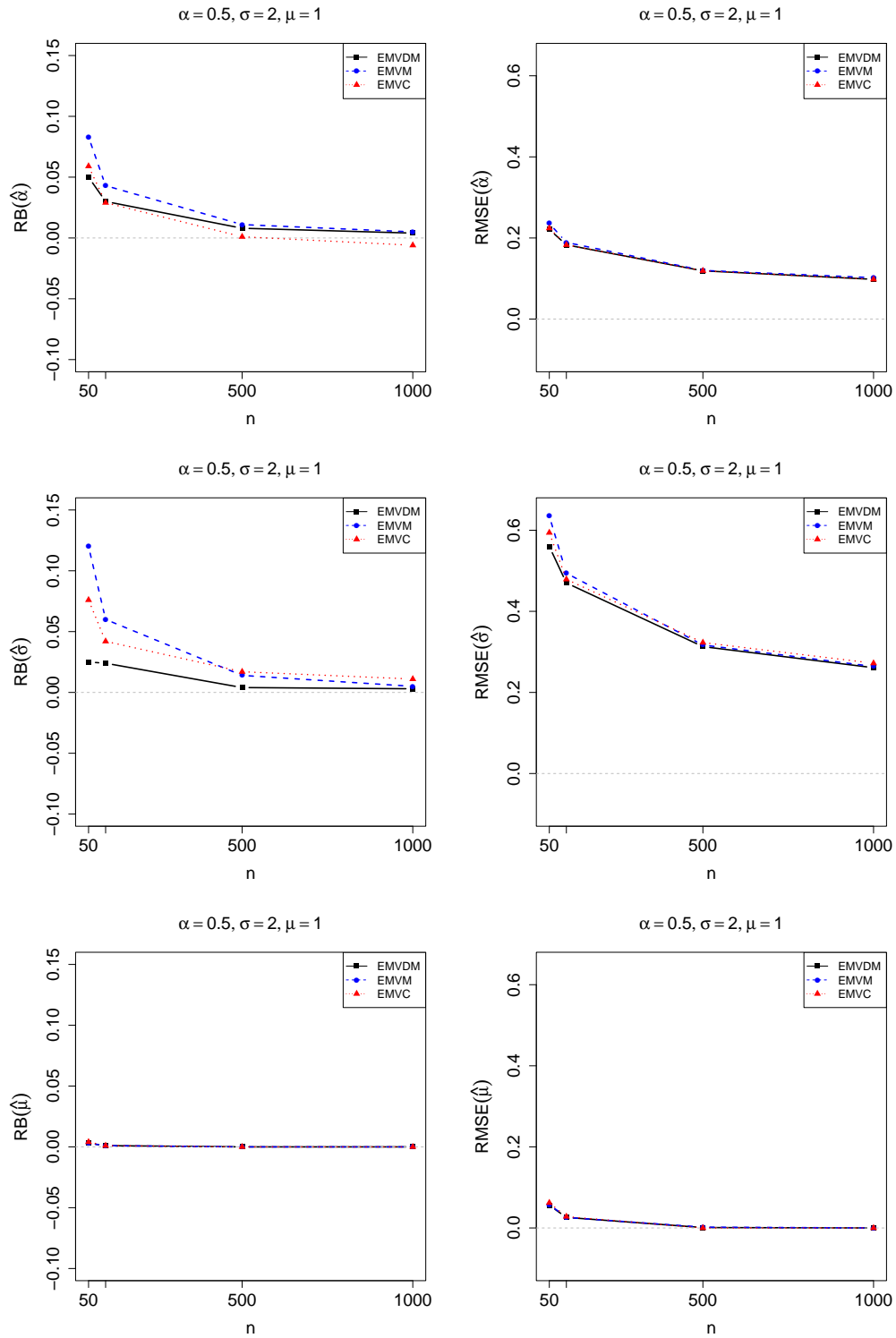
Os valores de CP estão apresentados na Figura 4.4 para o cenário  $\alpha = 1, 5$  e  $\sigma = 2$ , indicam que o EMVDM teve melhor desempenho quando comparado com o desempenho dos demais

estimadores. Porém, para  $\alpha = 0,5$  e  $\sigma = 2$ , teve o pior desempenho em relação aos demais métodos. No cenário  $\alpha = 1$  e  $\sigma = 1$ , para os três métodos, a probabilidade de cobertura fica em torno de 95%, principalmente quando o tamanho da amostra está acima de 500. Em todos os cenários, o CP tende a ser maior que 90%, para todos os métodos de estimação, à medida que o tamanho da amostra  $n$  aumenta.

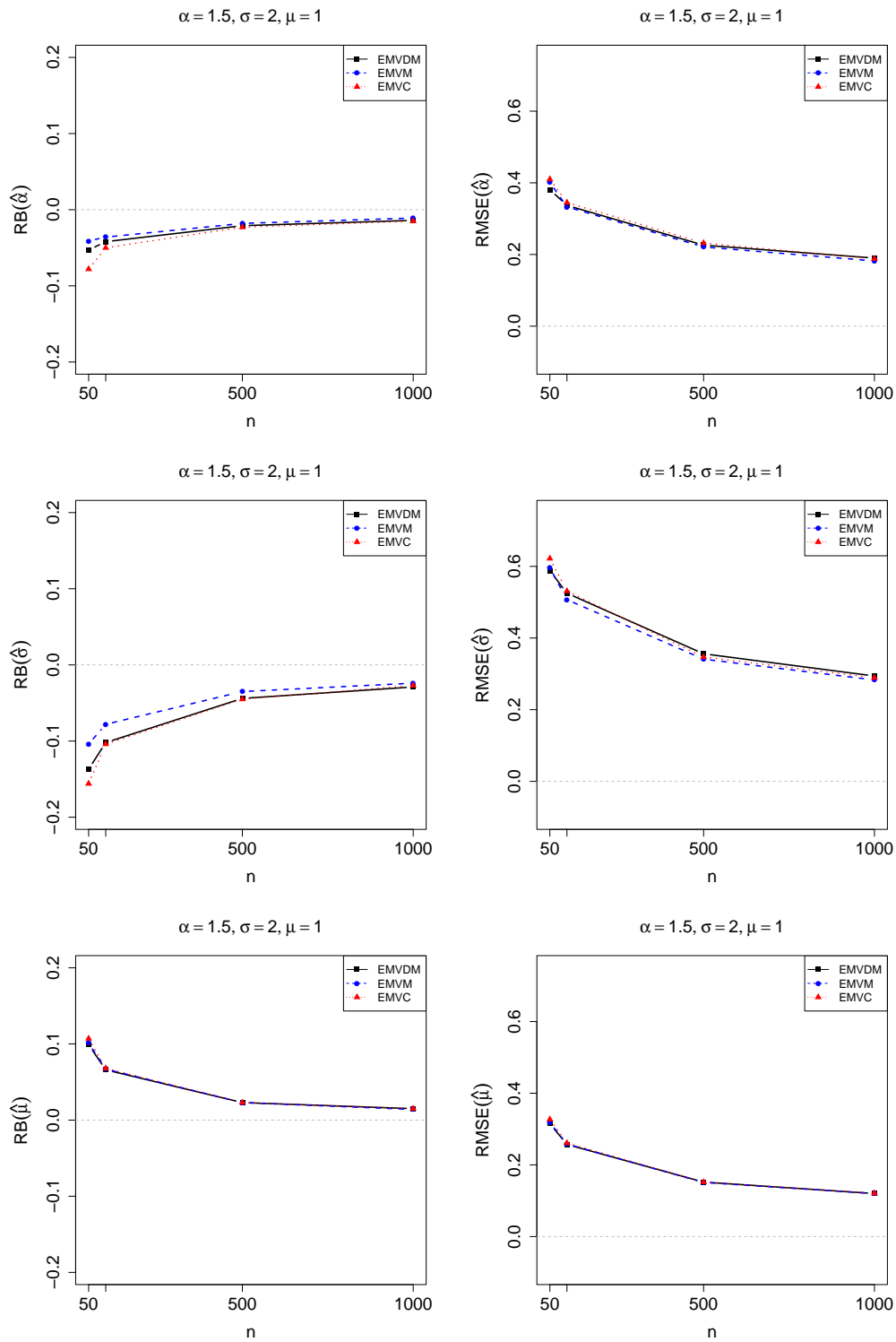


**Figura 4.1:** Viés (esquerda) e RMSE (direita) para as estimativas de  $\alpha = 1.0$  (acima),  $\sigma = 1.0$  (centro), and  $\mu = 1.0$  (abaixo).

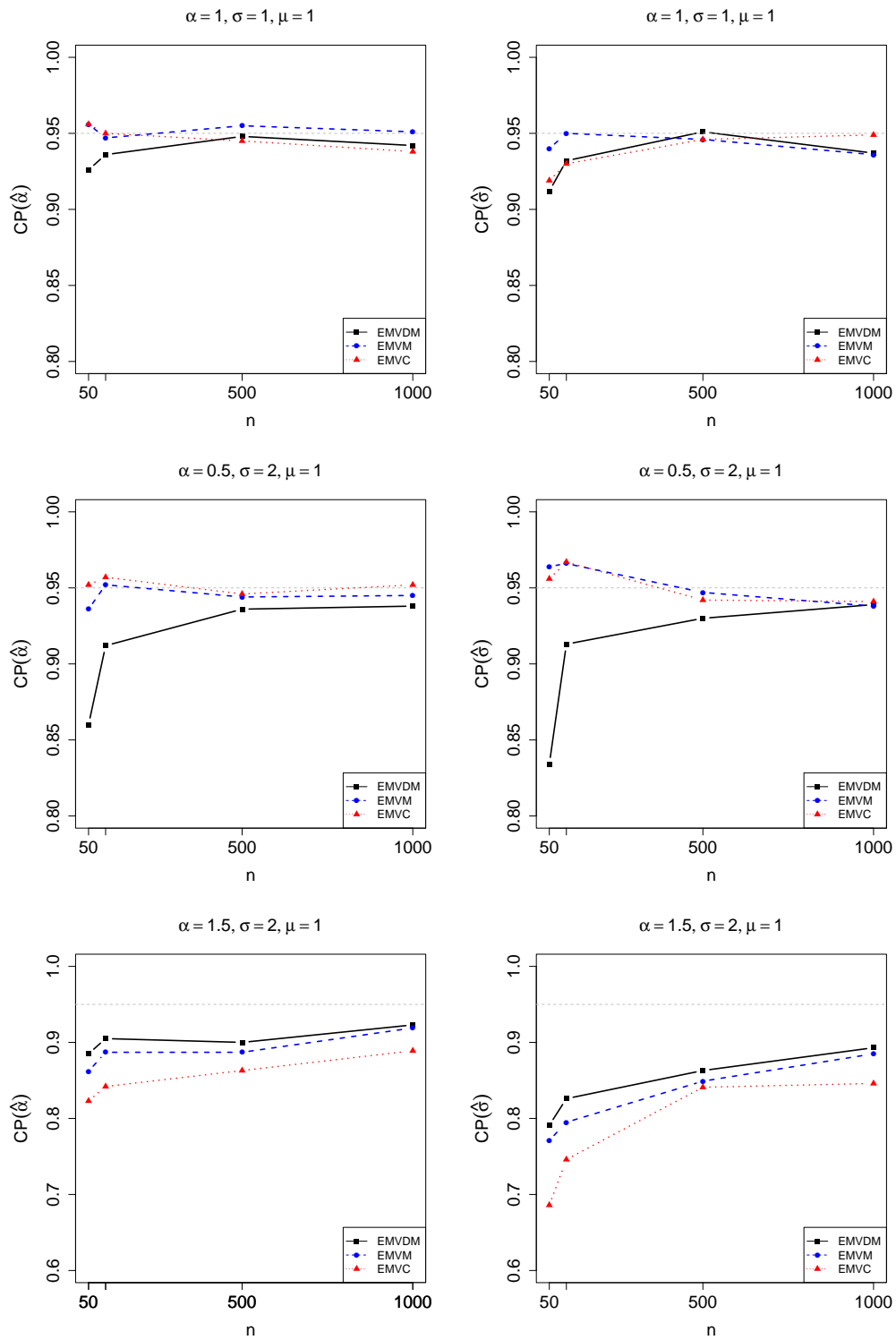




**Figura 4.2:** Viés (esquerda) e RMSE (direita) para as estimativas de  $\alpha = 0.5$  (acima),  $\sigma = 2.0$  (centro), and  $\mu = 1.0$  (abaixo).



**Figura 4.3:** Viés (esquerda) e RMSE (direita) para as estimativas de  $\alpha = 1.5$  (acima),  $\sigma = 2.0$  (centro), and  $\mu = 1.0$  (abaixo).



**Figura 4.4:** Probabilidade de cobertura das estimativas de  $\alpha$  (esquerda) e  $\sigma$  (direita).

# Capítulo 5

## Aplicações a Dados Reais

O presente capítulo foi reservado para apresentar resultados das aplicações em dados reais. A primeira base de dados foi analisada por Kundu e Raqab (2009), dentre outros autores que serão detalhados a seguir. A segunda base de dados refere-se à investimentos estrangeiros que geram emprego ou renda no país para concessão de autorização de residência ao imigrante. Essa é a primeira vez na literatura que tais dados são modelados via distribuições de probabilidade.

### 5.1 Resistência de Fibras de Carbono

A primeira base de dados estudada foi a utilizada por Kundu e Raqab (2009) relativa a resistência de fibras de carbono, que são utilizadas na indústria aeronáutica e náutica por ser um material com boa resistência elétrica e térmica, mais leve e mais durável. Os dados de resistência das fibras de carbono são originais da pesquisa de Bader e Priest (1982).

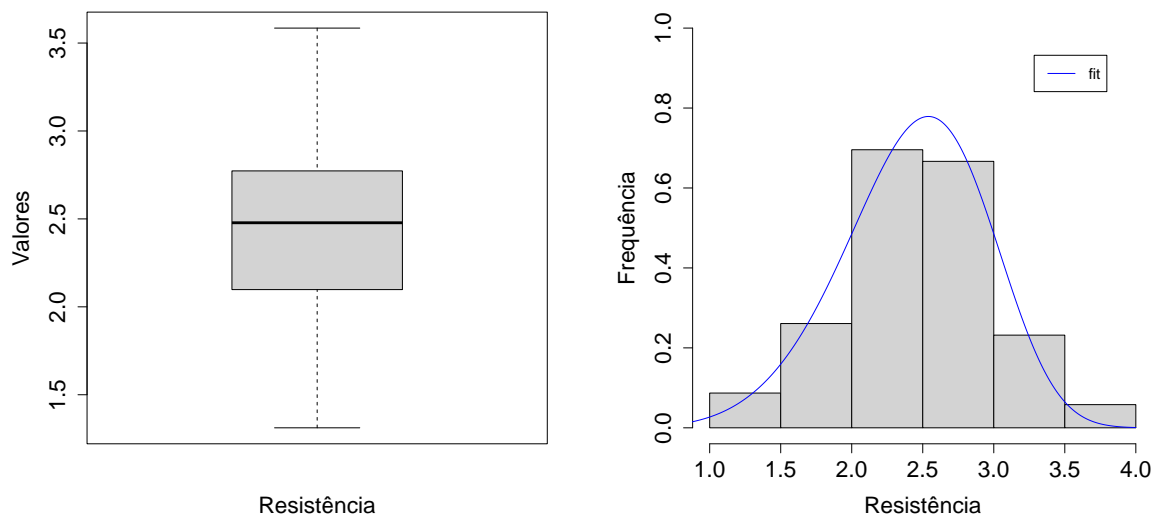
Os dados representam a resistência, medida em GPa (gigapascal), para fibras de carbono simples e fios com 1000 fibras de carbono. As fibras simples foram testadas sob tensão em comprimentos de calibre de 1, 10, 20 e 50 mm. Fios de 1000 fibras foram testados em comprimentos de calibre de 20, 50, 150 e 300 mm. Esses dados foram utilizados nos estudos de Montoya et al. (2019) para estimação dos parâmetros da distribuição Weibull de três

parâmetros pelo método EMVC utilizando várias configurações de parâmetros. Kundu e Raqab (2013) utilizaram esses mesmos dados para estimação dos parâmetros da distribuição Rayleg generalizada de três parâmetros pelo método EMVM, dentre outros.

A análise neste estudo refere-se aos dados de resistência das fibras de carbono simples sob tensão, em comprimento de calibre de 20mm, cuja amostra tem tamanho  $n = 69$ . Para conveniência do leitor, apresentamos os dados no Apêndice.

Os dados contêm 69 observações e variam de 1,312 a 3,585, com valor médio igual a 2,451. As estimativas de confiabilidade dos valores da resistência de carbono com intervalo de confiança de Bootstrap com 95% é (2.330; 2.568), podemos afirmar que em 95% dos casos o valor médio da resistência das fibras de carbono de 20 mm estará nesse intervalo.

A Figura 5.1 apresenta a distribuição dos dados de resistência no Boxplot e no histograma que contém a curva de ajuste para a distribuição teórica Weibull de 2 parâmetros.



**Figura 5.1:** Boxplot e histograma da resistência das fibras de carbono com calibre de 20mm com fit da distribuição Weibull de 2 parâmetros

Pelos resultados apresentados na Tabela 5.1 o logaritmo da função de verossimilhança (llmax) é melhor utilizando os métodos EMVM e EMVDM. Foram obtidas as medidas AIC

(Akaike Information Criterion) e BIC (Bayesian Information Criterion) para comparação dos modelos ajustados.

**Tabela 5.1:** Parâmetros estimados, Máxima verossimilhança, AIC e BIC para os dados de resistência

Método de estimação	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\mu}$	$\ell(\hat{\theta})$	AIC	BIC
EMVM	2.3804	1.8015	1.3120	-51.6745	106.3490	116.0514
EMVC	2.2752	1.7046	1.3120	-55.9871	114.9743	124.6766
EMVDM	2.3423	1.7145	1.3120	-51.7290	106.4580	116.1604
EMV2p	5.2615	2.6909	–	-50.0386	98.0773	91.6091

Os erros padrão (EPs) para os parâmetros de forma  $\alpha$  e escala  $\sigma$  foram calculados com base nas estimativas  $(\hat{\alpha}, \hat{\sigma})$  e na informação de Fisher. A Tabela 5.1 mostra seus valores para cada método de estimação. Com base no EMVDM, os intervalos de confiança de 95% são  $(1,9082; 2,7763)$  e  $(1,2051; 2,2238)$ , respectivamente para  $\alpha$  e  $\sigma$ .

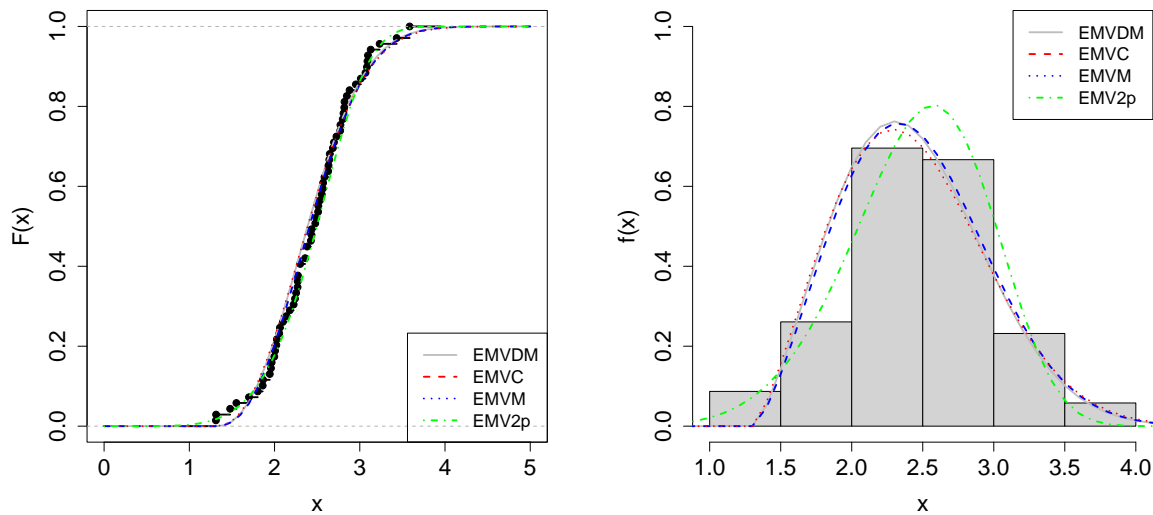
Analisando os dados da Tabela 5.1, podemos verificar que a inclusão do parâmetro de localização não melhora a estimação dos parâmetros de forma e escala, sendo o modelo Weibull de 2 parâmetros o que melhor ajusta os dados. Foram realizados os testes de Cramér-von Mises (CVM) e Kolmogorov-Smirnov (KS) para verificar o ajuste dos dados à distribuição Weibull. O teste de CVM não rejeita a hipótese de que os dados ajustam-se à uma distribuição Weibull, assim como o teste de KS (Tabela 5.2).

**Tabela 5.2:** P-valor dos testes Kolmogorov-Smirnov (KS) e Cramér-von Mises (CVM) para os dados de resistência

Método de estimação	KS	CVM
EMVM	0.8895	0.7829
EMVC	0.6493	0.5009
EMVDM	0.6985	0.5369
EMV2p	0.6436	0.5665

Na Figura 5.2, à esquerda, observa-se as FDAs teóricas ajustam a função de distribuição acumulada empírica (ECDF) dos dados de forma similar. À direita, o ajuste das densidades

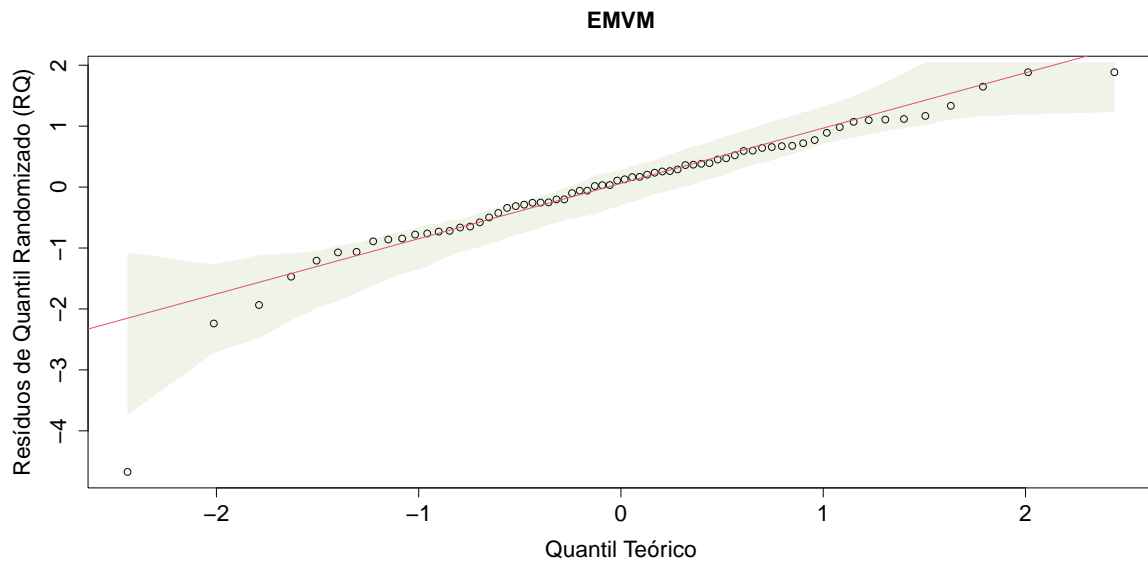
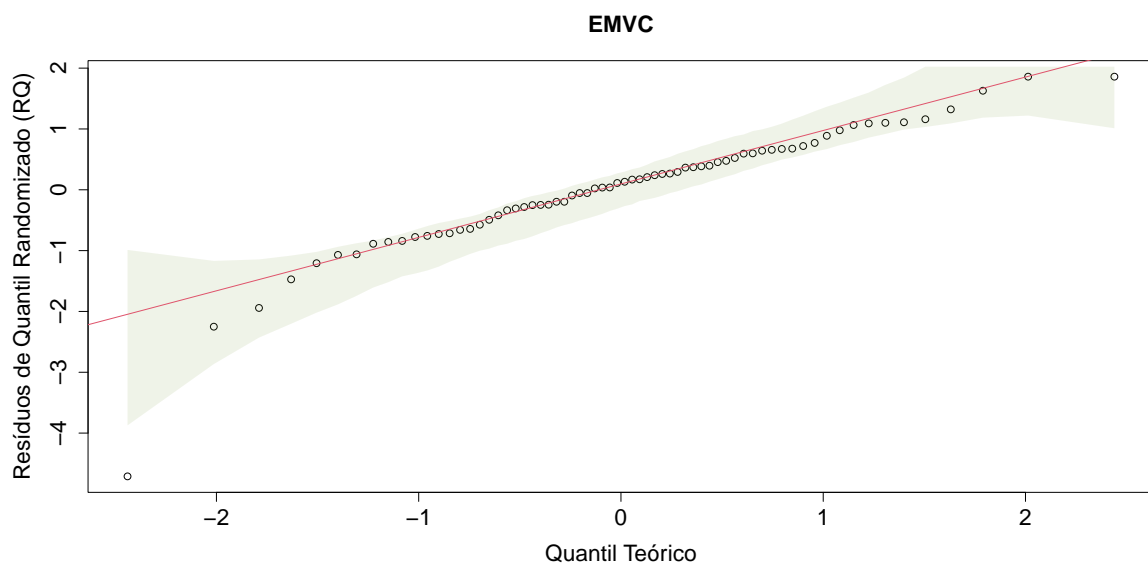
obtidas via estimação dos três métodos é semelhante, porém um pouco diferente da densidade Weibull de dois parâmetros ajustada. Uma análise de resíduos é apresentada a seguir para ajudar na escolha do melhor modelo para os dados.



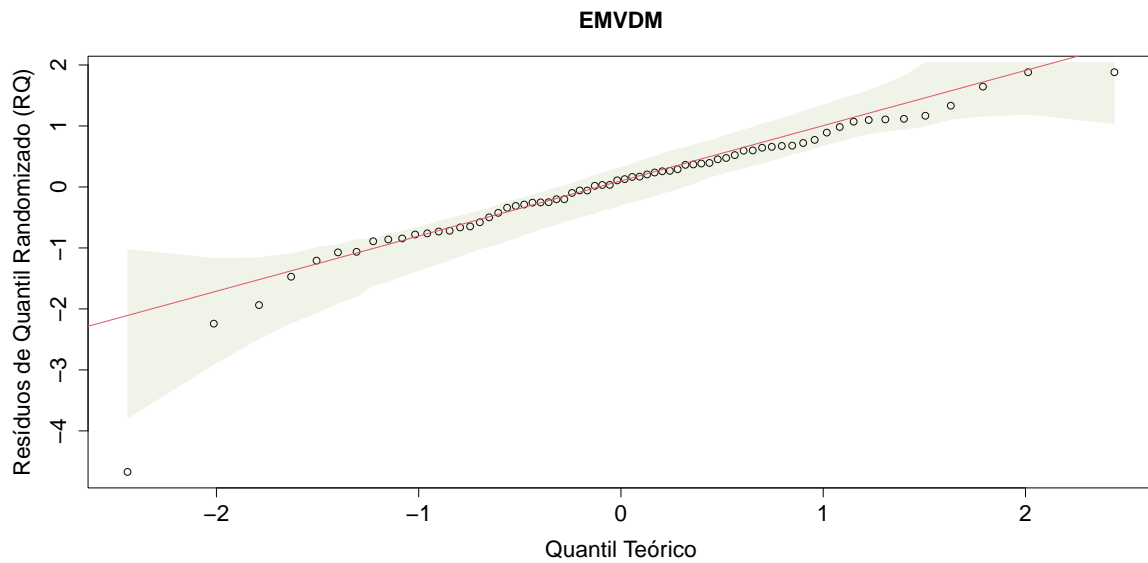
**Figura 5.2:** Ajuste do modelo Weibull pelos métodos de estimação via ECDF e Histograma da resistência das fibras de carbono com calibre de 20mm

Para uma visão mais completa do ajuste da distribuição usamos resíduos de Quantil Randomizado (RQ), conforme definido por (DUNN; SMYTH, 1996). Os RQs são calculados por meio da formula  $R_i = \Phi^{-1}(G(y_i; \hat{\theta}))$ , onde  $\hat{\theta}$  é o vetor dos parâmetros estimados,  $G(y_i; \hat{\theta})$  é a FDA do modelo ajustado para cada observação  $y_i$ , e  $\Phi^{-1}$  representa os quantis da distribuição normal padrão  $N(0, 1)$ . Quando a distribuição  $F$  é contínua, os resíduos RQ tem distribuição normal padrão, excluindo o impacto da variabilidade amostral nos parâmetros estimados.

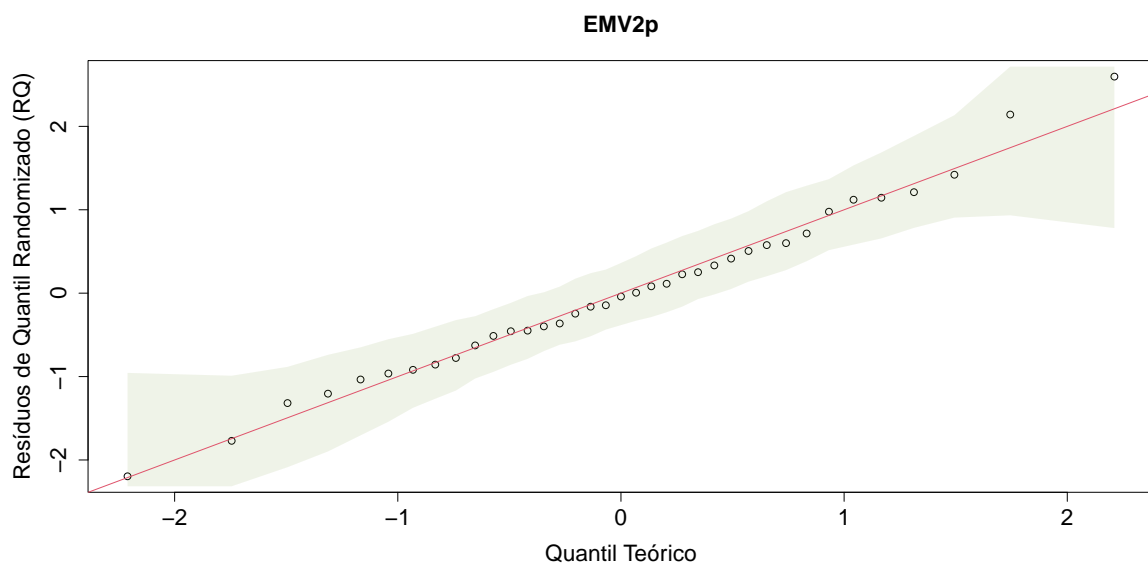
Nas Figuras 5.3-5.6 são apresentados os QQ-plot dos resíduos para os 4 métodos estudados. Esses gráficos ajudam a visualizar se os dados seguem a distribuição desejada ou se duas amostras têm distribuição semelhante. A linha de referência indica a posição ideal onde os pontos estariam seguindo a distribuição especificada. Pela análise gráfica, os dados se ajustam à distribuição Weibull.

**Figura 5.3:** QQ-plot resíduos - EMVM**Figura 5.4:** QQ-plot resíduos - EMVC





**Figura 5.5:** QQ-plot resíduos - EMVDM



**Figura 5.6:** QQ-plot resíduos - EMV2p

## 5.2 Investimentos Estrangeiros

A concessão de autorização de residência para realização de investimento de pessoa física em pessoa jurídica no País é de responsabilidade do Conselho Nacional de Imigração (CNIg),

que é um órgão do Ministério da Justiça e Segurança Pública que tem caráter deliberativo, normativo e consultivo. Este tema foi disciplinado na Resolução Normativa nº 13 (RN 13), de 12 de dezembro de 2017. O valor estabelecido na RN 13, deve ser igual ou superior a R\$ 500.000,00 (quinhentos mil reais) para concessão de vistos permanentes e entre R\$ 150.000,00 e R\$ 500.000,00 para concessão de vistos temporários.

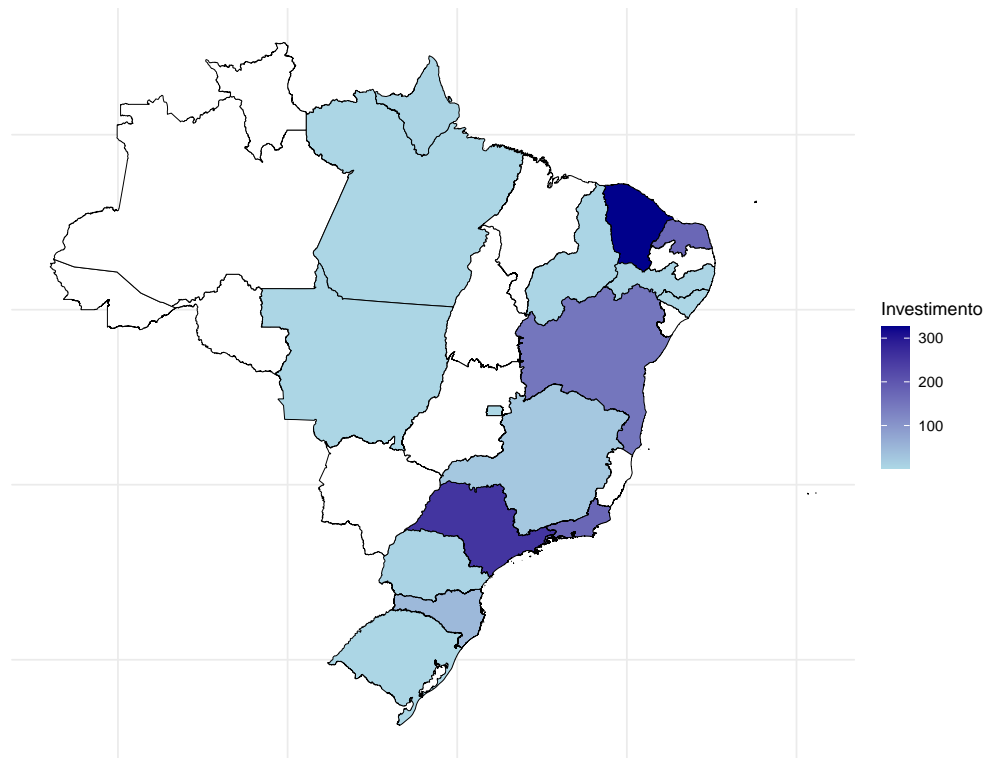
Os dados analisados neste trabalho, correspondem ao valor autorizado para investimento pelo imigrante no Brasil<sup>1</sup>, para investimentos acima de R\$ 500.000,00 e compreendem o período de janeiro a setembro de 2024. Foram desconsiderados da análise os valores de investimento iguais a zero ou não informados. Os dados contêm 132 observações e 4 variáveis, sejam elas, o mês de concessão da autorização, a unidade da federação de residência do estrangeiro, o país de origem e o valor do investimento em Reais (R\$), que por simplicidade da modelagem, foram divididos por 10<sup>5</sup>. Apresentamos os dados do valor do investimento no Apêndice.

De janeiro a setembro de 2024 os imigrantes interessados na autorização de residência, investiram um total de R\$ 114,8 milhões de reais no Brasil. O valor mínimo observado foi de R\$ 501.200,00, enquanto o máximo corresponde a R\$ 6,8 milhões. O valor médio foi de R\$ 869.473,00, pouco mais que o valor do 3º quartil, que foi de R\$ 791.347,00, com desvio padrão de R\$ 801.829,50s. Considerando que a média do conjunto de dados foi de R\$ 869.473,00, as estimativas de confiabilidade dos valores de investimento com intervalo de confiança de Bootstrap com 95% é (R\$ 746.200,00;R\$ 1.020.200,00), podemos afirmar que em 95% dos casos o valor médio de investimentos estará nesse intervalo.

O Ceará foi o estado que recebeu maior investimento (R\$ 31,5 milhões), seguido por São Paulo (R\$ 22,4 milhões) e Rio Grande do Norte (R\$ 17,4 milhões) como observado na Figura 5.7:

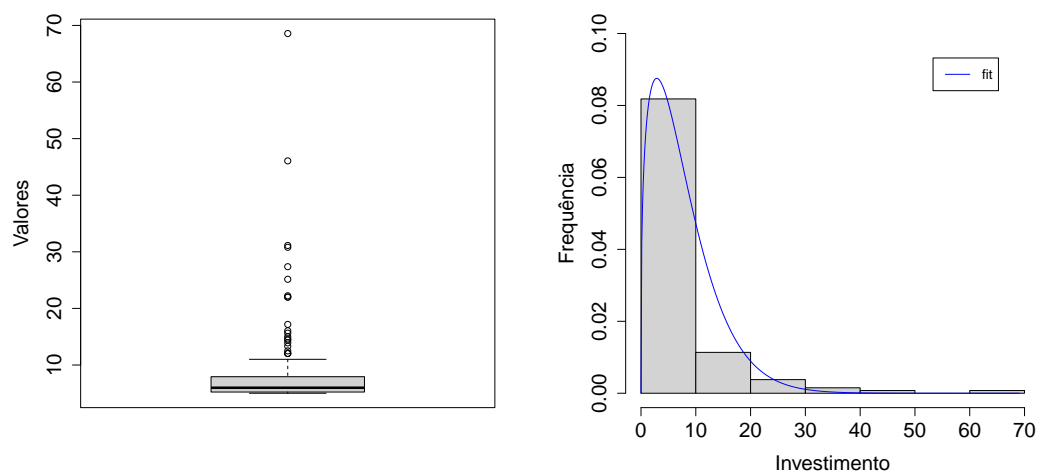
---

<sup>1</sup>Obtidos no site: <https://portaldeimigracao.mj.gov.br/pt/base-de-dados/datamigra?id=401202:cgil-cnig&catid=1733:microdados> (acesso em 07/10/2024)



**Figura 5.7:** Valor do Investimento (x R\$ 100.000,00) por UF

A Figura 5.8 apresenta a distribuição dos dados de resistência no Boxplot e no histograma que contém a curva de ajuste para a distribuição teórica Weibull de 2 parâmetros.



**Figura 5.8:** Boxplot e histograma dos valores do investimento

A Tabela 5.3 apresenta os 10 maiores valores investidos por país de origem, cujo montante equivale a 80,5% do valor total de investimentos. O coeficiente de assimetria é 4,5, indicando assimetria positiva para a distribuição dos dados, que tem uma cauda mais longa à direita (para valores maiores). Esse fato é confirmado pelo valor da curtose, 25,6, indicando a presença de valores extremos.

**Tabela 5.3:** 10 maiores valores de investimentos por País de origem

País de origem	Valor (R\$)
França	209,65
Itália	174,58
China	169,68
Alemanha	133,79
Holanda	73,64
Reino Unido	48,64
Portugal	47,05
Bélgica	42,19
Espanha	38,09
Romênia	33,45

O valor estimado para os parâmetros da Weibull pelos métodos EMVM, EMVC e EMVDM são muito próximos. O ajuste pelos métodos EMVM e EMVDM têm log-verossimilhanças mais próximas entre si e maiores que o valor de EMVC, sugerindo que esses métodos podem ajustar melhor a distribuição aos dados.

Analisando os dados da Tabela 5.4, observamos que a inclusão do parâmetro de localização melhora o ajuste, sendo o modelo Weibull com dois parâmetros o que apresenta o pior ajuste aos dados, conforme indicado pelos critérios AIC e BIC. De acordo com o método EMVDM, os intervalos de confiança de 95% são (0,4606; 0,6702) para  $\alpha$  e (1,0747; 1,9223) para  $\sigma$ , respectivamente.

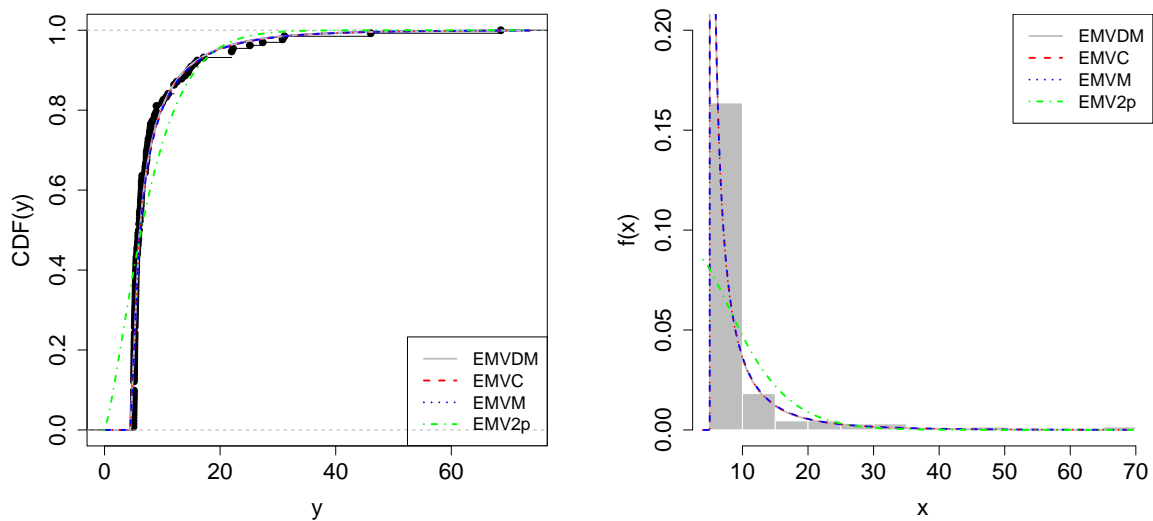
Por fim, apresentamos o histograma dos dados e as curvas de ajuste (ECDF) dos três métodos estudados. Como já evidenciado pelas medidas descritivas, 75% dos dados são de até R\$ 791.347,00, onde se observa grande concentração de dados. Com relação aos ajustes, o método

**Tabela 5.4:** Parâmetros estimados, Máxima verossimilhança, AIC e BIC para os dados de investimento

Método de estimação	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\mu}$	$\ell(\hat{\theta})$	AIC	BIC
EMVM	0.5691	1.5350	5.012	-250.8152	504.6303	514.3327
EMVC	0.5653	1.5158	5.012	-252.2789	507.5579	517.2602
EMVDM	0.5654	1.4985	5.012	-250.8422	504.6844	514.3867
EMV2p	1.4847	8.0858	–	-409.8814	803.2061	796.7379

EMVDM trouxe bom ajuste dos dados à distribuição Weibull.

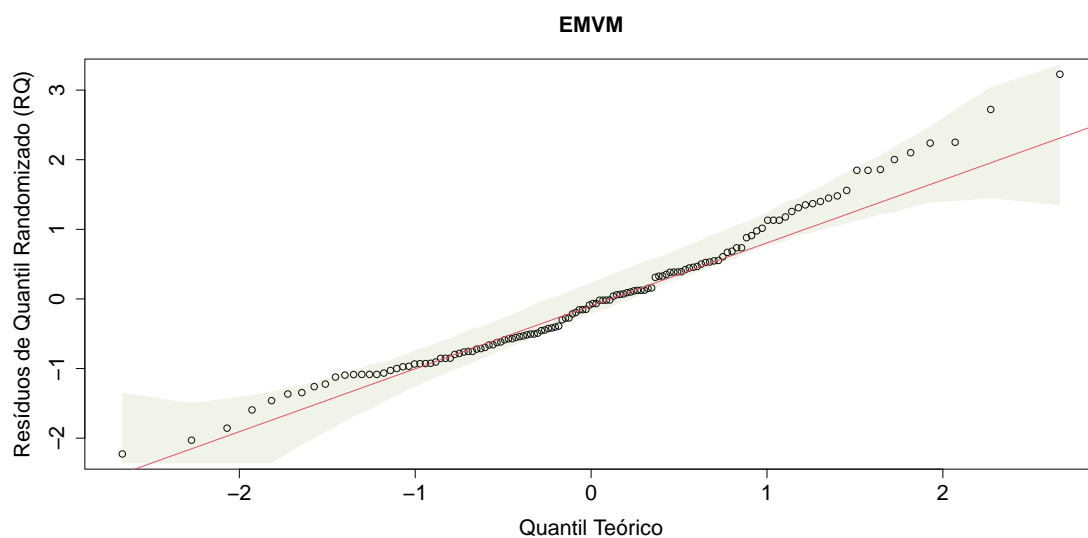
Na Figura 5.9, à esquerda, observa-se a ECDF, que os três métodos, ajustam os dados de investimento à Weibull de forma similar. À direita, a densidade de probabilidade dos dados também é similar para os três métodos, a maioria dos dados se acumulam entre o intervalo 0 e 10, já que o valor mínimo para investimento é de R\$ 500 mil reais, e 75% dos dados tem valor até R\$ 800 mil, verifica-se também a presença de valores extremos, que podem ser analisados de forma mais criteriosa em outro momento.

**Figura 5.9:** Ajuste do modelo Weibull pelos métodos de estimação via ECDF e Histograma dos dados de investimento

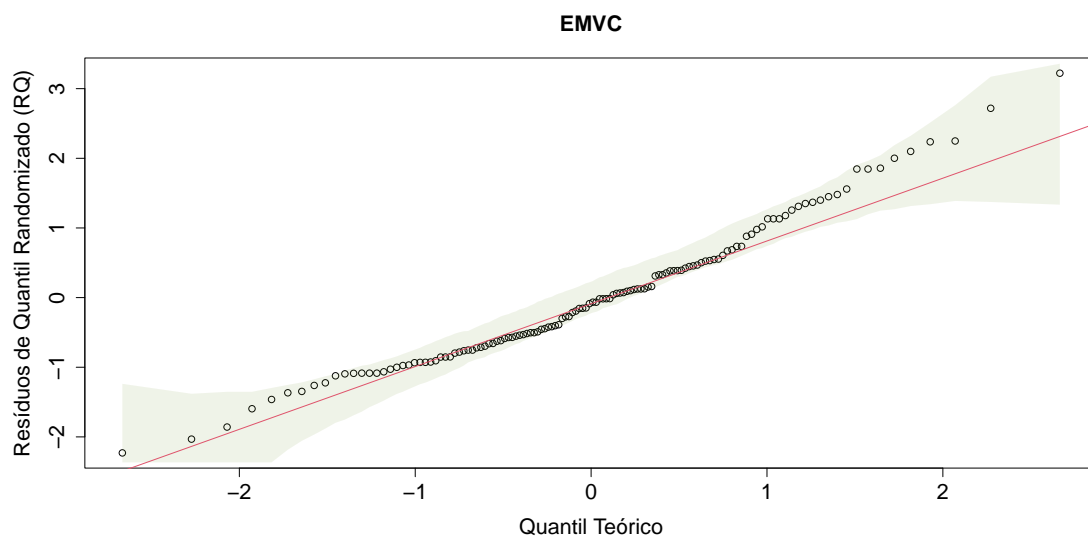
Já pelo gráfico QQ-plot dos resíduos comparamos os quantis dos resíduos com os quantis

teóricos da Weibull estimados que parecem se ajustar bem aos dados de investimento, dessa forma, podemos afirmar que os dados de investimento podem se ajustar a uma distribuição Weibull.

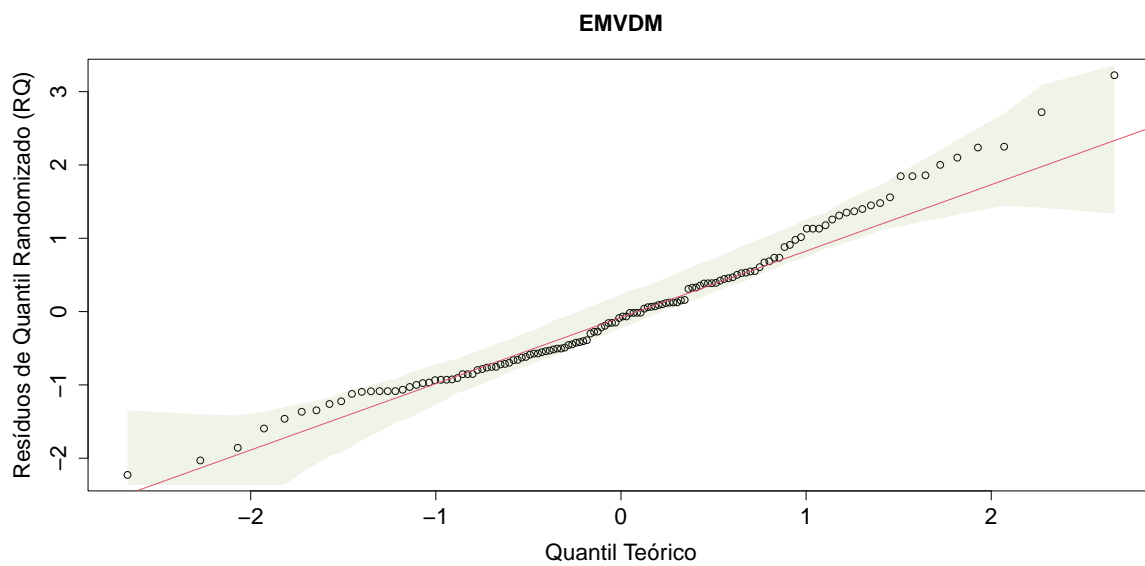
Pela análise das Figuras 5.10-5.13 verificamos melhor qualidade do ajuste dos dados à distribuição teórica nos métodos EMVM, EMVC e EMVDM.



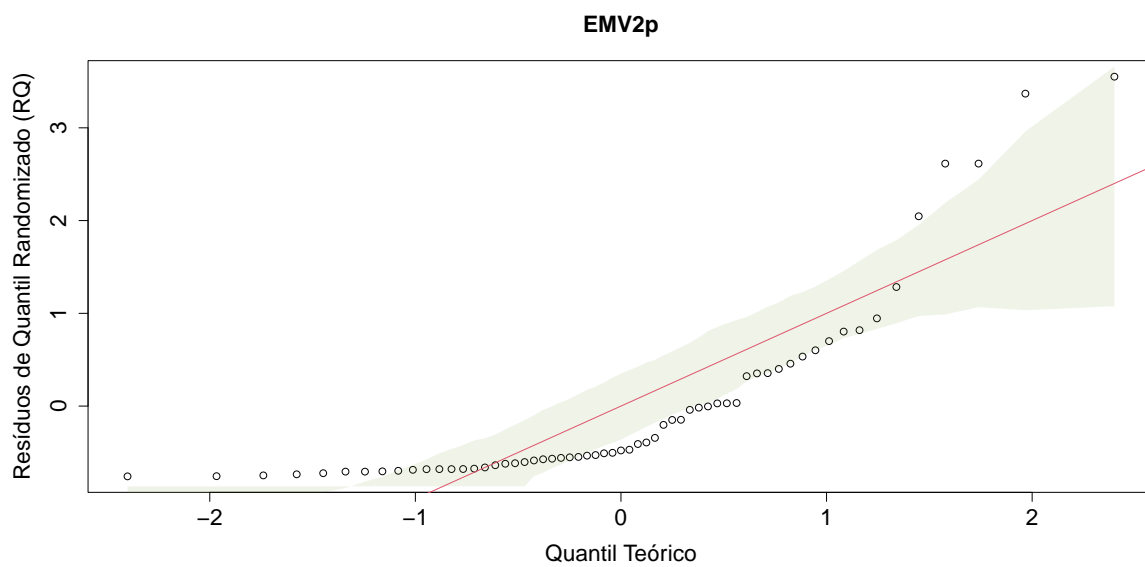
**Figura 5.10:** QQ-plot resíduos - EMVM



**Figura 5.11:** QQ-plot resíduos - EMVC



**Figura 5.12:** QQ-plot resíduos - EMVDM



**Figura 5.13:** QQ-plot resíduos - EMV2p

# Capítulo 6

## Considerações Finais

Nos casos em que o suporte da função densidade de probabilidade depende de parâmetros desconhecidos, os métodos convencionais de obtenção de estimativas dos parâmetros se tornam inviáveis, fazendo com que o estimador de máxima verossimilhança possa não assumir valores finitos. Para situações como esta, em que a estimativa de máxima verossimilhança não existe, estudamos, nesta dissertação, três métodos de estimação para a distribuição Weibull de três parâmetros.

O primeiro modelo, EMVM, é o proposto por Kundu e Raqab (2009), onde o logaritmo da função de verossimilhança modificada é baseado em  $(n - 1)$  observações após ignorar a menor observação e substituir  $\mu$  pelo seu estimador natural  $\hat{\mu} = X_{(1)}$ , modificando o vetor escore. O segundo método, proposto por Cheng e Iles (1987), é o EMVC, que ao integrar a densidade de probabilidade sobre um pequeno intervalo  $[X_{(1)}, X_{(1)} + h]$ , onde  $h$  é um parâmetro de correção, há redução no impacto de qualquer singularidade ou anomalia associada à menor observação. Por fim, inspirados nas ideias de Firth (1993), introduzimos um terceiro método, EMVDM, baseado na penalização do vetor escore modificado calculado no primeiro método.

Verificamos o desempenho dos três métodos na estimação dos parâmetros da distribuição Weibull através de um estudo de simulações de Monte Carlo. Em determinados cenários, o EMVDM teve melhor desempenho do que os outros estimadores. Os estimadores estudados



também foram aplicados em dados reais de resistência de fibras de carbono e de investimento estrangeiro no Brasil.

Para trabalhos futuros, sugerimos a utilização das ideias de Firth (1993) em problemas de estimação da distribuição Weibull fazendo a penalização da verossimilhança corrigida de Cheng e Iles (1987). Outra possibilidade, é propor uma nova distribuição para resolver o problema da exclusão da menor observação e ainda garantir uma melhor estimação dos parâmetros. Por fim, essa metodologia pode ser utilizada em outras áreas da Teoria de Probabilidade, como em Processos Estocásticos, por exemplo.

# Apêndice

## Dados 1 - Resistência de fibras de carbono 20mm <sup>1</sup>

X= (1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.140, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.570, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.880, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.585).

## Dados 2 - Investimentos Estrangeiros<sup>2</sup>

X=(8.958070, 25.150000, 5.015640, 11.006760, 5.279225, 5.013624, 6.435200, 5.835900, 7.000000, 30.801862, 7.266010, 5.019073, 8.000000, 22.000000, 5.514300, 5.302930, 7.968871, 6.000000, 7.058850, 15.000000, 5.155455, 5.243266, 5.219835, 9.950000, 12.041000, 7.300000, 5.743445, 6.272980, 5.070780, 6.529300, 6.000000, 8.697271, 46.067600, 17.171580, 6.098290, 10.166405, 5.445229, 7.281530, 7.757070, 6.413050, 5.331970, 6.225743, 14.399190, 5.850000, 5.653820, 5.484998, 5.257000, 5.841250, 8.265000, 27.373000, 5.412000, 5.099400, 5.145000, 5.344151, 7.064000, 7.608040, 6.346600, 5.052410, 5.623980, 14.615000, 5.358124, 5.210000, 68.570220, 5.097540, 5.400696, 5.959290, 5.148630, 7.860000, 5.104935, 22.000000, 8.985627,

---

<sup>1</sup>Dados disponíveis, por exemplo, em Kundu e Raqab (2009)

<sup>2</sup>Obtidos no site: <https://portaldeimigracao.mj.gov.br/pt/base-de-dados/datamigra?id=401202:cgil-cnig&catid=1733:microdados> (acesso em 07/10/2024)

5.176840, 5.177960, 5.040000, 7.419586, 15.634430, 5.146730, 5.131308, 7.153940, 6.105075,  
5.100000, 5.202800, 6.371200, 5.477140, 5.309140, 5.776270, 6.301900, 5.100000, 12.041000,  
5.123362, 5.133710, 5.225000, 5.030000, 6.508635, 5.012000, 5.064560, 5.500000, 12.041000,  
5.115196, 13.318230, 5.147040, 7.895000, 5.050000, 5.400000, 6.441160, 5.090940, 5.369000,  
8.607168, 5.455010, 22.230579, 5.376720, 7.275880, 13.920000, 5.225000, 31.122660, 12.4994,  
5.342750, 5.177051, 7.556150, 5.390042, 6.433160, 7.514053, 10.695323, 8.985627, 5.247600,  
6.287972, 16.052030, 5.659609, 5.100000, 6.098817, 6.101590, 5.281640)

# Referências

- Ali, M Masoom; Pal, Manisha; Woo, Jungsoo. Estimation of  $P(Y < X)$  in a four-parameter generalized gamma distribution. **Austrian Journal of Statistics**, v. 41, n. 3, p. 197–210, 2012.
- Almeida, Frederico M.; Colosimo, Enrico A.; Mayrink, Vinícius D. Prior specifications to handle the monotone likelihood problem in the Cox regression model. **Statistics and Its Interface**, v. 11, n. 4, p. 687–698, 2018.
- \_\_\_\_\_. Firth adjusted score function for monotone likelihood in the mixture cure fraction model. **Lifetime Data Analysis**, v. 27, p. 131–155, 2021.
- \_\_\_\_\_. Modified score function for monotone likelihood in the semiparametric mixture cure model. **Biometrical Journal**, v. 64, n. 3, p. 635–654, 2022.
- Arrué, Jaime; Arellano-Valle, Reinaldo B; Gómez, Héctor W. Bias reduction of maximum likelihood estimates for a modified skew-normal distribution. **Journal of Statistical Computation and Simulation**, v. 86, n. 15, p. 2967–2984, 2016.
- Arrué, Jaime et al. Likelihood Based Inference and Bias Reduction in the Modified Skew-t-Normal Distribution. **Mathematics**, v. 11, n. 15, p. 3287, 2023.
- Bader, M.; Priest, A. Statistical Aspects of Fiber and Bundle Strength in Hybrid Composites. Edição: T. Hayashi, S. Kawata e S. Umekawa. **Progress in Science and Engineering Composites**, ICCM–IV, 1129–1136, 1982.
- Brown, L.; Zhao, Y. Improving Bayesian inference with corrected likelihoods. **Bayesian Analysis**, v. 1, p. 22–35, 1996.
- Bull, Shelley B; Mak, Carmen; Greenwood, Celia MT. A modified score function estimator for multinomial logistic regression in small samples. **Computational Statistics & Data Analysis**, v. 39, n. 1, p. 57–74, 2002.
- Burden, R.L.; Faires, J.D.; Burden, A.M. **Numerical Analysis**. [S.l.]: Cengage Learning, 2015. ISBN 9781305465350. Disponível em: <<https://books.google.com.br/books?id=9DV-BAAAQBAJ>>.
- Casella, G.; Berger, R.L. **Statistical Inference**. [S.l.]: Duxbury Thomson Learning, 2002. (Duxbury advanced series). ISBN 9780495391876. Disponível em: <<https://books.google.com.br/books?id=ZpkPPwAACAAJ>>.
- Cheng, R.C.H.; Iles, T.C. Corrected maximum likelihood in non-regular problems. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 49, n. 1, p. 95–101, 1987.

- Cheng, R.C.H.; Traylor, L. Non-regular maximum likelihood problems. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 57, n. 1, p. 3–24, 1995.
- Clarke, Robin T. Estimating trends in data from the Weibull and a generalized extreme value distribution. **Water Resources Research**, v. 38, n. 6, p. 25–1–25–10, 2002.
- Dunn, P. K.; Smyth, G. K. Randomized quantile residuals. **Journal of Computational and graphical statistics**, 236–244, 1996.
- Firth, David. Bias reduction of maximum likelihood estimates. **Biometrika**, v. 80, n. 1, p. 27–38, 1993.
- Fisher, R. A.; Tippett, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. **Mathematical Proceedings of the Cambridge Philosophical Society**, v. 24, n. 2, 180–190, 1928.
- Galambos, J. **The Asymptotic Theory of Extreme Order Statistics**. [S.l.]: R.E. Krieger Publishing Company, 1978. (Wiley series in probability and mathematical statistics).
- Gebizlioglu, Omer; Şenoğlu, Birdal; Kantar, Yeliz Mert. Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution. **Journal of Computational and Applied Mathematics**, v. 235, n. 11, p. 3304–3314, 2011.
- Heinze, Georg; Ploner, Meinhard. Fixing the nonconvergence bug in logistic regression with SPLUS and SAS. **Computer methods and programs in biomedicine**, v. 71, n. 2, p. 181–187, 2003.
- Heinze, Georg; Schemper, Michael. A solution to the problem of monotone likelihood in Cox regression. **Biometrics**, v. 57, n. 1, p. 114–119, 2001.
- Johnson, N.L.; Kotz, S. **Continuous Univariate Distributions**. [S.l.]: J. Wiley, 1970.
- Jones, B.; Smith, C. Estimation in interval-censored models using corrected likelihood. **Statistics in Medicine**, v. 10, p. 1089–1100, 1991.
- Kohansal, Akram; Rezakhah, Saeid. Inference of  $R = P(Y < X)$  for two-parameter Rayleigh distribution based on progressively censored samples. **Statistics**, v. 53, n. 1, p. 81–100, 2019.
- Kundu, Debasis; Raqab, Mohammad Z. Estimation of  $R = P(Y < X)$  for three-parameter Weibull distribution. **Statistics & Probability Letters**, v. 79, n. 17, p. 1839–1846, 2009.
- \_\_\_\_\_. Estimation of  $R = P[Y < X]$  for three-parameter generalized Rayleigh distribution. **Journal of Statistical Computation and Simulation**, v. 85, n. 4, p. 725–739, 2013.
- Lima, Verônica MC; Cribari-Neto, Francisco. Penalized maximum likelihood estimation in the modified extended Weibull distribution. **Communications in Statistics-Simulation and Computation**, v. 48, n. 2, p. 334–349, 2019.
- Matsushita, Satoru et al. Lifetime data analysis of disease and aging by the Weibull probability distribution. **Journal of clinical epidemiology**, v. 45, n. 10, p. 1165–1175, 1992.
- Mittnik, Stefan; Rachev, Svetlozar T. Modeling asset returns with alternative stable distributions. **Econometric reviews**, v. 12, n. 3, p. 261–330, 1993.

- Montoya, José A; Díaz-Francés, Eloísa; Figueroa, Gudelia. Estimation of the reliability parameter for three-parameter Weibull models. **Applied Mathematical Modelling**, v. 67, p. 621–633, 2019.
- Nagatsuka, Hideki; Kamakura, Toshinari; Balakrishnan, N. A consistent method of estimation for the three-parameter Weibull distribution. **Computational Statistics & Data Analysis**, v. 58, p. 210–226, 2013.
- Ng, H.K.T et al. Parameter estimation of three-parameter Weibull distribution based on progressively Type-II censored samples. **Journal of Statistical Computation and Simulation**, v. 82, n. 11, p. 1661–1678, 2012.
- Papalexiou, Simon Michael; AghaKouchak, Amir; Foufoula-Georgiou, Efi. A Diagnostic Framework for Understanding Climatology of Tails of Hourly Precipitation Extremes in the United States. **Water Resources Research**, v. 54, n. 9, p. 6725–6738, 2018.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.
- Resnick, S. **Extreme values, regular variation, and point processes**. [S.l.]: Springer, 2008. v. 4.
- Smith, A. Adjusted maximum likelihood estimation for heavy-tailed distributions. **Journal of Statistical Planning and Inference**, v. 25, p. 45–56, 1990.
- Smith, Richard L. Maximum likelihood estimation in a class of nonregular cases. **Biometrika**, v. 72, n. 1, p. 67–90, 1985.