



GUSTAVO HENRIQUE TASCA

INFERÊNCIA BAYESIANA PARA DISTRIBUIÇÕES DE CAUDA  
LONGA

CAMPINAS  
2015





UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística  
e Computação Científica

GUSTAVO HENRIQUE TASCA

INFERÊNCIA BAYESIANA PARA DISTRIBUIÇÕES DE CAUDA  
LONGA

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em estatística.

**Orientadora: Laura Letícia Ramos Rifo**

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO GUSTAVO HENRIQUE TASCA, E ORIENTADA PELA PROFA. DRA. LAURA LETÍCIA RAMOS RIFO.

Assinatura da Orientadora

A handwritten signature in black ink is written over a horizontal line. The signature is cursive and appears to read "Laura Letícia Ramos Rifo".

CAMPINAS

2015

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Ana Regina Machado - CRB 8/5467

T181i Tasca, Gustavo Henrique, 1990-  
Inferência bayesiana para distribuições de cauda longa / Gustavo Henrique  
Tasca. – Campinas, SP : [s.n.], 2015.

Orientador: Laura Leticia Ramos Rifo.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de  
Matemática, Estatística e Computação Científica.

1. Inferência bayesiana. 2. Distribuição (Probabilidades). I. Rifo, Laura Leticia  
Ramos, 1970-. II. Universidade Estadual de Campinas. Instituto de Matemática,  
Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Bayesian inference for long-tailed distributions

**Palavras-chave em inglês:**

Bayesian inference

Distribution (Probability theory)

**Área de concentração:** Estatística

**Titulação:** Mestre em Estatística

**Banca examinadora:**

Laura Leticia Ramos Rifo [Orientador]

Victor Fossaluzza

Verónica Andrea González-López

**Data de defesa:** 23-02-2015

**Programa de Pós-Graduação:** Estatística

**Dissertação de Mestrado defendida em 23 de fevereiro de 2015 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**



---

**Prof(a). Dr(a). LAURA LETICIA RAMOS RIFO**



---

**Prof(a). Dr(a). VICTOR FOSSALUZA**



---

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



## Abstract

In this work, we study Bayesian inference methods for long-tailed distributions that don't involve the evaluation of the likelihood function. Initially, we present an analysis of the properties of heavy-tailed distributions and particular cases, as long-tailed, subexponential and regular variation families. Some statistics are presented and their sampling behavior studied, in order to develop diagnostic measures. For obtaining posterior inferences, we discuss the minimum entropy ABC and others likelihood-free algorithms, aiming model checking and model selection. We introduce a new model selection algorithm based on the posterior predictive distribution, the results of which are validated through simulations and real data related to river flow.

**Keywords:** bayesian inference, heavy-tailed distributions, approximate bayesian computation.

## Resumo

Neste trabalho, estudamos métodos de inferência bayesiana para distribuições de cauda longa, que não envolvam o cálculo da função de verossimilhança. Inicialmente, apresentamos uma análise das propriedades de distribuições de cauda pesada e seus casos particulares, como as famílias de distribuições de cauda longa, subexponenciais e de variação regular. Apresentamos algumas estatísticas e seus comportamentos amostrais, a fim de desenvolvermos medidas de diagnóstico. Para obtenção de inferências a posteriori, discutimos o método ABC de mínima entropia e outros algoritmos para verificação e seleção de modelos, que não utilizam o cálculo da função de verossimilhança. Introduzimos um novo algoritmo para seleção de modelos baseado na distribuição preditiva a posteriori, cujos resultados são validados através de simulações e análises de dados reais relacionados à hidrologia.

**Palavras-chave:** inferência bayesiana, distribuições de cauda pesada, aproximação computacional bayesiana.



# Sumário

Agradecimentos	xi
Lista de Ilustrações	xv
Lista de Tabelas	xvii
<b>1 Distribuições de cauda pesada</b>	<b>1</b>
1.1 Introdução . . . . .	1
1.1.1 Preliminares . . . . .	1
1.2 Classes de cauda pesada . . . . .	4
1.2.1 Classe de cauda longa . . . . .	4
1.2.2 Classe subexponencial . . . . .	5
1.2.3 Classe de variação regular . . . . .	6
1.3 Medidas de evidência de cauda pesada . . . . .	10
1.3.1 Razão entre estatísticas de ordem . . . . .	11
1.3.2 Índice de Obesidade . . . . .	14
1.3.3 Função de excesso médio . . . . .	18
<b>2 Inferência bayesiana</b>	<b>23</b>
2.1 Terminologias e preliminares . . . . .	23
2.1.1 Distribuição preditiva a posteriori . . . . .	25
2.2 Métodos MCMC . . . . .	26
2.3 Aproximação bayesiana computacional . . . . .	28
2.4 Verificação do modelo . . . . .	31
2.5 Seleção de modelos . . . . .	33
<b>3 Aplicações</b>	<b>36</b>
3.1 Estimação . . . . .	36

3.2	Exemplos simulados . . . . .	38
3.3	Aplicação em dados reais . . . . .	47
<b>4</b>	<b>Considerações finais</b>	<b>54</b>
	<b>Referências</b>	<b>56</b>
<b>A</b>	<b>Figuras</b>	<b>59</b>
<b>B</b>	<b>Tabelas</b>	<b>72</b>

# Agradecimentos

Agradeço a Deus, por todas as oportunidades com as quais tem me presenteado.

Aos meus pais, Natale e Cida, e minhas irmãs, Michelle e Vanessa, por todo apoio oferecido.

À Thaíse Mendes, por ser uma namorada companheira e amiga nestes difíceis anos.

À professora Laura Rifo, por toda a atenção, compreensão e auxílio, dentro e fora do programa de pós-graduação.

Aos professores Ademir Petenate, Jorge Mujica e Vera Carvalho, por serem excelentes modelos de educadores.

Aos professores Verónica González-López e Victor Fossaluzza, por terem concordado em fazer parte da banca examinadora, e pelas valiosas sugestões para o melhoramento deste trabalho.

Aos amigos Fernando El Kadri, Gabriel Franco, Claudia Koda e Lígia Silveira, que, de uma forma ou de outra, me apoiaram e contribuíram para a conclusão do curso de mestrado.

À CAPES, sem a qual eu não teria obtido este título.



“Live to win...” - Paul Stanley



# Lista de Ilustrações

1	Gráficos de índices de obesidade generalizados das distribuições $Weibull(1, k)$ e $Burr(2, \alpha)$ . . . . .	18
2	Histogramas de amostras de tamanho 100 de uma distribuição $Gama(2, 3)$ . . . . .	32
3	Histogramas de amostras de tamanho 100 de uma distribuição $Log - Gama(0.5, 5)$ . . . . .	33
4	Gráfico da densidade a posteriori exata e suas aproximações para o Exemplo 36. . . . .	40
5	Gráficos das densidades a posteriori estimadas para o Exemplo 38. . . . .	44
6	Gráfico da distribuição a priori para os modelos apresentados na Tabela B.9. . . . .	46
7	Gráfico da distribuição a posteriori pela abordagem ABC-MC para os modelos apresentados na Tabela B.9. . . . .	47
8	Gráfico da distribuição a posteriori pela abordagem ABC-PPMC para os modelos apresentados na Tabela B.9. . . . .	48
9	Gráfico de traços para a amostra do exemplo de dados reais. . . . .	49
10	Gráfico de autocorrelações para a amostra do exemplo de dados reais. . . . .	50
11	Gráfico da função de excesso médio estimada para a amostra do exemplo de dados reais. . . . .	52
12	Gráfico da distribuição a posteriori para os modelos apresentados na Tabela B.10. . . . .	52
13	Histogramas de 10000 simulações das estatísticas de interesse via distribuição preditiva a posteriori sob o modelo M21. . . . .	53
A.1	Histogramas de 10000 simulações de $\hat{R}a_{2,2}(X)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	59
A.2	Histogramas de 10000 simulações de $\hat{R}a_{15,2}(X)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	60

A.3	Histogramas de 10000 simulações de $\hat{Ra}_{30,2}(X)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	61
A.4	Histogramas de 10000 simulações de $\hat{O}b(X, 4)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	62
A.5	Histogramas de 10000 simulações de $\hat{O}b(X, 15)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	63
A.6	Histogramas de 10000 simulações de $\hat{O}b(X, 30)$ para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	64
A.7	Gráficos de $\hat{e}_X(x)$ para diversas distribuições considerando tamanho de amostra 1000.	65
A.8	Histogramas de 10000 simulações da estimativa do índice de variação regular, $\alpha$ , via FEM, $\hat{\alpha}_{FEM}$ , considerando tamanho de amostra 1000. . . . .	66
A.9	Gráficos de autocorrelação e de traços para estimação MCMC para o Exemplo 36. . . . .	67
A.10	Histogramas das amostras MCMC da distribuição marginal a posteriori dos índice de variação regular $\alpha_1$ e $\alpha_2$ que compõe a distribuição dos dados do Exemplo 38. . . . .	68
A.11	Curvas de nível da densidade conjunta da amostra MCMC da distribuição a posteriori dos índice de variação regular $\alpha_1$ e $\alpha_2$ , que compõe a distribuição dos dados do Exemplo 38. . . . .	69
A.12	Histograma da amostra da distribuição a posteriori do índice de variação regular, $\alpha$ , via método MCMC para o Exemplo 38. . . . .	70
A.13	Gráfico da função de excesso médio estimada para o Exemplo 39. . . . .	71

# Lista de Tabelas

1	Exemplos de distribuições subexponenciais. . . . .	6
2	Exemplos de distribuições de variação regular. . . . .	10
3	Valores numéricos de $Ra_{n,2}(X)$ calculados para algumas distribuições. . . . .	15
4	Funções de perda e seus respectivos estimadores de Bayes. . . . .	25
5	Valores das estimativas observadas para o índice de variação, $\alpha$ , para cem repetições do Exemplo 36. . . . .	41
6	Medidas de evidência de cauda pesada para o exemplo de dados reais. . . . .	51
B.1	Viés médios observados em 10000 simulações de $\hat{R}a_{2,2}(X)$ , $\hat{R}a_{15,2}(X)$ e $\hat{R}a_{30,2}(X)$ para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	72
B.2	EQM observados em 10000 simulações de $\hat{R}a_{2,2}(X)$ , $\hat{R}a_{15,2}(X)$ e $\hat{R}a_{30,2}(X)$ para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	73
B.3	Viés médios observados em 10000 simulações de $\hat{O}b(X, 4)$ , $\hat{O}b(X, 15)$ e $\hat{O}b(X, 30)$ para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	73
B.4	EQM observados em 10000 simulações de $\hat{O}b(X, 4)$ , $\hat{O}b(X, 15)$ e $\hat{O}b(X, 30)$ para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens. . . . .	73
B.5	Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 36. . . . .	74
B.6	Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 37. . . . .	75

B.7	Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 38. . . . .	76
B.8	Medidas de evidência de cauda pesada para o Exemplo 39. . . . .	77
B.9	Conjunto de modelos candidatos para o Exemplo 39. . . . .	78
B.10	Conjunto de modelos candidatos para o exemplo de dados reais. . . . .	79

# Lista de Algoritmos

1	: MCMC-MH para estimação da distribuição a posteriori de $\theta$ . . . . .	28
2	: ABC para estimação da distribuição a posteriori de $\theta$ . . . . .	29
3	: ABC para estimação da distribuição a posteriori de $\theta$ (Nunes e Balding (2010)). . . . .	30
4	: Estimação de $p_B$ para uma medida de teste $T(\mathbf{x}, \theta)$ . . . . .	34
5	: ABC-MC. . . . .	34
6	: ABC-PPMC. . . . .	35



# Capítulo 1

## Distribuições de cauda pesada

### 1.1 Introdução

É comum encontrarmos na natureza exemplos de fenômenos que superam de maneira exarcebada seu histórico de ocorrências. Por exemplo, até 2005, o furacão que mais havia ocasionado gastos nos EUA foi o nomeado Andrew (1992), custando cerca de 42.7 bilhões de dólares (2014 USD). Em 2005, o furacão Katrina passou a ser o novo recordista em gastos, custando em torno de 129 bilhões de dólares (2014 USD), ou seja, três vezes o valor do recordista anterior. Neste tipo de problema, a modelagem usando distribuições de cauda pesada surge como uma opção natural.

Neste texto, temos como objetivo reunir e desenvolver técnicas de inferência bayesiana para distribuições de cauda pesada, que não envolvam o cálculo explícito da função de verossimilhança. Estaremos interessados em fazer inferência para os parâmetros que representam o peso caudal das distribuições. Inicialmente, exploraremos propriedades de algumas classes de distribuições de cauda pesada, e desenvolveremos algumas medidas associadas ao peso caudal. Discutiremos as características da técnica ABC para a aproximação da distribuição a posteriori, e introduziremos seu uso, aliado à distribuição preditiva a posteriori, para seleção de modelos. Apresentaremos exemplos simulados que exibirão diversas características das técnicas desenvolvidas ao longo do texto. E, por fim, apresentaremos uma aplicação em dados reais relacionados à hidrologia.

#### 1.1.1 Preliminares

Neste capítulo, exploraremos propriedades de algumas classes de distribuições de variáveis aleatórias estritamente positivas. Mais especificamente, estamos interessados em características da cauda direita dessas distribuições. Na Seção 1.2 apresentaremos as três subclasses de cauda pesada que serão o foco de nosso estudo: a classe de cauda longa; a classe subexponencial; e a

classe de variação regular. Veremos que existe uma hierarquia nesses três conjuntos, sendo a classe de cauda longa a mais ampla. A Seção 1.3 apresenta algumas possíveis medidas de diagnóstico gerais para cauda pesada.

Na seguinte definição, introduzimos o conceito de função cauda que será amplamente utilizado neste texto.

**Definição 1:** Para qualquer distribuição  $F$  com suporte em  $\mathbb{R}$ , definimos a função cauda (ou função de sobrevivência)  $\bar{F}$  por

$$\bar{F}(x) = 1 - F(x), \quad x \in \mathbb{R}.$$

Como na literatura o termo “distribuição de cauda pesada” varia de acordo com a área de interesse do estudo, nos ateremos àquela que será dada na Definição 2, em função da ausência de momentos exponenciais (positivos).

**Definição 2** (Foss et al. (2013)): Dizemos que  $F$  é uma distribuição de cauda pesada (à direita) se e somente se

$$\limsup_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} = \infty, \quad \forall \lambda > 0.$$

A classe formada por todas essas distribuições é denominada classe de distribuições de cauda pesada e será denotada  $\mathcal{CP}$ .

**Exemplo 3:** A distribuição Pareto pertence a  $\mathcal{CP}$ . De fato, seja  $X$  uma variável aleatória com distribuição Pareto( $\mu, \alpha$ ),  $\alpha > 0$  e  $\mu > 0$ , com

$$\bar{F}(x) = \begin{cases} \left(\frac{\mu}{x}\right)^\alpha, & \text{se } x \geq \mu, \\ 1 & \text{se } x < \mu. \end{cases}$$

Assim, para  $\lambda > 0$  fixado,

$$\begin{aligned} \limsup_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} &= \limsup_{x \rightarrow \infty} \left(\frac{\mu}{x}\right)^\alpha e^{\lambda x} \\ &= \mu^\alpha \limsup_{x \rightarrow \infty} \frac{e^{\lambda x}}{x^\alpha} \end{aligned}$$

$$=: \mu^\alpha \limsup_{x \rightarrow \infty} g(x) = \infty,$$

pois a função  $g(x) = e^{\lambda x}/x^\alpha$  é crescente para  $x > \frac{\alpha}{\lambda}$ . Logo, a distribuição Pareto pertence a  $\mathcal{CP}$ .

Neste texto, chamaremos de classe de distribuições de cauda leve ( $\mathcal{CL}$ ) a classe formada por todas as distribuições em  $\mathbb{R}_+$ , cujo limite apresentado na Definição 2 assume valor finito para algum  $\lambda > 0$ , ou seja, as distribuições de  $\mathcal{CL}$  possuem pelo menos um momento exponencial finito. Comumente iremos comparar os resultados obtidos para distribuições de cauda pesada e leve. Escolhemos para isso a distribuição exponencial para representar os elementos de  $\mathcal{CL}$ , pois, como veremos adiante, algumas probabilidades relacionadas à cauda dessa distribuição independem de sua média.

**Exemplo 4:** A distribuição exponencial pertence a  $\mathcal{CL}$ . De fato, seja  $X$  uma variável aleatória com distribuição  $\text{Exp}(\theta)$ ,  $\theta > 0$ , com

$$\bar{F}(x) = \begin{cases} e^{-\theta x}, & \text{se } x > 0, \\ 1, & \text{se } x \leq 0. \end{cases}$$

Assim, para  $\lambda > 0$  fixado,

$$\begin{aligned} \limsup_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} &= \limsup_{x \rightarrow \infty} e^{-\theta x} e^{\lambda x} \\ &= \limsup_{x \rightarrow \infty} e^{x(\lambda - \theta)} = 0, \quad \text{se } \theta > \lambda. \end{aligned}$$

Logo, a distribuição exponencial pertence a  $\mathcal{CL}$ .

Faremos também comparações entre caudas de distribuições de  $\mathcal{CP}$ , mediante algum critério, para julgarmos qual cauda é mais pesada do que outra. Para isso, apresentamos a seguinte definição.

**Definição 5** (Halliwell (2013)): *Sejam  $F_1$  e  $F_2$  duas distribuições distintas de variáveis aleatórias positivas. Diremos que a distribuição  $F_2$  possui cauda mais pesada que a distribuição  $F_1$ , se*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}_1(x)}{\bar{F}_2(x)} = 0.$$

**Exemplo 6:** *Sejam  $X_1$  e  $X_2$  variáveis aleatórias com distribuição Weibull( $\lambda, k_1$ ) e Weibull( $\lambda, k_2$ ),  $\lambda > 0$ , respectivamente. Denotemos por  $F_i$  a função de distribuição de  $X_i$ ,  $i = 1, 2$ . Suponhamos que  $k_1 < k_2$ . Então  $F_1$  possui cauda mais pesada que  $F_2$ .*

*De fato, seja  $X_i$  uma variável aleatória com distribuição Weibull( $\lambda, k_i$ ), com*

$$\bar{F}_i(x) = e^{-\left(\frac{x}{\lambda}\right)^{k_i}}, \quad i = 1, 2.$$

*Então*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}_2(x)}{\bar{F}_1(x)} = \lim_{x \rightarrow \infty} \frac{e^{-\left(\frac{x}{\lambda}\right)^{k_2}}}{e^{-\left(\frac{x}{\lambda}\right)^{k_1}}} = \lim_{x \rightarrow \infty} \exp \left[ -\frac{1}{\lambda^{k_2}} x^{k_2} + \frac{1}{\lambda^{k_1}} x^{k_1} \right] = 0,$$

*uma vez que  $k_1 < k_2$ .*

## 1.2 Classes de cauda pesada

### 1.2.1 Classe de cauda longa

**Definição 7** (Foss et al. (2013)): *Dizemos que  $F$  é uma distribuição de cauda longa se e somente se*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1, \quad \forall y > 0.$$

*A classe formada por todas essas distribuições é denominada classe de distribuições de cauda longa e a denotaremos por  $\mathcal{L}$ .*

Como dito anteriormente, esta é a mais ampla classe de distribuições de cauda pesada com a qual trabalharemos. Em Foss et al. (2013) é demonstrado que  $\mathcal{L} \subset \mathcal{CP}$ . Tal demonstração é feita em duas etapas. A primeira consiste em mostrar que se  $F$  é uma distribuição de cauda longa, temos convergência uniforme do limite apresentado na Definição 7, para  $y$  em intervalos compactos. Em seguida, encontra-se limitantes para  $\bar{F}(x)e^{\lambda x}$ , com  $\lambda > 0$  fixado, e prova-se que  $\liminf_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} = \infty$ , e conseqüentemente,  $\limsup_{x \rightarrow \infty} \bar{F}(x)e^{\lambda x} = \infty$ .

## 1.2.2 Classe subexponencial

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com distribuição  $F$ . Denotemos por  $F^{n*}$  a  $n$ -ésima convolução de  $F$  e suponhamos que

$$\overline{F^{n*}}(x) \sim n\bar{F}(x) \sim Pr(\text{Max}\{X_1, X_2, \dots, X_n\} > x), \quad x \rightarrow \infty, \quad (1)$$

onde “ $\sim$ ” significa que a razão entre as quantidades de cada lado do símbolo converge para um, quando  $x \rightarrow \infty$ .

Notemos que, intuitivamente, a equação (1) nos diz que a distribuição da soma  $X_1 + X_2 + \dots + X_n$  possui o mesmo comportamento assintótico que a do máximo dessas variáveis. Ou seja, dado um alto limiar estabelecido, o excesso dessa soma é devido, praticamente, ao maior valor da amostra. Com isso, podemos considerar uma subclasse das distribuições de cauda pesada tais que a cauda da soma é essencialmente representada pela cauda do máximo. Essa é a caracterização intuitiva para a classe de distribuições subexponenciais. Apresentamos a seguir a definição formal de uma distribuição subexponencial.

**Definição 8** (Foss et al. (2013)): *Uma função de distribuição  $F$  com suporte em  $(0, \infty)$  é uma distribuição subexponencial, se*

$$\lim_{x \rightarrow \infty} \frac{\overline{F^{2*}}(x)}{\bar{F}(x)} = 2. \quad (2)$$

*A classe formada por todas essas distribuições é denominada classe de distribuições subexponenciais e será denotada por  $\mathcal{S}$ .*

A classe subexponencial tem sua origem nos anos de 1960 e seu conceito foi introduzido por Chistyakov (1964) no contexto de processos estocásticos, mais especificamente, em processos de ramificação. Aplicações podem ser encontradas em teoria de filas, passeios aleatórios e teoria de probabilidade aplicada. A Tabela 1 apresenta algumas distribuições de  $\mathcal{S}$ . Mais detalhes da classe  $\mathcal{S}$  podem ser encontrados em Embrechts et al. (1979), Foss et al. (2013) e Klüppelberg (1989).

**Proposição 9** (Santos (2007)): *Se  $F \in \mathcal{S}$ , então  $F \in \mathcal{L}$ .*

*Demonstração.* A demonstração pode ser encontrada em Santos (2007), e baseia-se em considerar duas variáveis aleatórias,  $X_1$  e  $X_2$ , positivas e independentes com distribuição  $F \in \mathcal{S}$ , e, utilizando

**Tabela 1:** Exemplos de distribuições subexponenciais.

Distribuição	$\bar{F}(x)$ ou $f(x)$	Parâmetros
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{[\log(x)-\mu]^2}{2\sigma^2}}$	$\mu \in \mathbb{R}, \sigma > 0$
Benktander tipo I	$\bar{F}(x) = \left(1 + 2\frac{\beta}{\alpha} \log(x)\right) e^{-\beta(\log(x))^2} x^{-(\alpha+1)}$	$\alpha, \beta > 0$
Benktander tipo II	$\bar{F}(x) = e^{\frac{\alpha}{\beta}(1-x^\beta)} x^{-1+\beta}$	$\alpha > 0, 0 < \beta \leq 1$
Weibull	$\bar{F}(x) = e^{-\left(\frac{x}{\lambda}\right)^k}$	$\lambda > 0, 0 < k < 1$

o fato de  $\bar{F}(x)$  ser decrescente, encontrar um limitante inferior para a razão

$$\frac{\bar{F}^{2*}(x)}{\bar{F}(x)} = \frac{1 - \int_0^x Pr(X_1 \leq x-t | X_2 = t) dF(t)}{\bar{F}(x)}.$$

Por fim, se considerarmos  $y$  fixo,  $x \geq y > 0$ , notaremos que o limitante inferior citado anteriormente é função de  $\frac{\bar{F}(x-y)}{\bar{F}(x)}$  e que esta converge para 1, quando  $x \rightarrow \infty$ .

□

Em Embrechts e Goldie (1982) e Foss et al. (2009) encontram-se discussões aprofundadas sobre condições suficientes e necessárias para que misturas de distribuições subexponenciais continuem sendo subexponenciais.

### 1.2.3 Classe de variação regular

Nesta seção, discutiremos uma classe de distribuições muito importante: a classe de distribuições de variação regular. Contudo, devemos explorar um pouco a teoria de funções de variação regular. Uma discussão mais detalhada sobre o assunto pode ser encontrada em Bingham et al. (1989) e Feller (1971).

**Definição 10** (Bingham et al. (1989)): *Uma função  $h$ , mensurável e positiva em  $(0, \infty)$ , é de variação regular no infinito (ou simplesmente de variação regular) com índice  $\alpha \in \mathbb{R}$  se*

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha, \quad t > 0.$$

Usaremos a notação  $h \in \mathcal{R}_\alpha$ . Se  $\alpha = 0$ , dizemos que a função é de variação lenta.

**Exemplo 11:** Sejam  $\mu, \alpha > 0$ . A função  $h(x) = (\mu/x)^\alpha$ , definida para  $x \geq \mu$ , pertence a  $\mathcal{R}_{-\alpha}$ . De fato, se tomarmos  $t > 0$ , temos que

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = \lim_{x \rightarrow \infty} \frac{\left(\frac{\mu}{tx}\right)^\alpha}{\left(\frac{\mu}{x}\right)^\alpha} = \lim_{x \rightarrow \infty} \frac{x^\alpha}{x^\alpha t^\alpha} = t^{-\alpha}.$$

**Proposição 12** (Bingham et al. (1989)): Se  $L$  é uma função de variação lenta e  $\alpha > 0$ , então

$$x^\alpha L(x) \rightarrow \infty, \quad x^{-\alpha} L(x) \rightarrow 0, \quad \text{quando } x \rightarrow \infty.$$

*Demonstração.* Bingham et al. (1989) apresentam um possível caminho para esta demonstração. Prova-se, inicialmente, que uma função de variação lenta,  $L$ , pode ser escrita como

$$L(x) = c(x) \exp \left\{ \int_a^x \frac{\varepsilon(u)}{u} du \right\}, \quad x \geq a,$$

para algum  $a > 0$ , onde  $c(\cdot)$  e  $\varepsilon(\cdot)$  são funções tais que  $c(x) \rightarrow k \in (0, \infty)$  e  $\varepsilon(x) \rightarrow 0$ , quando  $x \rightarrow \infty$ . Em seguida, usa-se tal notação para mostrar que, para  $\alpha > 0$  fixo,

$$\lim_{x \rightarrow \infty} x^\alpha L(x) = \infty, \quad \lim_{x \rightarrow \infty} x^{-\alpha} L(x) = 0.$$

□

Notemos que, se decomposmos a função  $h$  da Definição 10 como  $h(x) = x^\alpha g(x)$ , onde  $g$  é uma função real apropriada, temos, para  $t > 0$ ,

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha \iff \lim_{x \rightarrow \infty} \frac{(tx)^\alpha g(tx)}{x^\alpha g(x)} = t^\alpha \iff t^\alpha \lim_{x \rightarrow \infty} \frac{g(tx)}{g(x)} = t^\alpha \iff \lim_{x \rightarrow \infty} \frac{g(tx)}{g(x)} = 1.$$

Assim, podemos reescrever a função  $h(x)$  como

$$h(x) = x^\alpha L(x), \tag{3}$$

onde  $L(x)$  é uma função de variação lenta apropriada. A equação (3) será usada nas demonstrações

das duas proposições que introduzimos a seguir, que constataam que a classe  $\mathcal{R}_\alpha$  é fechada para combinações lineares.

**Proposição 13:** *Sejam  $\alpha_1, \alpha_2 \in \mathbb{R}$ , e sejam  $f_1 \in \mathcal{R}_{\alpha_1}$  e  $f_2 \in \mathcal{R}_{\alpha_2}$ . Então  $f_1 + f_2 \in \mathcal{R}_\alpha$  com  $\alpha = \max\{\alpha_1, \alpha_2\}$ .*

*Demonstração.* Sem perda de generalidade, suponhamos que  $\alpha_1 \leq \alpha_2$ . Da equação (3), existem funções de variação lenta,  $L_1$  e  $L_2$ , tais que  $f_1(x) = x^{\alpha_1}L_1(x)$ ,  $f_2(x) = x^{\alpha_2}L_2(x)$ . Assim,

$$\begin{aligned} f(x) = f_1(x) + f_2(x) &= x^{\alpha_1}L_1(x) + x^{\alpha_2}L_2(x) \\ &= x^{\alpha_2} \left[ L_2(x) + x^{\alpha_1 - \alpha_2}L_1(x) \right] \\ &=: x^{\alpha_2}L(x). \end{aligned}$$

Como  $\alpha_1 - \alpha_2 \leq 0$ , temos, pela Proposição 12, que  $x^{\alpha_1 - \alpha_2}L_1(x) \rightarrow 0$ , quando  $x \rightarrow \infty$ . Logo,  $L \sim L_2$  é uma função de variação lenta. E portanto,  $f$  é de variação regular com índice de variação  $\alpha = \alpha_2$ .

□

**Proposição 14:** *Seja  $\alpha \in \mathbb{R}$  e seja  $f_1 \in \mathcal{R}_\alpha$ . Então  $f = pf_1 \in \mathcal{R}_\alpha$ ,  $\forall p \in \mathbb{R}$ ,  $p \neq 0$ .*

*Demonstração.* Como  $f_1 \in \mathcal{R}_\alpha$ , podemos escrever, pela equação (3),  $f_1(x) = x^\alpha L_1(x)$ , onde  $L_1$  é função de variação lenta apropriada. Assim,

$$f(x) = pf_1(x) = x^\alpha(pL_1(x)) =: x^\alpha L(x).$$

Para que  $f$  seja de variação regular, devemos mostrar que  $L$  é de variação lenta. De fato, como  $p \neq 0$ ,

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = \lim_{x \rightarrow \infty} \frac{pL_1(tx)}{pL_1(x)} = \lim_{x \rightarrow \infty} \frac{L_1(tx)}{L_1(x)} = 1, \quad \forall t > 0.$$

□

A seguir, apresentamos a definição de distribuições de variação regular, encontrada em Mikosch (1999).

**Definição 15** (Mikosch (1999)): *Uma variável aleatória não-negativa  $X$  é dita ter uma distribui-*

ção de variação regular com índice de variação  $\alpha \geq 0$ , se  $\bar{F}(x)$  for uma função de variação regular com índice  $-\alpha$ .

A classe formada por todas essas distribuições é denominada classe de distribuições de variação regular com índice de variação  $\alpha$  e será denotada por  $\mathcal{VR}_\alpha$ .

É interessante notar que quanto menor for o valor do índice de variação  $\alpha$ , mais lentamente  $\bar{F}(x)$  converge para zero, ou seja, mais pesada é a cauda da distribuição. A Tabela 2 apresenta algumas distribuições da classe de variação regular.

**Corolário 16:** A classe  $\mathcal{VR}_\alpha$  é fechada com relação a misturas finitas.

*Demonstração.* Segue direto das Proposições 13 e 14. □

**Proposição 17:** Suponha que  $X$  é uma variável aleatória não-negativa com distribuição de variação regular com índice  $\alpha > 0$ . Então

$$\begin{aligned} E(X^\beta) &< \infty & \text{se } \beta < \alpha, \\ E(X^\beta) &= \infty & \text{se } \beta > \alpha. \end{aligned}$$

*Demonstração.* Para esta demonstração, faremos uso de dois lemas. O primeiro encontra-se em Bingham et al. (1989) e o segundo em Feller (1971).

**Lema 18:** Para qualquer lei  $F$  em  $[0, \infty)$ , e qualquer  $\alpha > 0$ , temos que

$$\int_{[0, \infty)} x^\alpha dF(x) = \alpha \int_0^\infty x^{\alpha-1} (1 - F(x)) dx.$$

**Lema 19:** Seja  $L$  uma função de variação lenta e  $p \in \mathbb{R}$ . Então,

$$\lim_{x \rightarrow \infty} \int_0^x y^p L(y) dy$$

converge para  $p < -1$  e diverge para  $p > -1$ .

Assim, seja  $X$  uma variável aleatória não-negativa com distribuição de variação regular com índice  $\alpha > 0$ . Então, pelo Lema 18, temos que, para  $\beta > 0$ , e para uma função de variação

**Tabela 2:** Exemplos de distribuições de variação regular.

Distribuição	$\bar{F}(x)$ ou $f(x)$	Parâmetros	Índice de variação
Pareto	$\bar{F}(x) = \left(\frac{x}{\mu}\right)^{-\alpha}$	$\mu, \alpha > 0$	$\alpha$
Burr tipo XII	$\bar{F}(x) = (1 + x^c)^{-\alpha}$	$\alpha, c > 0$	$c\alpha$
Log-Gama	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)} [\log(x)]^{\beta-1} x^{-\alpha-1}$	$\alpha, \beta > 0$	$\alpha$

lenta  $L$  apropriada (ver equação (3)),

$$\begin{aligned} E(X^\beta) &= \int_{[0, \infty)} x^\beta dF(x) = \beta \int_0^\infty x^{\beta-1} (\bar{F}(x)) dx \\ &= \beta \int_0^\infty x^{\beta-1} x^{-\alpha} L(x) dx = \beta \int_0^\infty x^{\beta-\alpha-1} L(x) dx. \end{aligned}$$

Pelo Lema 19, temos que  $E(X^\beta)$  converge para  $\beta < \alpha$  e diverge para  $\beta > \alpha$ .

□

**Proposição 20** (Santos (2007)): *Seja  $\alpha > 0$ . Se  $F \in \mathcal{VR}_\alpha$ , então  $F \in \mathcal{S}$ .*

*Demonstração.* A demonstração pode ser encontrada em Santos (2007), e baseia-se em considerar  $X_1, X_2$  variáveis aleatórias independentes com distribuição  $F \in \mathcal{VR}_\alpha$ ,  $\alpha > 0$ , e encontrarmos limitantes inferior e superior para  $\overline{F^{2*}}(x) = Pr(X_1 + X_2 > x)$ . Com esses limitantes, e fazendo uso da definição de uma distribuição de variação regular, chega-se a

$$\lim_{x \rightarrow \infty} \frac{\overline{F^{2*}}(x)}{\bar{F}(x)} = 2.$$

□

### 1.3 Medidas de evidência de cauda pesada

Nesta seção exploraremos propriedades de algumas medidas que podem evidenciar se uma certa distribuição possui cauda pesada, e também nos dar uma ideia da magnitude do peso

da mesma a partir de uma amostra aleatória da distribuição. Como apresentado na seção anterior, o índice de variação  $\alpha$  representa o peso da cauda de uma distribuição de variação regular. No entanto, em distribuições exclusivamente subexponenciais ou de cauda longa não conseguimos construir um parâmetro, cuja relação com a existência de momentos e peso da cauda seja tão clara quanto o índice de variação regular. Daí a necessidade de explorarmos outras características da cauda da distribuição. A seguir estudaremos o comportamento de três delas: razão entre estatísticas de ordem; índice de obesidade; e função de excesso médio. Exploraremos essas medidas para o caso de uma distribuição exponencial, que possui cauda leve, e as compararemos com os resultados obtidos para distribuições de cauda pesada. É importante ressaltar que as duas primeiras medidas apresentadas não dependem de nenhuma suposição quanto à distribuição dos dados, o que torna seus usos indicados em diversas situações.

### 1.3.1 Razão entre estatísticas de ordem

Uma característica de amostragens provenientes de uma distribuição de cauda pesada é que geralmente existem alguns valores observados que são excessivamente grandes quando comparados com toda amostra. Isso sugere que as distâncias entre os maiores valores tendem a ser maiores do que aquelas para valores menores. Dessa forma, faz sentido estudarmos o comportamento da razão entre os dois maiores valores observados de uma amostra. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias i.i.d. com distribuição  $F$  e densidade  $f$ . Sejam  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes à amostra. Iremos explorar a seguinte probabilidade:

$$Ra_{n,k}(X) = Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > k\right), \quad k > 1, \quad (4)$$

ou seja, a probabilidade de que o maior valor observado seja pelo menos  $k$  vezes o valor do segundo. Para isso, necessitaremos dos seguintes teoremas, extraídos de Ross (2009).

**Teorema 21:** *Sejam  $X_1, \dots, X_n$  amostra aleatória de  $X$  (variável positiva) com distribuição  $F$  e densidade  $f$ . Sejam  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes à amostra. Temos então que a densidade de  $X_{(l)}$ ,  $1 \leq l \leq n$ , é dada por*

$$f_{X_{(l)}}(x) = \frac{n!}{(l-1)!(n-l)!} [F(x)]^{l-1} [1-F(x)]^{n-l} f(x) \mathbb{1}_{\{x>0\}}.$$

**Teorema 22:** *Consideremos as mesmas condições do Teorema 21. Temos então que a densidade*

conjunta de  $X_{(l)}$  e  $X_{(m)}$ , para  $1 \leq l \leq m \leq n$ , é dada por

$$f_{X_{(l)}, X_{(m)}}(x, y) = \frac{n!}{(l-1)!(m-l-1)!(n-m)!} [F(x)]^{l-1} [F(y) - F(x)]^{m-l-1} \\ \times [1 - F(y)]^{n-m} f(x) f(y) \mathbb{1}_{\{x < y\}}.$$

Temos pelo teorema anterior que

$$f_{X_{(n-1)}, X_{(n)}}(x, y) = n(n-1) [F(x)]^{n-2} f(x) f(y) \mathbb{1}_{\{x < y\}}.$$

Dessa forma, dado  $k > 1$ ,

$$\begin{aligned} Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > k\right) &= Pr(X_{(n)} > kX_{(n-1)}) \\ &= \int_0^\infty \int_{kx}^\infty f_{X_{(n-1)}, X_{(n)}}(x, y) dy dx \\ &= n(n-1) \int_0^\infty [F(x)]^{n-2} f(x) \left[ \int_{kx}^\infty f(y) dy \right] dx \\ &= n(n-1) \int_0^\infty [F(x)]^{n-2} f(x) [1 - F(kx)] dx. \end{aligned} \quad (5)$$

Em Cooke e Nieboer (2011) é estudado o uso de  $k = 2$  para análises de distribuições de cauda pesada, que utilizaremos neste trabalho.

**Exemplo 23:** Seja  $X$  uma variável aleatória com distribuição exponencial com média  $\frac{1}{\theta}$ ,  $\theta > 0$ . Sejam  $X_1, \dots, X_n$  amostra aleatória de  $X$ , e  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes a amostra. Então

$$Ra_{n,2}(X) = Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) = \frac{2}{n+1}.$$

De fato, pela equação (5), temos que

$$Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) = n(n-1) \int_0^\infty [1 - e^{-\theta x}]^{n-2} \theta e^{-\theta x} e^{-2\theta x} dx$$

$$\begin{aligned}
&= n(n-1) \int_0^\infty [1 - e^{-\theta x}]^{n-2} \theta e^{-3\theta x} dx \\
&= n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k \int_0^\infty (e^{-\theta x})^k \theta e^{-3\theta x} dx \\
&= n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k \int_0^\infty \theta e^{-\theta x(k+3)} dx \\
&= n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k \frac{1}{k+3} \\
&= n(n-1) \frac{2}{(n-1)n(n+1)} = \frac{2}{n+1}.
\end{aligned}$$

**Exemplo 24:** Seja  $X$  uma variável aleatória com distribuição Pareto( $1, \alpha$ ),  $\alpha > 0$ . Sejam  $X_1, \dots, X_n$  amostra aleatória de  $X$ , e  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes a amostra. Então

$$Ra_{n,2}(X) = Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) = \frac{1}{2^\alpha},$$

ou seja,  $Ra_{n,2}(X)$  não depende do tamanho da amostra considerada.

De fato, pela equação (5) e pela Tabela 2, temos que

$$\begin{aligned}
Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) &= n(n-1) \int_1^\infty [1 - x^{-\alpha}]^{n-2} \alpha x^{-(\alpha+1)} (2x)^{-\alpha} dx \\
&= n(n-1) \int_1^\infty [1 - x^{-\alpha}]^{n-2} \alpha 2^{-\alpha} x^{-2\alpha-1} dx \\
&= 2^{-\alpha} n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k \alpha \int_1^\infty x^{-(\alpha k + 2\alpha + 1)} dx \\
&= 2^{-\alpha} n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k \alpha \frac{1}{\alpha(k+2)} \\
&= 2^{-\alpha} n(n-1) \frac{1}{n(n-1)} = \frac{1}{2^\alpha}.
\end{aligned}$$

**Exemplo 25:** *Seja  $X$  uma variável aleatória com distribuição Weibull(1,  $k$ ),  $k > 0$ . Sejam  $X_1, \dots, X_n$  amostra aleatória de  $X$ , e  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes a amostra. Então*

$$Ra_{n,2}(X) = Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) = n(n-1) \text{Beta}(1 + 2^k, n-1),$$

onde *Beta* denota a função beta, ou seja,

$$\text{Beta}(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1} du, \quad \text{para } a, b > 0.$$

De fato, pela equação (5) e pela Tabela 2, temos que

$$\begin{aligned} Pr\left(\frac{X_{(n)}}{X_{(n-1)}} > 2\right) &= n(n-1) \int_0^\infty [1 - e^{-x^k}]^{n-2} kx^{k-1} e^{-x^k} e^{-2^k x^k} dx \\ &= n(n-1) \int_0^1 [1-u]^{n-2} u^{2^k} du \\ &= n(n-1) \text{Beta}(1 + 2^k, n-1). \end{aligned}$$

A Tabela 3 apresenta o valor numérico de  $Ra_{n,2}(X)$  para algumas distribuições. Pode-se facilmente mostrar que  $Ra_{n,2}(X)$  vai para zero mais rápido, quando  $n \rightarrow \infty$ , no caso de uma amostra exponencial do que nos casos das distribuições *Weibull*(1,  $k$ ),  $0 < k < 1$ , e *Pareto*(1,  $\alpha$ ),  $\alpha > 0$ . Podemos também notar que, dentro de uma mesma família, quanto mais pesada a cauda da distribuição, maior o valor associado a  $Ra_{n,2}(X)$ , para  $n$  fixo.

### 1.3.2 Índice de Obesidade

O índice de obesidade é mais uma tentativa de determinar, através de uma probabilidade, quão pesada é a cauda de uma distribuição dada. A definição a seguir foi extraída de Cooke e Nieboer (2011).

**Definição 26** (Cooke e Nieboer (2011)): *Sejam  $X_1, X_2, X_3, X_4$  amostra aleatória de  $X$  com distribuição  $F$ . Sejam  $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$  as estatísticas de ordem referentes à amostra. Definimos*

**Tabela 3:** Valores numéricos de  $Ra_{n,2}(X)$  calculados para algumas distribuições.

Distribuição	$n = 10$	$n = 20$	$n = 40$	$n = 80$
$Exp(\theta), \theta > 0$	0.1818	0.0952	0.0488	0.0247
$Pareto(1, 1)$	0.5000	0.5000	0.5000	0.5000
$Pareto(1, 3)$	0.1250	0.1250	0.1250	0.1250
$Weibull(1, 0.8)$	0.2724	0.1680	0.1021	0.0616
$Weibull(1, 0.2)$	0.7551	0.6893	0.6182	0.5583

o índice de obesidade de  $X$  por

$$Ob^*(X) = Pr(X_{(4)} + X_{(1)} > X_{(2)} + X_{(3)}).$$

Neste texto, introduzimos o conceito de índice de obesidade generalizado que é apresentado na próxima definição. A intenção é acrescentar a informação do tamanho da amostra na investigação sobre o peso caudal da distribuição em questão.

**Definição 27:** *Sejam  $X_1, \dots, X_n$  amostra aleatória de  $X$  com distribuição  $F$ . Sejam  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes à amostra. Definimos o índice de obesidade generalizado de  $X$  por*

$$Ob(X, n) = Pr(X_{(n)} + X_{(n-3)} > X_{(n-2)} + X_{(n-1)}).$$

Notemos que podemos calcular o índice de obesidade generalizado de  $X$  da seguinte forma:

$$Ob(X, n) = Pr(W > Z) = \int_0^\infty Pr(W > Z | Z = z) f_Z(z) dz = \int_0^\infty Pr(W > z) f_Z(z) dz, \quad (6)$$

onde  $W = X_{(n)} - X_{(n-1)}$  e  $Z = X_{(n-2)} - X_{(n-3)}$ .

Temos pelo Teorema 22 (ver página 11) que

$$f_{X_{(n-1)}, X_{(n)}}(x_1, x_2) = \frac{n!}{(n-2)!} [F(x_1)]^{n-2} f(x_1) f(x_2) \mathbb{1}_{\{x_1 < x_2\}}.$$

Definamos a transformação

$$h(x_1, x_2) = (x_2 - x_1, x_2)$$

com inversa

$$h^{-1}(w, y) = (y - w, y).$$

Notemos que  $0 < w < y < \infty$ . Denotemos por  $J$  a matriz jacobiana referente à transformação  $h$ . Assim,

$$J = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Portanto,  $\det(J) = -1$ . Como  $h$  é uma transformação um a um, pela regra do jacobiano, a densidade conjunta de  $(W, Y)$  é

$$\begin{aligned} g_{W,Y}(w, y) &= |\det(J)| f_{X_{(n-1)}, X_{(n)}}(y - w, w) \\ &= n(n-1)[F(y-w)]^{n-2} f(y-w) f(y) \mathbb{1}_{\{0 < w < y < \infty\}}. \end{aligned}$$

Dessa forma,

$$\begin{aligned} f_W(w) &= \int_w^\infty g_{W,Y}(w, y) dy \\ &= n(n-1) \int_w^\infty [F(y-w)]^{n-2} f(y-w) f(y) dy \\ &= n(n-1) \int_0^\infty [F(x)]^{n-2} f(x) f(x+w) dx. \end{aligned} \tag{7}$$

De maneira análoga, chegamos à densidade de  $Z = X_{(n-2)} - X_{(n-3)}$ :

$$f_Z(z) = \frac{n(n-1)(n-2)(n-3)}{2} \int_0^\infty [F(x)]^{n-4} [1 - F(x+z)]^2 f(x) f(x+z) dx. \tag{8}$$

**Exemplo 28:** Seja  $X$  uma variável aleatória com distribuição exponencial com média  $\frac{1}{\theta}$ ,  $\theta > 0$ . Então  $Ob(X, n) = \frac{3}{4}$ .

De fato, se definirmos  $X_1, \dots, X_n$  amostra aleatória de  $X$ , e  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem referentes à amostra, teremos pelas equações (7) e (8) que

$$\begin{aligned}
 f_W(w) &= n(n-1) \int_0^\infty [1 - e^{-\theta x}]^{n-2} \theta e^{-\theta x} \theta e^{-\theta(x+w)} dx \\
 &= n(n-1)\theta^2 \left( \frac{e^{-\theta w}}{n(n-1)\theta} \right) \\
 &= \theta e^{-\theta w} \mathbb{1}_{\{w>0\}}
 \end{aligned} \tag{9}$$

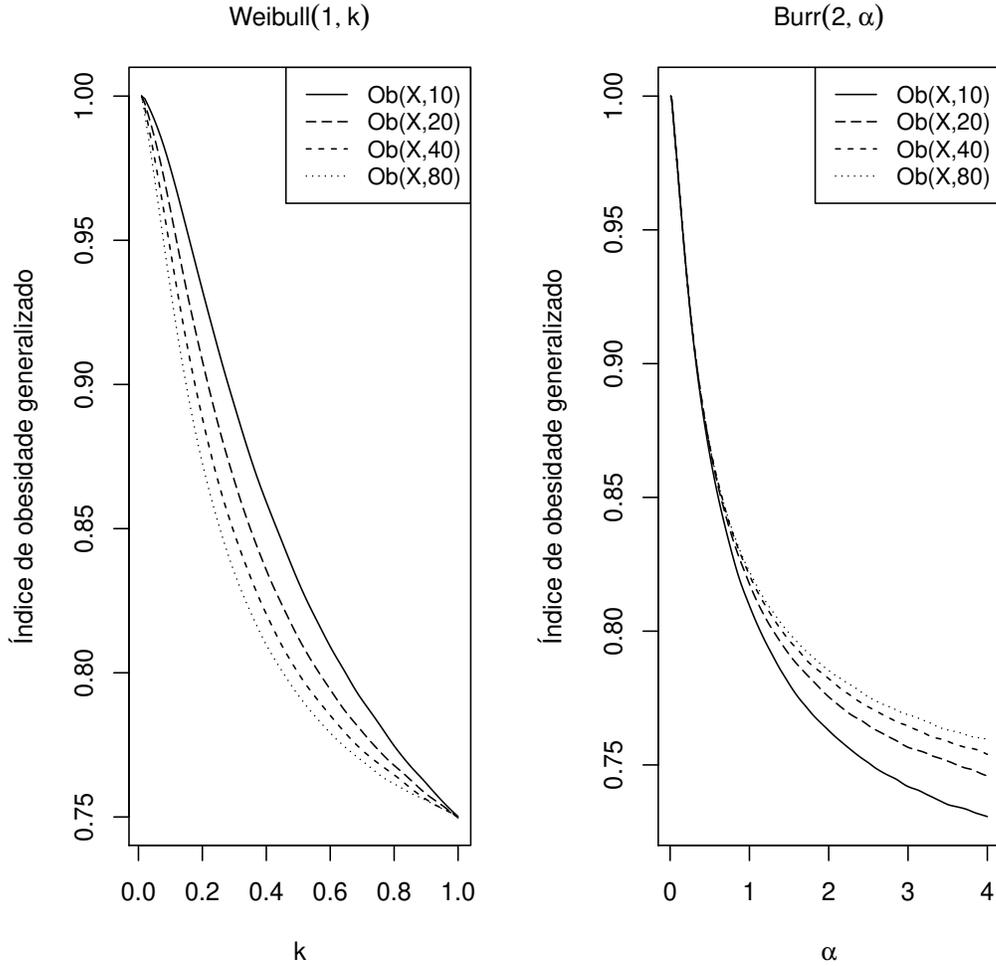
e

$$\begin{aligned}
 f_Z(z) &= \frac{n(n-1)(n-2)(n-3)}{2} \int_0^\infty [1 - e^{-\theta x}]^{n-4} [e^{-\theta(x+z)}]^2 \theta e^{-\theta x} \theta e^{-\theta(x+z)} dx \\
 &= \frac{n(n-1)(n-2)(n-3)\theta^2}{2} \left( \frac{6e^{-3\theta z}}{n(n-1)(n-2)(n-3)\theta} \right) \\
 &= 3\theta e^{-3\theta z} \mathbb{1}_{\{z>0\}}.
 \end{aligned} \tag{10}$$

Assim, pelas equações (6), (9) e (10), temos que

$$\begin{aligned}
 Ob(X, n) &= \int_0^\infty Pr(W > z) f_Z(z) dz \\
 &= \int_0^\infty e^{-\theta z} 3\theta e^{-3\theta z} dz \\
 &= \frac{3}{4} \int_0^\infty 4\theta e^{-4\theta} dz = \frac{3}{4}.
 \end{aligned}$$

Observemos que no exemplo anterior obtivemos que  $Ob(X, n)$  é constante em relação a  $n$ . No entanto, isso não é sempre verdade. Algumas famílias de distribuições parecem ser mais sensíveis ao tamanho amostral que outras. A Figura 1 apresenta os índices de obesidade generalizados para amostras de diferentes tamanhos provenientes de distribuições  $Weibull(1, k)$  e  $Burr(2, \alpha)$ . Notemos que nos casos apresentados, quanto mais pesada a cauda, maior o valor



**Figura 1:** Gráficos de índices de obesidade generalizados das distribuições  $Weibull(1, k)$  e  $Burr(2, \alpha)$ .

estimado de  $Ob(X, n)$ .

O índice de obesidade generalizado possui um grande inconveniente que é sua manipulação analítica, pois nem sempre a equação (6) é de fácil trato (sendo em alguns casos analiticamente não tratável). Por isso, integrações numéricas e simulações apresentam papéis fundamentais neste procedimento.

### 1.3.3 Função de excesso médio

A função de excesso médio (FEM) representa o valor esperado do quanto uma variável aleatória  $X$  excede um determinado limiar  $u$ , dado que esta é maior que  $u$ . Neste texto utilizaremos a definição formal de função de excesso médio para distribuições de variáveis positivas cujos

suportes não possuem limitante superior.

**Definição 29** (Cooke e Nieboer (2011)): *Seja  $X$  uma variável aleatória assumindo valores em  $(a, \infty)$ ,  $a \geq 0$ , e  $E(X) < \infty$ . Definimos a função de excesso médio de  $X$  como*

$$e_X(u) = E(X - u | X > u), \quad u \geq a.$$

Existem diversas formas de se calcular  $e_X(u)$ . A proposição a seguir apresenta uma forma simples e que utilizaremos neste texto.

**Proposição 30:** *Seja  $X$  uma variável aleatória com suporte em  $[a, \infty)$ ,  $a \geq 0$ , e função sobrevivência  $\bar{F}$ . Temos que*

$$e_X(u) = \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(y) dy, \quad u \geq 0. \quad (11)$$

*Demonstração.* Dado  $u \geq a$ ,

$$\begin{aligned} e_X(u) &= E(X - u | X > u) = \int_0^\infty Pr(X - u > t | X > u) dt \\ &= \int_0^\infty \frac{Pr(X - u > t, X > u)}{Pr(X > u)} dt \\ &= \frac{1}{\bar{F}(u)} \int_0^\infty Pr(X - u > t, X > u) dt \\ &= \frac{1}{\bar{F}(u)} \int_0^\infty Pr(X > u + t) dt \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty Pr(X > y) dy \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(y) dy. \end{aligned}$$

□

**Exemplo 31:** Seja  $X$  uma variável aleatória com distribuição  $Exp(\theta)$ ,  $\theta > 0$ . Então

$$e_X(x) = \frac{1}{\theta}.$$

De fato, pela Proposição 30, temos que

$$\begin{aligned} e_X(x) &= \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(t) dt = \frac{1}{e^{-\theta x}} \int_x^\infty e^{-\theta t} dt = e^{\theta x} \frac{1}{\theta} \left( -e^{-\theta t} \Big|_x^\infty \right) \\ &= e^{\theta x} \frac{1}{\theta} e^{-\theta x} = \frac{1}{\theta}. \end{aligned}$$

**Exemplo 32:** Seja  $X$  uma variável aleatória com distribuição  $Weibull(\lambda, k)$ ,  $\lambda, k > 0$ . Então

$$e_X(x) = \frac{\lambda e^{(\frac{x}{\lambda})^k}}{k} \gamma \left( \frac{1}{k}, \left( \frac{x}{\lambda} \right)^k \right),$$

onde  $\gamma$  denota a função gama incompleta superior, ou seja,

$$\gamma(a, b) = \int_b^\infty e^{-u} u^{a-1} du.$$

De fato, pela Proposição 30, temos que

$$e_X(x) = \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(t) dt = \frac{1}{e^{-(\frac{x}{\lambda})^k}} \int_x^\infty e^{-(\frac{t}{\lambda})^k} dt.$$

Fazendo a mudança de variável  $u = \left(\frac{t}{\lambda}\right)^k$  na integral, temos que

$$e_X(x) = e^{(\frac{x}{\lambda})^k} \frac{\lambda}{k} \int_{(\frac{x}{\lambda})^k}^\infty e^{-u} u^{\frac{1}{k}-1} du = \frac{\lambda e^{(\frac{x}{\lambda})^k}}{k} \gamma \left( \frac{1}{k}, \left( \frac{x}{\lambda} \right)^k \right).$$

Utilizando propriedades do limite da função de excesso médio, Beirlant e Teugels (1992) encontram uma relação linear entre  $\log e_X(x)$  e  $\log x$ , quando  $x \rightarrow \infty$ , para  $0 < k < 1$ .

A seguir, apresentamos o Teorema de Karamata, extraído de Bingham et al. (1989), que é um dos pilares fundamentais no estudo de funções de variação regular. Tal resultado será utilizado para demonstrar que existe uma relação assintótica entre o índice de variação regular e

a função de excesso médio. No Capítulo 3, discutiremos o uso da FEM como uma técnica gráfica para explorarmos características assintóticas de uma amostra observada.

**Teorema 33** (Teorema de Karamata): *Seja  $L$  uma função de variação lenta definida em  $[x_0, \infty)$ ,  $x_0 \geq 0$ , tal que para todo  $y \in [x_0, \infty)$  exista uma vizinhança  $V$  em que, para algum  $M > 0$ ,  $|L(x)| \leq M, \forall x \in V$ . Então*

1. para  $\alpha > -1$ ,

$$\int_{x_0}^x t^\alpha L(t) dt \sim (\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \rightarrow \infty;$$

2. para  $\alpha = -1$ ,

$$\frac{1}{L(x)} \int_x^\infty t^\alpha L(t) dt \rightarrow \infty, \quad x \rightarrow \infty;$$

3. para  $\alpha < -1$

$$\frac{x^{\alpha+1} L(x)}{\int_x^\infty t^\alpha L(t) dt} \rightarrow -(\alpha + 1), \quad x \rightarrow \infty.$$

**Proposição 34:** *Seja  $X$  uma variável aleatória pertencente a  $\mathcal{VR}_\alpha$ , com  $\alpha > 1$ . Então*

$$e_X(x) \sim \frac{x}{\alpha - 1}, \quad x \rightarrow \infty.$$

*Demonstração.* Pela Proposição 30 temos que

$$e_X(x) = \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(t) dt.$$

Como  $\bar{F} \in \mathcal{VR}_\alpha$ , sabemos que existe uma função de variação lenta,  $L$ , que nos permite escrever  $\bar{F}(x) = x^{-\alpha} L(x)$ . Então,

$$e_X(x) = \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(t) dt = \frac{\int_x^\infty t^{-\alpha} L(t) dt}{x^{-\alpha} L(x)} = x \frac{\int_x^\infty t^{-\alpha} L(t) dt}{x^{-\alpha+1} L(x)}.$$

Assim, temos pelo Teorema de Karamata

$$x \frac{\int_0^\infty t^{-\alpha} L(t) dt}{x^{-\alpha+1} L(x)} \rightarrow \frac{x}{\alpha - 1}, \quad x \rightarrow \infty.$$

□

# Capítulo 2

## Inferência bayesiana

### 2.1 Terminologias e preliminares

Iniciamos este capítulo definindo um modelo estatístico. Seguindo a definição apresentada por McCullagh (2002), um modelo estatístico  $\mathbf{P}$  é uma coleção de distribuições de probabilidade sobre o espaço amostral,  $\Omega$ , associado a um experimento aleatório. Podemos indexar os elementos de  $\mathbf{P}$  por um parâmetro  $\theta$ , assumindo valores em um espaço parâmetro  $\Theta$ , ou seja,  $\mathbf{P} = \{P_\theta \mid \theta \in \Theta\}$ , em que cada  $P_\theta$  é uma distribuição de probabilidade.

Dizemos que um modelo  $\mathbf{P}$  é paramétrico se  $\Theta$  possui dimensão finita, ou seja,  $\Theta \subseteq \mathbb{R}^d$ , para algum  $d \in \mathbb{N}$ . No caso em que  $\Theta$  possui dimensão infinita, dizemos que  $\mathbf{P}$  é um modelo não paramétrico.

Neste texto, consideraremos  $X_1, \dots, X_n$  variáveis aleatórias i.i.d. provenientes de alguma distribuição de probabilidade de  $\mathbf{P}$ , ou seja,

$$X_1, \dots, X_n \mid \theta \stackrel{i.i.d.}{\sim} P_\theta, \quad \theta \in \Theta,$$

e queremos tirar conclusões sobre o valor de  $\theta$  a partir das observações  $x_1, \dots, x_n$ .

Na teoria bayesiana de inferência estatística, inserimos uma distribuição de probabilidade a priori para o valor do parâmetro  $\theta$ ,  $\pi(\theta)$ . Em DeGroot e Schervish (2011) e Gelman et al. (2013) encontram-se discussões sobre o uso da distribuição a priori como uma medida de incerteza sobre o “verdadeiro” valor de  $\theta$ , antes de observarmos o resultado do experimento aleatório. Depois de observarmos uma amostra, podemos atualizar a distribuição do parâmetro, ou seja, podemos construir uma distribuição posterior de  $\theta$  após a observação de  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\pi(\theta \mid \mathbf{x})$ . A maneira de fazermos tal atualização é a aplicação direta do Teorema de Bayes, da seguinte forma (ver

DeGroot e Schervish (2011)):

$$\pi(\theta | \mathbf{x}) = \frac{\pi(\theta) f(\mathbf{x} | \theta)}{f(\mathbf{x})} = \frac{\pi(\theta) f(\mathbf{x} | \theta)}{\int_{\Theta} \pi(\theta) f(\mathbf{x} | \theta) d\theta}, \quad (12)$$

onde  $f(\mathbf{x} | \theta)$  denota a verossimilhança de  $\theta$  dada a amostra  $\mathbf{x}$ . Além disso, vale notar que o denominador da equação anterior é uma constante em  $\theta$ , ou seja, podemos escrever

$$\pi(\theta | \mathbf{x}) \propto \pi(\theta) f(\mathbf{x} | \theta).$$

É interessante perceber que a distribuição a posteriori é a união de duas fontes distintas de informação: o conhecimento prévio do observador que se manifesta na distribuição a priori; e a informação contida nas observações que se expressa na função de verossimilhança.

A escolha de distribuições a priori adequadas apresenta um papel muito importante em inferência bayesiana. Geralmente, desejamos que  $\pi(\cdot)$  seja uma distribuição que nos permita obter alguma distribuição a posteriori de fácil tratamento analítico. Uma possibilidade para obtermos posterioris conhecidas é a consideração de famílias conjugadas.

**Definição 35** (Gelman et al. (2013)): *Uma família  $\mathcal{F}$  de distribuições a priori para  $\theta$  é dita fechada, ou conjugada para amostragem de  $f(\mathbf{x} | \theta)$ , se para toda a distribuição a priori  $\pi(\theta) \in \mathcal{F}$ , a distribuição a posteriori  $\pi(\theta | \mathbf{x}) \in \mathcal{F}$ .*

Notemos que, em inferência bayesiana, a distribuição a posteriori representa toda a informação disponível sobre o parâmetro  $\theta$ , ou seja, essa distribuição é um estimador para  $\theta$ . Assim, conhecendo  $\pi(\theta | \mathbf{x})$  podemos obter estimativas pontuais, por região ou realizar testes de hipóteses.

Os problemas de inferência podem ser tratados sob a abordagem de teoria de decisão. Nesse contexto, o problema de estimação pontual pode ser visto como um procedimento de decisão  $\delta : \Omega \rightarrow \mathcal{A}$ , com espaço de ações  $\mathcal{A} = \Theta$ . Ou seja, um estimador para o parâmetro  $\theta$ , baseado na amostra aleatória  $\mathbf{X} = (X_1, \dots, X_n)$ , é uma função  $\delta(\mathbf{X})$  que especifica o valor estimado de  $\theta$  para cada possível valor de  $\mathbf{X}$ . Além disso, assumamos que para cada valor de  $\theta \in \Theta$  e para cada estimativa  $a \in \Theta$ , existe um número  $l(\theta, a)$  que mede a perda, ou custo, por utilizar a estimativa  $a$ , quando o valor “verdadeiro” do parâmetro é  $\theta$ . Assim, definimos uma função de perda  $l$ , como

$$l : \Theta \times \mathcal{A} \rightarrow \mathbb{R}_+$$

**Tabela 4:** Funções de perda e seus respectivos estimadores de Bayes.

Função de perda	$l(\theta, a)$	Estimador
Quadrática	$(\theta - a)^2$	$E[\theta   \mathbf{x}]$
Absoluta	$ \theta - a $	Mediana de $\pi(\theta   \mathbf{x})$
0 - 1	$\begin{cases} 0, & \text{se }  \theta - a  \leq b, \\ 1, & \text{se }  \theta - a  > b. \end{cases}$	Moda de $\pi(\theta   \mathbf{x})$

$$(\theta, a) \mapsto l(\theta, a)$$

Utilizaremos como critério de comparação entre estimadores, a perda esperada do procedimento  $\delta$  após observada a amostra  $\mathbf{x}$ , ou seja,

$$E[l(\theta, \delta(\mathbf{x})) | \mathbf{x}] = \int_{\Theta} l(\theta, \delta(\mathbf{x})) \pi(\theta | \mathbf{x}) d\theta. \quad (13)$$

Chamaremos de estimador de Bayes o procedimento de decisão que minimiza a equação (13). Assim, diferentes funções de perda ou distribuições priori, levam a diferentes estimadores de Bayes. A Tabela 4 apresenta a forma do estimador para algumas funções de perda.

### 2.1.1 Distribuição preditiva a posteriori

Antes de observamos uma amostra aleatória  $\mathbf{x}$ , a distribuição de uma observação  $x$  é dada por

$$f(x) = \int_{\Theta} \pi(\theta) f(x | \theta) d\theta \quad (14)$$

A equação (14) é chamada de distribuição marginal de  $x$ , ou distribuição preditiva a priori de  $x$  (ver Gelman et al. (2013)).

A distribuição preditiva a posteriori é a distribuição de dados não observados condicionada ao conjunto observado  $\mathbf{x}$ . Ou seja, após a amostra  $\mathbf{x}$  ser observada, podemos prever o valor de “futuras” observações,  $\mathbf{x}^{pred}$ . A distribuição de  $\mathbf{x}^{pred}$  é chamada de distribuição preditiva

a posteriori e é definida como

$$\begin{aligned} f(\mathbf{x}^{pred} | \mathbf{x}) &= \int_{\Theta} f(\mathbf{x}^{pred}, \theta | \mathbf{x}) d\theta = \int_{\Theta} f(\mathbf{x}^{pred} | \theta, \mathbf{x}) \pi(\theta | \mathbf{x}) d\theta \\ &= \int_{\Theta} f(\mathbf{x}^{pred} | \theta) \pi(\theta | \mathbf{x}) d\theta. \end{aligned} \tag{15}$$

Em análise bayesiana de dados, o uso da distribuição preditiva a posteriori é comumente relacionado a verificação de ajuste do modelo. Na Seção 2.5, introduziremos um possível método de seleção de modelos, fazendo uso implícito de  $f(\mathbf{x}^{pred} | \mathbf{x})$ .

## 2.2 Métodos MCMC

Como visto na seção anterior, se queremos fazer inferências sobre o parâmetro  $\theta$  de uma distribuição, ou seja, se queremos tirar conclusões sobre nosso parâmetro de interesse, devemos conhecer a nossa distribuição a posteriori. Pela equação (12), vemos que a função de verossimilhança é de grande importância para a construção e conhecimento de  $\pi(\theta | \mathbf{x})$ . Em modelos simples, podemos encontrar fórmulas analiticamente tratáveis para a verossimilhança, a fim de que a distribuição posterior de  $\theta$  seja de fácil manuseio. Em casos mais complexos, uma forma analítica para  $\pi(\theta | \mathbf{x})$  é inviável ou seu cálculo é muito caro computacionalmente. Para isso, uma abordagem muito difundida é a utilização de métodos de Monte Carlo via Cadeias de Markov (MCMC). Tais métodos consistem em simulações baseadas em Cadeias de Markov ergódicas, onde a distribuição estacionária do processo estocástico é a distribuição a posteriori de interesse. Dessa forma, o objetivo é obter uma amostra de  $\pi(\theta | \mathbf{x})$ .

Dentre as diversas abordagens MCMC, encontra-se o algoritmo Metropolis-Hasting (M-H), inicialmente desenvolvido por Metropolis et al. (1953) e posteriormente generalizado por Hastings (1970). Denotemos por  $g$ , a distribuição estacionária da cadeia de Markov. O algoritmo M-H inicia com a suposição de que, para cada passo  $i$  de nossa cadeia,  $i \in \mathbb{N}$ , seja possível simularmos um valor  $y$ , de uma determinada distribuição conhecida, segundo um método simples, por exemplo, de uma distribuição uniforme ou de uma distribuição normal. Seja  $q(z, y)$  a função de densidade dessa distribuição que pode ou não depender do estado anterior do processo,  $z$ . Ou seja, os candidatos  $y$  são gerados a partir de  $q(z, y)$ , e então a cadeia irá se movimentar do estado  $z$  para o estado  $y$ , ou irá permanecer em  $z$ .

Chib e Greenberg (1995) postulam que, dado que a cadeia estava no estado  $z$  no passo

anterior, ela deverá se movimentar para o candidato  $y$  com probabilidade  $p(z, y)$  dada por

$$p(z, y) = \begin{cases} \min \left\{ \frac{g(y)q(y, z)}{g(z)q(z, y)}, 1 \right\} & , \text{ se } g(z)q(z, y) > 0, \\ 1 & , \text{ caso contrário.} \end{cases} \quad (16)$$

Tais probabilidades são chamadas de probabilidades de movimento. Também em Chib e Greenberg (1995) foi mostrado que nos casos em que temos a função  $g$  definida em suporte limitado ou no mesmo suporte da distribuição alvo  $g$ , as probabilidades de movimento definidas pela equação (16) fazem com que a cadeia seja reversível e possua distribuição estacionária  $g$ .

Por fim, após simularmos a cadeia por vários passos, esperamos que as observações possuam aproximadamente distribuição  $g$ . No entanto, existem dois pontos que devem ser considerados. O primeiro é a dependência da cadeia quanto ao seu estado inicial. Uma possível abordagem é considerarmos um período de aquecimento (*burn-in*) do processo, digamos  $b$  passos, e apenas começarmos a amostragem após um determinado número de iterações.

O segundo tópico é a existência de correlação entre os passos da cadeia. Para ser contornado de forma satisfatória, amostramos, após o período de *burn-in*, valores que possuam um determinado espaçamento, digamos  $l$ , entre os passos em que foram gerados.

Em inferência bayesiana, o uso do algoritmo M-H está, geralmente, relacionado a estimar a distribuição a posteriori do parâmetro de interesse. Ou seja, estamos interessados em construir uma cadeia de Markov que possua  $\pi(\cdot | \mathbf{x})$  como sua distribuição estacionária. Para isso, basta utilizarmos as probabilidades de movimento como

$$p(z, y) = \begin{cases} \min \left\{ \frac{\pi(y|\mathbf{x})q(y, z)}{\pi(z|\mathbf{x})q(z, y)}, 1 \right\} & , \text{ se } \pi(z|\mathbf{x})q(z, y) > 0, \\ 1 & , \text{ caso contrário.} \end{cases} \quad (17)$$

Notemos que, pela equação (12), o denominador da distribuição a posteriori é constante em  $\theta$ . Assim, podemos calcular facilmente as probabilidades dadas na equação (17), como

$$p(z, y) = \begin{cases} \min \left\{ \frac{\pi(y)f(\mathbf{x}|y)q(y, z)}{\pi(z)f(\mathbf{x}|z)q(z, y)}, 1 \right\} & , \text{ se } \pi(z)f(\mathbf{x}|z)q(z, y) > 0, \\ 1 & , \text{ caso contrário.} \end{cases}$$

Em resumo, podemos construir o Algoritmo 1 para estimarmos a distribuição a posteriori de  $\theta$ .

---

**Algoritmo 1** : MCMC-MH para estimação da distribuição a posteriori de  $\theta$ .

---

Inicializar com um valor arbitrário  $\theta^{(0)}$

**Para**  $i \in \{0, 1, \dots, N\}$  **Faça**

Gerar candidata  $\theta^*$  a partir de  $q(\theta^i, \cdot)$  e  $u \sim Unif(0, 1)$

**Se**  $u \leq p(\theta^{(i)}, \theta^*)$ , **Então**

$$\theta^{(i+1)} = \theta^*$$

**Caso Contrário**

$$\theta^{(i+1)} = \theta^{(i)}$$

**Finalize Se**

**Finalize Para**

Desconsiderar os  $b$  primeiros valores obtidos e amostrar observações separadas por  $l$  passos;

Utilizando os valores amostrados, aproximar a distribuição a posteriori  $\pi(\theta | \mathbf{x})$ .

---

## 2.3 Aproximação bayesiana computacional

A metodologia chamada de aproximação bayesiana computacional (ABC) possui suas raízes nos anos 1980 com Rubin (1984), e ideias de Diggle e Gratton (1984). Tal procedimento contorna o cálculo exato da função de verossimilhança. A ideia básica de um algoritmo de rejeição ABC consiste em obtermos uma amostra da distribuição a posteriori de  $\theta$ . Para isso, devemos amostrar um conjunto de valores da distribuição a priori  $\pi(\theta)$ . Dado um valor amostrado do parâmetro,  $\theta^*$ , devemos simular um conjunto de dados  $\hat{\mathbf{x}}$  de  $P_{\theta^*} \in \mathbf{P}$ . Se o conjunto gerado  $\hat{\mathbf{x}}$  for muito diferente da amostra original  $\mathbf{x}$ , o parâmetro amostrado  $\theta^*$  é descartado. De maneira mais formal, dada uma distância  $\rho$  e uma tolerância  $\varepsilon > 0$ , aceitamos  $\theta^*$  se  $\rho(\hat{\mathbf{x}}, \mathbf{x}) \leq \varepsilon$ . Contudo, se a dimensão do conjunto de dados original for muito grande, a probabilidade de gerarmos um conjunto de dados  $\hat{\mathbf{x}}$  “próximo” a  $\mathbf{x}$  é muito pequena, o que leva a uma diminuição substancial da eficiência do algoritmo ABC. Uma possível abordagem, e que adotaremos neste texto, é considerar uma medida-resumo apropriada de  $\mathbf{x}$ ,  $S(\mathbf{x})$ . Assim, teremos que um valor  $\theta^*$  é aceito se  $\rho(S(\hat{\mathbf{x}}), S(\mathbf{x})) \leq \varepsilon$ , ou seja, é aceito se a medida-resumo escolhida de  $\hat{\mathbf{x}}$  está suficientemente próxima da de  $\mathbf{x}$ .

Dada uma amostra  $x_1, \dots, x_n$ , podemos resumir o procedimento ABC no Algoritmo 2.

Posteriormente, percebeu-se que ajustar a tolerância  $\varepsilon$  pode ser uma tarefa muito trabalhosa. Dessa forma, seguindo os passos de Nunes e Balding (2010), fixamos a quantidade de valores de  $\theta$  amostrados da distribuição a priori,  $n_{sim}$ , e aceitamos os  $N$  valores que produziram as estatísticas  $S(\hat{\mathbf{x}})$  mais próximas de  $S(\mathbf{x})$ . O Algoritmo 3 apresenta o algoritmo ABC segundo

---

**Algoritmo 2** : ABC para estimação da distribuição a posteriori de  $\theta$ .

---

Calcular a medida  $S(\mathbf{x})$  para os dados observados

**Enquanto** não forem aceitos  $N$  valores para  $\theta$  **Faça**

Amostrar  $\theta^*$  da distribuição a priori  $\pi(\theta)$  e simular uma amostra,  $\hat{\mathbf{x}}$ , segundo o modelo e a distribuição especificada por  $\theta^*$

Calcular a medida-resumo  $S(\hat{\mathbf{x}})$

**Se**  $\rho(S(\mathbf{x}), S(\hat{\mathbf{x}})) \leq \varepsilon$ , **Então**

    aceito o valor  $\theta^*$

**Caso Contrário**

$\theta^*$  é descartado

**Finalize Se**

**Finalize Enquanto**

Utilizando os valores aceitos, aproximar a distribuição a posteriori  $\pi(\theta | \mathbf{x})$ .

---

as modificações de Nunes e Balding (2010).

Em Joyce e Marjoram (2008), os autores discutem sobre a escolha da estatística  $S(\mathbf{x})$  a ser utilizada. Declaram que, idealmente,  $S(\mathbf{x})$  deve ser uma estatística suficiente para o parâmetro de interesse, contudo, isso nem sempre é alcançável na prática. Por isso, em Nunes e Balding (2010), foi apresentado o conceito de ABC de mínima entropia (ABCME). Tal variação do método consiste em escolher, dentro de um conjunto de estatísticas,  $\Omega_S$ , aquela que produzir a distribuição a posteriori  $\pi(\theta | \mathbf{x})$  com menor entropia. Como introduzido em Shannon (1948), a entropia de uma distribuição de probabilidade é uma medida de informação: alta entropia corresponde a pouca informação e vice-versa. Ou seja, estamos selecionando a distribuição a posteriori que carregue mais informação sobre o observado  $\mathbf{x}$ . Shannon (1948) define a entropia para uma variável aleatória discreta  $Y$ , como

$$H(Y) = - \sum_y Pr(Y = y) \log_2 Pr(Y = y).$$

No entanto, em nosso texto, todas as variáveis consideradas possuem distribuição contínua. Para contornar tal questão, Vasicek (1976) apresenta a generalização da entropia de Shannon para variáveis contínuas. Seja  $X$  uma variável aleatória contínua com suporte  $\mathbb{X}$  e densidade  $f(\cdot)$ .

---

**Algoritmo 3** : ABC para estimação da distribuição a posteriori de  $\theta$  (Nunes e Balding (2010)).

---

Calcular  $S(\mathbf{x})$  para os dados observados

**Para**  $i \in \{1, \dots, n_{Sim}\}$  **Faça**

Amostrar  $\theta^{(i)}$  da distribuição a priori  $\pi(\theta)$

Simular uma amostra,  $\hat{\mathbf{x}}^{(i)}$ , segundo o modelo e a distribuição especificada por  $\theta^{(i)}$

Calcular a medida-resumo  $S(\hat{\mathbf{x}}^{(i)})$

**Finalize Para**

Declare  $M$  como o conjunto dos  $N$  valores de  $\theta^{(i)}$ , cujas simulações minimizam  $\rho(S(\mathbf{x}), S(\hat{\mathbf{x}}^{(i)}))$

Utilizando os valores de  $M$ , aproximar a distribuição a posteriori  $\pi(\theta | \mathbf{x}^{(i)})$ .

---

A entropia contínua, ou entropia diferencial, é definida como

$$h(X) = - \int_{\mathcal{X}} f(x) \log f(x) dx. \quad (18)$$

Na literatura, diversos métodos para estimação de  $h(X)$  foram construídos. Aqui, para  $\theta_1, \dots, \theta_t$  amostra aleatória de  $\pi(\theta | \mathbf{x})$ , adotaremos o estimador da entropia baseado na técnica do k-ésimo vizinho mais próximo, definido em Nunes e Balding (2010) como

$$\hat{h}(\theta | \mathbf{x}) = \log \left[ \frac{\pi^{1/2}}{\Gamma(3/2)} \right] - \psi(k) + \log t + \frac{1}{t} \sum_{i=1}^t \log R_{i,k}, \quad (19)$$

onde  $R_{i,k}$  é a distância euclidiana de  $\theta_i$  até seu k-ésimo vizinho mais próximo na amostra da distribuição a posteriori, e  $\psi(k) = \Gamma'(k)/\Gamma(k)$  é a função digama. Em Singh et al. (2003), é demonstrado que o estimador  $\hat{h}(\theta | \mathbf{x})$  é não viesado, e aconselha o uso de  $k = 4$ , devido ao baixo erro quadrático médio associado.

Consideremos o problema de encontrar em  $\Theta$  as candidatas mais apropriadas a serem a densidade geradora da amostra observada  $\mathbf{x}$ . Nos exemplos do Capítulo 3, assumiremos que a distribuição “verdadeira”  $F(\cdot | \theta)$  pertence à classe  $\mathcal{VR}_\alpha$ ,  $\alpha > 0$ . Neste texto, iremos nos focar em tirar conclusões sobre o peso caudal das distribuições, ou seja, faremos inferência para parâmetros relacionados a ele.

## 2.4 Verificação do modelo

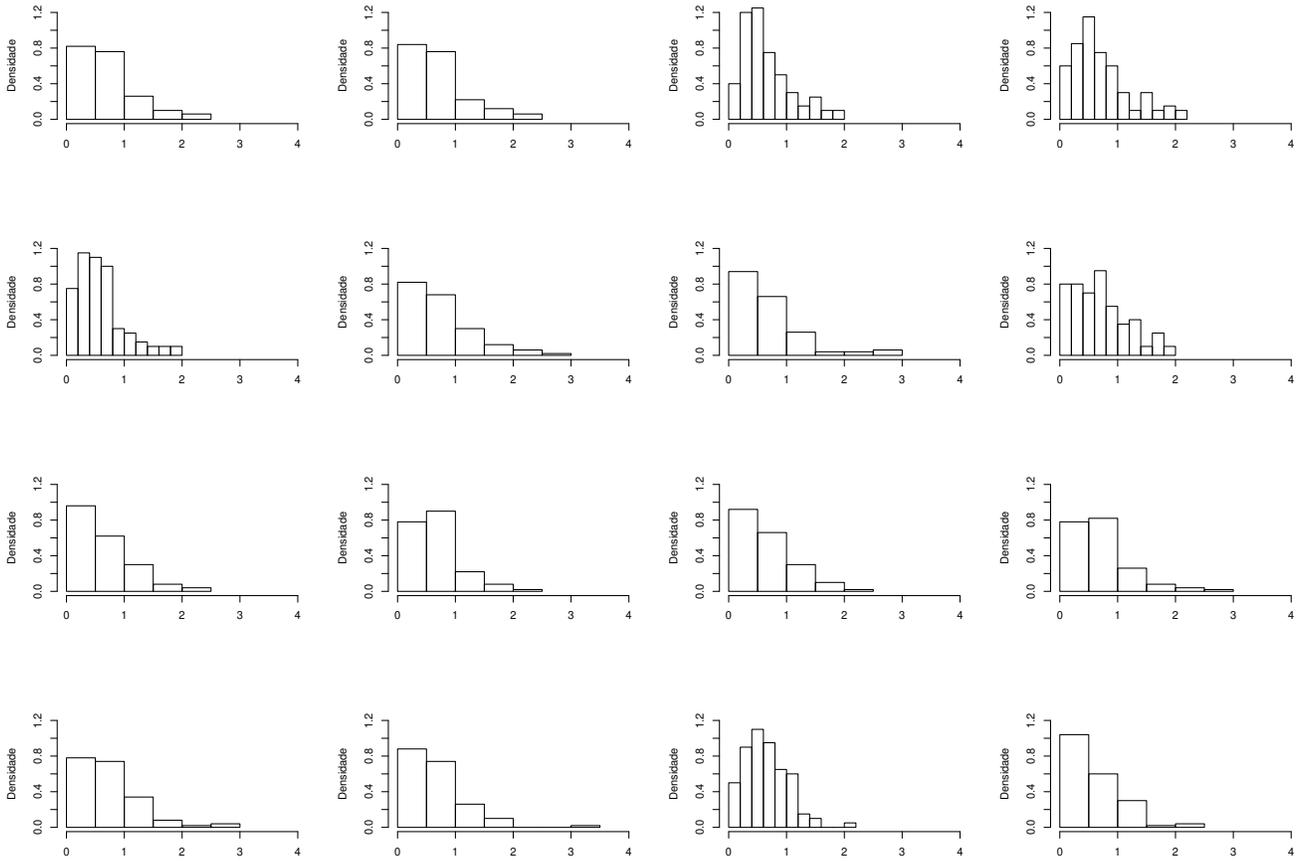
Após definirmos um modelo e obtermos a distribuição a posteriori (de forma analítica ou por simulações) do parâmetro de interesse, devemos verificar se o modelo ajusta-se bem aos dados e ao nosso conhecimento prévio. Ou seja, devemos averiguar se o ajuste obtido é adequado e plausível segundo a natureza do problema. Seguindo a linha de pensamento da seção anterior, é razoável a conjectura de que se um modelo está bem ajustado, então dados replicados, gerados nas condições determinadas pelo modelo, devem ser “semelhantes” aos observados originalmente. Dessa forma, consideremos  $\mathbf{x}$  a amostra observada e  $\theta$  o parâmetro de interesse. Definamos  $\mathbf{x}^{rep}$  como sendo os dados replicados que poderiam ser observados, ou, como apontado em Gelman et al. (2013), a amostra que veríamos amanhã se o experimento que produziu  $\mathbf{x}$  hoje fosse replicado com o mesmo modelo e mesmo valor de  $\theta$  que produziu os dados observados. Trabalharemos com a distribuição de  $\mathbf{x}^{rep}$  dado nosso atual nível de conhecimento, ou seja, utilizaremos a distribuição preditiva a posteriori apresentada na equação (15).

Uma técnica gráfica de verificação do modelo comum na literatura é a comparação do histograma da amostra observada  $\mathbf{x}$  com histogramas de réplicas  $\mathbf{x}^{rep}$  (ver Gelman et al. (2013)). A Figura 2 apresenta histogramas de amostras de tamanho 100, simuladas a partir de uma distribuição Gama(2, 3). Podemos notar diversas semelhanças nas amostras observadas, o que justificaria a proposição de que são provenientes de uma mesma distribuição. No entanto, para amostras de distribuições de cauda muito pesada, a análise anterior torna-se complicada. A Figura 3 apresenta histogramas de amostras de tamanho 100, simuladas a partir de uma distribuição Log-gama(0.5, 5). Notemos que as amostras observadas apresentam grande variabilidade, o que dificulta suas comparações. Ou seja, esta técnica gráfica de verificação tem seu uso comprometido quando utilizamos distribuições de cauda pesada em nosso modelo.

Uma alternativa para verificarmos a plausibilidade do modelo, é definirmos medidas de interesse que indiquem as discrepâncias entre o modelo e os dados observados. Gelman et al. (2013), apresenta o conceito de medida de teste, uma medida-resumo escalar,  $T(\mathbf{x}, \theta)$ , que leva em consideração a amostra  $\mathbf{x}$  e o valor do parâmetro  $\theta$  utilizado para produzi-la. Um ajuste ruim do modelo aos dados, com respeito a distribuição preditiva a posteriori, pode ser quantificado pelo p-valor preditivo a posteriori, definido como

$$p_B = Pr(T(\mathbf{x}^{rep}, \theta) \geq T(\mathbf{x}, \theta) | \mathbf{x}), \quad (20)$$

onde a probabilidade é calculada sobre a distribuição a posteriori de  $\theta$  e a distribuição preditiva a posteriori de  $\mathbf{x}^{rep}$ . Valores próximos de zero ou de um, indicam discrepâncias entre o modelo



**Figura 2:** Histogramas de amostras de tamanho 100 de uma distribuição  $Gama(2, 3)$ .

ajustado e a amostra observada, pois se o modelo está bem ajustado com relação a característica da distribuição refletida na medida  $T(\cdot, \theta)$ , então, após observada a amostra  $\mathbf{x}$ ,  $T(\mathbf{x}^{rep}, \theta)$  não deve ter tendência de ser maior ou menor que  $T(\mathbf{x}^{rep}, \theta)$ . Gelman et al. (2013) argumenta que a escolha de  $T(\mathbf{x}, \theta)$  deve ser feita com cautela. A medida de teste deve refletir a natureza do parâmetro  $\theta$ . No entanto,  $T(\mathbf{x}, \theta)$  não deve ser uma estatística suficiente para o modelo, pois na ausência de prioris informativas, o valor de  $p_B$  nos indicaria sempre que o ajuste está bem feito. Neste texto, como estamos estudando distribuições de cauda pesada, a medida de teste deverá refletir o peso caudal das mesmas, como exemplo, altos quantis amostrais ou estimativas pontuais de probabilidades relacionadas à cauda da distribuição.

Na prática, o valor numérico da equação (20) é aproximado utilizando-se simulações. O Algoritmo 4 apresenta uma fácil maneira de estimarmos  $p_B$ .

Para outros métodos de verificação do modelo, ver Gelman et al. (2013), Geisser e Eddy (1979), Gelfand (1996) e Laud e Ibrahim (1995).

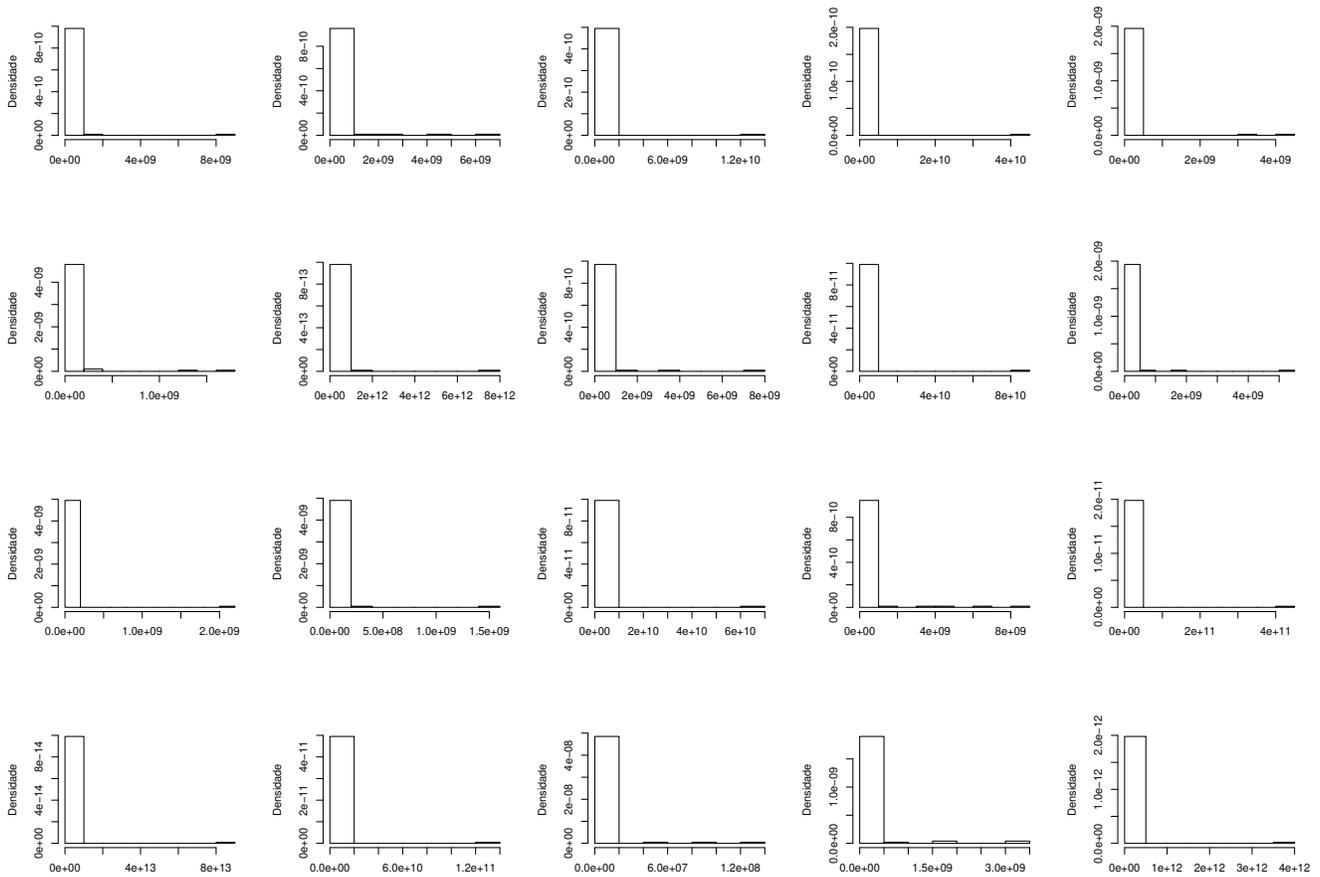


Figura 3: Histogramas de amostras de tamanho 100 de uma distribuição  $\text{Log} - \text{Gamma}(0.5, 5)$ .

## 2.5 Seleção de modelos

Em modelagem estatística, é muito comum que mais de um modelo probabilístico forneça um ajuste adequado aos dados em questão. Por isso, devemos nos atentar às mudanças na inferência posterior quando consideramos diferentes modelos. Diversas medidas de comparação de modelos, como Critério de Informação Bayesiano (BIC), Fator de Bayes, e Desvio Esperado (ver Gelman et al. (2013)), foram amplamente estudadas e aplicadas. No entanto, a grande maioria desses métodos necessitam do cálculo da função de verossimilhança em algum momento. Seguindo a motivação do algoritmo ABC, estamos interessados em critérios de seleção de modelos que não envolvam o cálculo de  $f(\mathbf{x} | \theta)$ .

Consideremos o problema de selecionar em um conjunto de modelos,  $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_d\}$ , o elemento que melhor se ajusta aos dados observados  $\mathbf{x}$ . Chamemos de  $\Psi = \{1, \dots, d\}$  o conjunto de índices dos modelos propostos. Queremos fazer inferência sobre  $\Psi$ . Utilizemos uma

---

**Algoritmo 4** : Estimaco de  $p_B$  para uma medida de teste  $T(\mathbf{x}, \theta)$ .

---

**Para**  $l \in \{1, 2, \dots, L\}$  **Faa**

Amostre  $\theta_l$  de  $\pi(\theta | \mathbf{x})$

Gere  $\mathbf{x}_l^{rep}$  segundo o modelo e o valor de  $\theta_l$

Calcule  $T(\mathbf{x}_l^{rep}, \theta_l)$  e  $T(\mathbf{x}, \theta_l)$

**Finalize Para**

Estime  $p_B$  como a proporo de vezes em que  $T(\mathbf{x}_l^{rep}, \theta_l) \geq T(\mathbf{x}, \theta_l)$ .

---

---

**Algoritmo 5** : ABC-MC.

---

Calcule  $S(\mathbf{x})$

**Para**  $n \in \{1, \dots, n_{Sim}\}$  **Faa**

Gere  $m^{(n)}$  da priori  $p(\Psi = m)$

Gere  $\theta_{m^{(n)}}$  da priori  $\pi_{m^{(n)}}(\theta_{m^{(n)}})$

Gere  $\mathbf{z}^{(n)}$  a partir do modelo  $m^{(n)}$  e de  $\theta_{m^{(n)}}$

Calcule  $S(\mathbf{z}^{(n)})$

**Finalize Para**

Declare  $M$  como o conjunto dos  ndices das  $N$  simulaoes que minimizam  $\rho(S(\mathbf{x}), S(\mathbf{z}^{(n)}))$

**Para**  $k \in \{1, \dots, d\}$  **Faa**

$$\hat{p}(\Psi = k | \mathbf{x}) = \frac{1}{N} \sum_{n \in M} \mathbb{1}_{\{m^{(n)}=k\}}$$

**Finalize Para**

Selecione o modelo  $k$  que maximize  $\hat{p}(\Psi = k | \mathbf{x})$ .

---

distribuio a priori para  $\Psi$ ,  $p(\Psi = m)$ , al m de uma distribuio a priori para o par metro de interesse, condicionada ao valor do modelo de  ndice  $m$ ,  $\pi_m(\theta_m)$ . Toni et al. (2009) propoem que o modelo que possuir a maior probabilidade a posteriori, ou seja, o valor de  $m$  que maximiza  $p(\Psi = m | \mathbf{x})$ , deve ser o  ndice do modelo selecionado como o mais apropriado para se explicar as observaoes  $\mathbf{x}$ . O Algoritmo 5 apresenta o m todo proposto por Toni et al. (2009) (chamado de ABC-MC), adaptado ao m todo introduzido por Nunes e Balding (2010). Sejam  $\rho$  uma dist ncia e  $S(\cdot)$  uma estat stica. Dessa forma, o Algoritmo 5   constru do como um poss vel m todo de seleo de modelos livre do uso da forma anal tica da funo de verossimilhana.

Podemos notar, no algoritmo supracitado, que a seleo do modelo mais apropriado ocorre sem levarmos em considerao a distribuio a posteriori de cada modelo. Neste texto, modificaremos o algoritmo ABC-MC incorporando o uso impl cito da distribuio preditiva a

---

**Algoritmo 6** : ABC-PPMC.

---

Calcule  $S(\mathbf{x})$

**Para**  $n \in \{1, \dots, n_{Sim}\}$  **Faça**

Gere  $m^{(n)}$  da priori  $p(\Psi = m)$

Gere  $\theta_{m^{(n)}}$  da posteriori  $\pi_{m^{(n)}}(\theta_{m^{(n)}} | \mathbf{x})$

Gere  $\mathbf{z}^{(n)}$  a partir do modelo  $m^{(n)}$  e de  $\theta_{m^{(n)}}$

Calcule  $S(\mathbf{z}^{(n)})$

**Finalize Para**

Declare  $M$  como o conjunto dos índices das  $N$  simulações que minimizam  $\rho(S(\mathbf{x}), S(\mathbf{z}^{(n)}))$

**Para**  $k \in \{1, \dots, d\}$  **Faça**

$$\hat{p}(\Psi = k | \mathbf{x}) = \frac{1}{N} \sum_{n \in M} \mathbb{1}_{\{m^{(n)}=k\}}$$

**Finalize Para**

Selecione o modelo  $k$  que maximize  $\hat{p}(\Psi = k | \mathbf{x})$ .

---

posteriori. Consideremos, além das condições do algoritmo anterior, a distribuição a posteriori do parâmetro  $\theta$ , obtida via ABCME para cada um dos  $d$  modelos, ou seja, para o modelo  $m$ ,  $m \in \Psi$ , dispomos de  $\pi_m(\theta_m | \mathbf{x})$ . Construimos agora o Algoritmo 6 e o definimos como o método ABC para seleção de modelos via distribuição preditiva a posteriori (ABC-PPMC), outra possível abordagem para seleção de modelos livre do uso da forma analítica da função de verossimilhança.

É importante ressaltar que devemos escolher uma estatística  $S(\cdot)$  que seja pouco sensível ao valor do parâmetro de interesse  $\theta$ , pois queremos que o modelo selecionado possa refletir toda a informação contida nos dados que esteja além da natureza de  $\theta$ . Por exemplo, neste texto, como estamos nos focando em distribuições de cauda pesada, e como o parâmetro  $\theta$  está de alguma forma refletindo o peso caudal, a escolha de  $S(\cdot)$  como um baixo quantil amostral, digamos 10%, pode ser razoável. O próximo capítulo apresentará um comparativo entre os métodos ABC-MC e ABC-PPMC.

# Capítulo 3

## Aplicações

### 3.1 Estimação

Nesta seção, iremos analisar o comportamento amostral de estimadores das medidas de evidência de cauda pesada apresentadas no Capítulo 1, através de simulações. Utilizaremos tais estimadores como estatísticas auxiliares no método ABC, e também como possíveis ferramentas para verificação do ajuste de modelos.

Por se tratar de uma probabilidade,  $Ra_{n,2}(X)$  (ver equação (4), página 11) pode ser estimado via *bootstrap*, dada uma amostra observada  $x_1, \dots, x_m$ . Neste texto, utilizaremos 20000 reamostragens para a estimação de tal quantidade. Denotemos tal estimador por  $\hat{R}a_{n,2}(X)$ . As Figuras A.1, A.2 e A.3 apresentam o comportamento amostral, para  $m = 500$ , de  $\hat{R}a_{2,2}(X)$ ,  $\hat{R}a_{15,2}(X)$  e  $\hat{R}a_{30,2}(X)$ , respectivamente, para diferentes distribuições. As Tabelas B.1 e B.2 apresentam o viés e o erro quadrático médio (EQM) observados para as distribuições consideradas nas Figuras A.1, A.2 e A.3. Pelos resultados obtidos, podemos notar que  $\hat{R}a_{n,2}(X)$  tende a subestimar o valor de  $Ra_{n,2}(X)$ . Além disso, quanto mais pesada a cauda da distribuição de  $X$ , maior o viés e menor o EQM observados. Outra peculiaridade ocorre quando comparamos os estimadores para diferentes tamanhos de subamostras, ou seja, comparamos  $\hat{R}a_{n_1,2}(X)$  com  $\hat{R}a_{n_2,2}(X)$ ,  $n_1 < n_2$ : o primeiro possui viés e EQM menores que o segundo.

Por também se tratar de uma probabilidade, podemos estimar  $Ob(X, n)$  (ver Definição 27, página 15) a partir de *bootstrap*, dada uma amostra observada  $x_1, \dots, x_m$ . Novamente utilizaremos 20000 reamostragens para a estimação de tal quantidade. Denotemos tal estimador por  $\hat{O}b(X, n)$ . As Figuras A.4, A.5 e A.6 apresentam o comportamento amostral, para  $m = 500$ , de  $\hat{O}b(X, 4)$ ,  $\hat{O}b(X, 15)$  e  $\hat{O}b(X, 30)$ , respectivamente, para diferentes distribuições. As Tabelas B.3 e B.4 apresentam o viés e o erro quadrático médio observados para as distribuições consideradas nas

Figuras A.4, A.5 e A.6. Pelos resultados obtidos, podemos notar que  $\hat{Ob}(X, n)$  se comporta de maneira semelhante a  $\hat{Ra}_{n,2}(X)$ , ou seja, tende a subestimar o valor de  $Ob(X, n)$ . Além disso, quanto mais pesada a cauda da distribuição de  $X$ , maior o viés e menor o EQM observados. A mesma peculiaridade anterior ocorre quando comparamos os estimadores  $\hat{Ob}(X, n_1)$  com  $\hat{Ob}(X, n_2)$ ,  $n_1 < n_2$ : o primeiro possui viés e EQM menores que o segundo.

O Exemplo 32 e a Proposição 34 (ver páginas 20 e 21) nos mostram que a função de excesso médio (ver Definição 29, página 19) pode nos trazer muita informação sobre a distribuição da variável  $X$ . Dessa forma, temos a necessidade de estimar a função  $e_X(x)$ . Dada uma amostra aleatória  $X_1, \dots, X_n$  de uma distribuição  $F$ , definamos o estimador da função de excesso médio,  $\hat{e}_X(x)$ , como

$$\hat{e}_X(x) = \frac{\sum_{i=1}^n X_i \mathbb{1}_{\{X_i > x\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > x\}}} - x \quad . \quad (21)$$

A Figura A.7 apresenta as FEM estimadas para algumas distribuições, considerando-se apenas os quantis amostrais de 65 a 90%. É interessante notar que para distribuições de cauda mais leve, o comportamento assintótico de  $e_X(x)$  pode não ser captado pelo estimador  $\hat{e}_X(x)$ .

Pela Proposição 34, temos que  $e_X(x)$  apresenta comportamento linear em  $x$ , quando  $x \rightarrow \infty$ , para distribuições de variação regular com índice de variação maior que um. Denotemos por  $\phi$  o coeficiente angular de  $e_X(x)$ , quando  $x \rightarrow \infty$ . Definimos um estimador para  $\phi$ ,  $\hat{\phi}$ , como sendo o coeficiente angular da reta de mínimos quadrados do estimador da função de excesso médio,  $\hat{e}_X(x)$ , restrito aos valores entre os quantis amostrais de 65 e 90%. Não consideramos valores acima do quantil de 90%, pois o pequeno número de observações nessa região pode afetar a qualidade da estimativa obtida. Dessa forma, estamos tentando reter o máximo de informação sobre o comportamento assintótico da distribuição contida na amostra, enquanto mantemos a qualidade da estimativa observada. Novamente pela Proposição 34, podemos propor um estimador para o índice de variação regular  $\alpha$ , da forma  $\hat{\alpha}_{FEM} = 1 + 1/\hat{\phi}$ . Como dito anteriormente,  $\hat{\alpha}_{FEM}$  pode nos fornecer estimativas ruins sobre o parâmetro  $\alpha$ , no caso de caudas menos pesadas. A Figura A.8 apresenta o comportamento amostral de  $\hat{\alpha}_{FEM}$  para a distribuição Log-gama com diferentes índices de variação regular. Podemos notar que quanto mais leve é a cauda da distribuição, maior a dispersão das estimativas em torno do verdadeiro valor do parâmetro. E além disso, para caudas muito leves, podemos obter estimativas que fogem do espaço paramétrico. Nas próximas seções, ilustraremos o uso da FEM no processo de inferência para distribuições de cauda pesada.

## 3.2 Exemplos simulados

Nesta seção, iremos comparar, através de alguns exemplos simulados, os resultados obtidos por inferência bayesiana tradicional e por métodos ABC. Inicialmente, consideraremos grandes amostras, uma vez que utilizaremos estimativas de probabilidades, como as citadas na seção anterior. Por esse mesmo motivo, iremos realizar o procedimento ABCME em duas etapas. Na primeira, iremos utilizar 20% dos valores simulados do parâmetro para selecionar a estatística que nos entrega a distribuição a posteriori de mínima entropia. Na segunda, continuamos o procedimento ABC, calculando apenas o valor da medida-resumo selecionada no passo anterior. Decidimos utilizar o procedimento em duas etapas, uma vez que estimativas de probabilidades são muito custosas computacionalmente.

**Exemplo 36:** Consideremos  $X_1, \dots, X_n$  uma amostra aleatória proveniente de uma distribuição Log-Gama(3, 2). Pelo apresentado na Tabela 2 (ver página 10), temos que a distribuição mencionada pertence à classe  $\mathcal{VR}_3$ . Assumamos que temos informação completa sobre o parâmetro de taxa da distribuição, ou seja, sabemos que a amostra provém de uma distribuição Log-Gama( $\alpha$ , 2),  $\alpha > 0$ . Para esse modelo, podemos escrever a função de verossimilhança para o parâmetro  $\alpha$  como

$$\begin{aligned} f(\mathbf{x} | \alpha) &= \prod_{i=1}^n f(x_i | \alpha) = \prod_{i=1}^n \frac{\alpha^2}{\Gamma(2)} [\log x_i] x_i^{-\alpha-1} \\ &= \frac{\alpha^{2n}}{[\Gamma(2)]^n} \left[ \prod_{i=1}^n \log x_i \right] \left[ \prod_{i=1}^n x_i \right]^{-\alpha-1} \\ &\propto \alpha^{2n} \left[ \prod_{i=1}^n x_i \right]^{-\alpha-1}. \end{aligned} \quad (22)$$

Pela equação (22), temos que a distribuição a priori conjugada para amostragem Log-Gama( $\alpha$ , 2) é a distribuição  $G(a, b)$  que possui densidade da forma

$$\pi(\alpha) \propto \alpha^a b^{-\alpha-1}, \quad \text{para } a > 0, b > 1. \quad (23)$$

Assim, dada a observação  $\mathbf{x} = (x_1, \dots, x_n)$ , temos que a densidade a posteriori de  $\alpha$  é

$$\pi(\alpha | \mathbf{x}) \propto \alpha^{2n} \left[ \prod_{i=1}^n x_i \right]^{-\alpha-1} \pi(\alpha) \propto \alpha^{2n+a} \left[ b \prod_{i=1}^n x_i \right]^{-\alpha-1}, \quad (24)$$

ou seja,  $\alpha|\mathbf{x} \sim G(2n + a, b \prod_{i=1}^n x_i)$ . Neste exemplo, assumiremos  $a = 1$  e  $b = 2$  e consideraremos uma amostra de tamanho 1000.

Para aplicarmos a técnica de ABCME (ver Algoritmo 3, página 30) para aproximarmos  $\pi(\alpha|\mathbf{x})$  devemos construir um conjunto de medidas-resumo  $\Omega_S$  apropriado. A Tabela B.5 apresenta as estatísticas escolhidas para  $\Omega_S$ . Para este exemplo, amostramos  $3 \times 10^6$  valores da distribuição a priori ( $n_{sim} = 3 \times 10^6$ ) e selecionamos os  $10^4$  valores de  $\alpha$  que produziram menor distância euclidiana entre cada uma das medidas-resumo observadas e simuladas ( $N = 10^4$ ). A Tabela B.5 também apresenta os valores observados das entropias das estimativas da distribuição a posteriori definidas para cada elemento de  $\Omega_S$ . Pelo apresentado, vemos que a estatística que define a densidade de mínima entropia é a média amostral,  $\frac{1}{1000} \sum_{i=1}^{1000} X_i$ .

Consideremos também uma aproximação de  $\pi(\alpha|\mathbf{x})$  via MCMC pelo algoritmo M-H (ver Algoritmo 1, página 28). Desejamos obter uma amostra da distribuição a posteriori de tamanho  $10^4$ . Pelo visto na Seção 2.2, devemos definir a densidade de transição  $q(z, y)$ . Assim, definamos

$$q(z, y) = ze^{-zy} \mathbb{1}_{(y>0)}$$

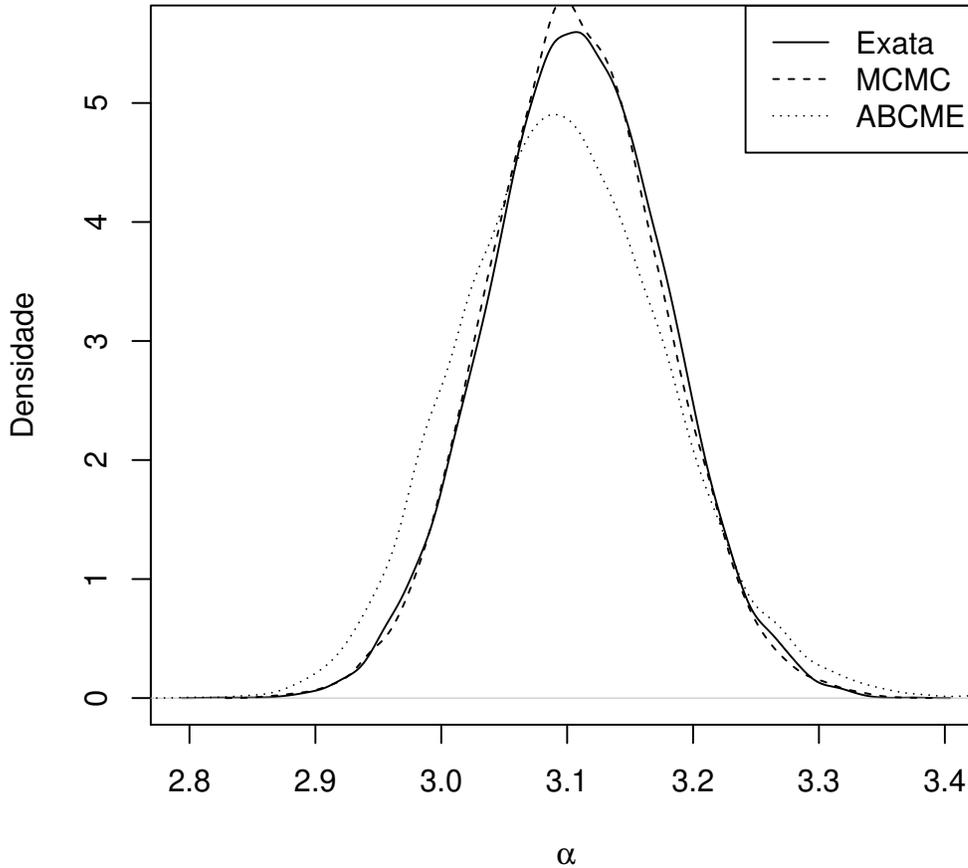
Consideremos um burn-in de  $3 \times 10^5$  passos e um espaçamento de amostragem de 200. A Figura A.9 apresenta os gráficos de autocorrelação (ACF) e de traços da amostra MCMC. Podemos notar que não há fortes indícios de que a cadeia não tenha atingido a estacionariedade, nem que exista correlação entre os valores amostrados.

A Figura 4 apresenta a densidade posterior exata, a aproximação MCMC e outra via ABCME.

Em Bishop (2006) é apresentada a divergência de Kullback–Leibler (KL) como medida de discrepância entre distribuições de probabilidade contínuas. Mais especificamente, dadas duas distribuições de probabilidade contínuas,  $P$  e  $Q$ , e suas respectivas densidades,  $p$  e  $q$ , medimos o quanto a distribuição  $Q$  está distante da distribuição  $P$ , calculando

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx.$$

Em Boltz et al. (2007) discute-se a estimação de  $D_{KL}(P||Q)$  via método do  $k$ -ésimo vizinho mais próximo. Neste exemplo, utilizaremos o procedimento levando em conta os 10 vizinhos mais próximos. A estimativa da divergência KL obtida ao usar o método MCMC para aproximar  $\pi(\theta|\mathbf{x})$  é de 0.0043, e de 0.0442 quando utilizamos o método ABCME. Logo, a distribuição aproximada pelo



**Figura 4:** Gráfico da densidade a posteriori exata e suas aproximações para o Exemplo 36.

*método MCMC apresenta menos perda de informação do que a aproximada pelo método ABCME.*

*Definamos agora, três estimadores pontuais para  $\alpha$ : um baseado na distribuição a posteriori exata,  $\hat{\alpha}_E$ ; um baseado na amostra obtida pelo método ABCME,  $\hat{\alpha}_{ABCME}$ ; e outro baseado na aproximação MCMC,  $\hat{\alpha}_{MCMC}$ . Consideremos o estimador de Bayes para perda quadrática, ou seja,  $\hat{\alpha}_E = E[\alpha | \mathbf{x}]$ . Para  $\hat{\alpha}_{ABCME}$  e  $\hat{\alpha}_{MCMC}$ , consideramos as médias amostrais dos valores obtidos para  $\alpha$ . Nesta simulação foram obtidas as seguintes estimativas:*

$$\begin{aligned}\hat{\alpha}_E &= 3.1079 \\ \hat{\alpha}_{ABCME} &= 3.0956 \\ \hat{\alpha}_{MCMC} &= 3.1089\end{aligned}$$

**Tabela 5:** Valores das estimativas observadas para o índice de variação,  $\alpha$ , para cem repetições do Exemplo 36.

Estimador	Média observada	EQM estimado
$\hat{\alpha}_E$	2.9996	0.0005
$\hat{\alpha}_{ABCME}$	2.9810	0.0061
$\hat{\alpha}_{MCMC}$	3.0000	0.0018

Repetimos o processo de simulação cem vezes, registrando em cada repetição os valores estimados de  $\alpha$ . A Tabela 5 apresenta os resultados.

Este exemplo foi construído com o intuito de ilustrar o procedimento ABCME para um caso simples, e verificar a consistência dos resultados obtidos. Pela Tabela 5, reparamos que o EQM estimado para  $\hat{\alpha}_{ABCME}$  é muito mais elevado do que para  $\hat{\alpha}_{MCMC}$ . Além disso, como dito anteriormente, a distribuição a posteriori estimada via MCMC apresenta menos perda de informação do que a estimada via ABCME. Tais resultados a favor do método MCMC já eram esperados, uma vez que tal procedimento conta com o cálculo direto da função de verossimilhança de  $\alpha$ , que é uma estatística suficiente para o parâmetro.

**Exemplo 37:** Consideremos  $X_1, \dots, X_{1000}$  uma amostra aleatória proveniente de uma distribuição Log – Gama(0.5, 2). Assim, temos que a distribuição mencionada pertence à classe  $\mathcal{VR}_{0.5}$ . Assumamos que temos informação completa sobre o parâmetro de taxa da distribuição, ou seja, sabemos que a amostra provém de uma distribuição Log – Gama( $\alpha$ , 2),  $\alpha > 0$ . Pelo apresentado no exemplo anterior, a distribuição a priori conjugada para amostragem Log-gama é a distribuição  $G(a, b)$  que possui densidade apresentada na equação (23). Consideremos, neste exemplo,  $a = 1$  e  $b = 10$ .

Seguindo o mesmo procedimento do exemplo anterior, utilizamos o método ABCME. A Tabela B.6 apresenta os valores observados das entropias das estimativas da distribuição a posteriori definidas para cada elemento de  $\Omega_S$ .

Podemos notar pelas Tabelas B.5 e B.6 que houve uma mudança em qual estatística produz a distribuição a posteriori de mínima entropia. Neste exemplo, tal estatística é  $\frac{1}{1000} \sum_{i=1}^{1000} \log X_i$ . Essa mudança é natural, uma vez que a distribuição Log – Gama(0.5, 2)  $\in \mathcal{VR}_{0.5}$  e, por consequência,  $E(X) = \infty$ . Notemos também que estatísticas referentes a cauda da distribuição, como  $\hat{R}_{a,2}(X)$ , apresentam-se mais expressivas nesta simulação. Isso se deve ao fato de que quanto

mais pesada a cauda em questão, mais facilmente o comportamento assintótico da distribuição é captado em uma amostra.

**Exemplo 38:** Consideremos  $X_1, \dots, X_n$  uma amostra aleatória proveniente de uma distribuição  $F$ , tal que

$$F(x) = 0.7F_1(x) + 0.3F_2(x), \quad (25)$$

em que  $F_1$  denota a distribuição Log – Gama(4, 2) e  $F_2$  a distribuição Log – Gama(6, 3). Assumamos que temos a informação completa sobre os parâmetros de taxas das distribuições e o peso da mistura. Assumamos também que temos conhecimento de que a distribuição possui pelo menos primeiro momento finito. Assim, temos que

$$F(x | \alpha_1, \alpha_2) = 0.7F_1(x | \alpha_1) + 0.3F_2(x | \alpha_2), \quad (26)$$

em que  $F_1(\cdot | \alpha_1)$  denota a distribuição Log–Gama( $\alpha_1$ , 2) e  $F_2(\cdot | \alpha_2)$  a distribuição Log–Gama( $\alpha_2$ , 3). Notemos que pela Proposição 13 e pelo Corolário 16 (ver página 9), temos que  $F(\cdot | \alpha_1, \alpha_2) \in \mathcal{VR}_\alpha$ ,  $\alpha = \min\{\alpha_1, \alpha_2\}$ . Suponha que nosso interesse é fazer inferência sobre o peso da cauda de  $F(\cdot | \alpha_1, \alpha_2)$ , ou seja, queremos obter informação sobre  $\alpha$ . Em casos como este, o cálculo exato da distribuição a posteriori  $\pi(\alpha | \mathbf{x})$  é muito custoso analiticamente, uma vez que a verossimilhança  $f(\mathbf{x} | \alpha)$  não possui uma forma simples. Assim, técnicas de aproximação de  $\pi(\alpha | \mathbf{x})$  podem ser convenientes.

Consideremos, a princípio, uma abordagem MCMC. Desejamos obter uma amostra da distribuição a posteriori de tamanho  $10^4$ . Pelo visto na Seção 2.2, devemos definir a densidade de transição  $q(\mathbf{z}, \mathbf{y}) = q((z_1, z_2), (y_1, y_2))$ . Assim, definamos  $q(\mathbf{z}, \mathbf{y}) = q((z_1, z_2), (y_1, y_2)) = q(z_1, y_1)q(z_2, y_2)$ , em que, para  $i = 1, 2$ ,

- Se  $1 \leq z_i \leq 2$ ,

$$q(z_i, y_i) = \left(1 - \frac{z_i}{2}\right) \frac{1}{z_i} \mathbb{1}_{\{y_i=1\}} + \frac{1}{z_i} \mathbb{1}_{(1, \frac{3z_i}{2})}(y_i);$$

- Se  $z_i > 2$ ,

$$q(z_i, y_i) = \frac{1}{z_i} \mathbb{1}_{(\frac{z_i}{2}, \frac{3z_i}{2})}(y_i).$$

Dessa forma, temos garantido que todos os possíveis valores de  $\alpha_1$  e  $\alpha_2$  serão maiores que 1. Consideremos também que, a priori,  $\alpha_1$  e  $\alpha_2$  são independentes e possuem distribuição Gama(6,2) deslocada em uma unidade para a direita. Nesta simulação, utilizamos burn-in de tamanho  $3 \times 10^5$ , amostramos valores com espaçamento de 200 passos, utilizamos os valores iniciais  $\alpha_1^{(0)} = \alpha_2^{(0)} = 5$ , e em cada repetição registramos  $\alpha^{(i)} = \min\{\alpha_1^{(i)}, \alpha_2^{(i)}\}$ .

A Figura A.10 apresenta histogramas das distribuições marginais amostrais de  $\alpha_1$  e  $\alpha_2$ , e a Figura A.11 apresenta a distribuição conjunta amostral dos mesmos. Notemos a existência de uma correlação negativa entre os parâmetros, o que é razoável que aconteça como um “balançamento” para o peso caudal da mistura. A Figura A.12 apresenta a amostra obtida de  $\pi(\alpha | \mathbf{x})$ . Definimos a estimativa do peso caudal obtida por este procedimento,  $\hat{\alpha}_{MCMC}$ , como a média da amostra da distribuição a posteriori obtida para  $\alpha$ . Neste caso, observamos  $\hat{\alpha}_{MCMC} = 4.3559$ .

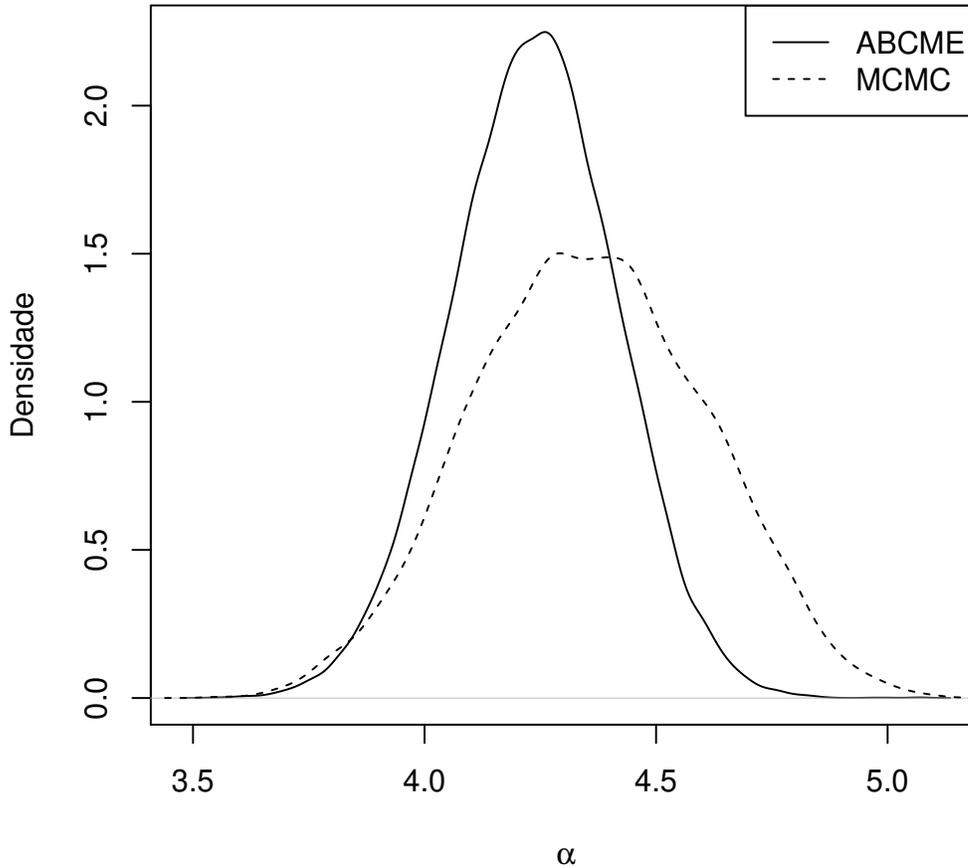
Consideremos, agora, uma possível abordagem ABC. Assumamos, a priori, que  $\alpha \sim \text{Gama}(6,2)$  deslocada em uma unidade para a direita. Desejamos obter uma amostra da distribuição a posteriori de tamanho  $10^4$ . Pelo visto na Seção 2.3, após amostrarmos  $\alpha^*$  de  $\pi(\cdot)$ , devemos gerar uma amostra aleatória,  $\hat{\mathbf{x}}$ , a partir da distribuição definida por ele. Para isso, realizamos os seguintes passos:

1. Com probabilidade  $\frac{1}{2}$ , sortear se  $F_1$  ou  $F_2$  terá índice de variação igual a  $\alpha^*$ ;
2. A distribuição não escolhida terá índice de variação igual a  $\alpha^* + \varepsilon$ , com  $\varepsilon \sim \text{Gama}(3,2)$ ;
3. Gerar uma amostra  $\hat{\mathbf{x}}$  a partir do modelo definido pelos passos 1 e 2.

Em sequência, devemos definir um conjunto de medidas-resumo,  $\Omega_S$ , conveniente. Em adição às medidas-resumo consideradas na Tabela B.5, incluiremos o coeficiente angular da função de excesso médio estimada,  $\hat{\phi}$ , apresentado na Seção 3.1, como elemento de  $\Omega_S$ .

Para este exemplo, utilizamos  $n_{sim} = 10^6$  e  $N = 10^4$ . A Tabela B.7 apresenta os valores observados das entropias das estimativas da distribuição a posteriori definidas para cada elemento de  $\Omega_S$ . Pelo apresentado, vemos que a estatística que define a densidade de mínima entropia é a média amostral. Definimos o estimador  $\hat{\alpha}_{ABCME}$  como sendo a média da distribuição amostral obtida pelo método ABCME. Neste caso, observamos  $\hat{\alpha}_{ABCME} = 4.2357$ .

A Figura 5 apresenta as densidades a posteriori estimadas para  $\alpha$  para as abordagens ABCME e MCMC. É possível notar que, além de uma ajuste de centralidade, a abordagem ABC proporcionou uma densidade posterior com menor variabilidade do que a obtida pela abordagem MCMC.



**Figura 5:** Gráficos das densidades a posteriori estimadas para o Exemplo 38.

**Exemplo 39:** Consideremos uma amostra aleatória  $X_1, \dots, X_{900}$  proveniente de uma distribuição  $\text{Log} - \text{Gama}(2, 4)$ . Suponhamos que temos interesse em fazer inferência sobre o peso caudal da distribuição. Primeiramente, queremos obter indícios de que a modelagem por uma distribuição de cauda pesada é razoável para este caso. A Tabela B.8 e a Figura A.13 apresentam algumas medidas de evidência de cauda pesada. Pelo apresentado na Tabela B.8, a suposição da distribuição em questão provir de um elemento da família  $\mathcal{CP}$  parece razoável, pois as medidas indicam que a distribuição da amostra possui cauda mais pesada que uma distribuição exponencial. E pelo apresentado na Figura A.13, um modelo de variação regular parece ser mais razoável que um modelo proveniente da família Weibull. Desejamos selecionar, em um conjunto de modelos  $\mathbf{P}$ , o elemento mais adequado para explicar o comportamento da amostra observada. A Tabela B.9

apresenta a descrição de  $\mathbf{P}$ . Consideremos a seguinte distribuição a priori para o conjunto de índices  $\Psi$ :

$$p(\Psi = m) = \begin{cases} \frac{1}{24}, & \text{se } m \in \{1, \dots, 8\}, \\ \frac{1}{12}, & \text{se } m \in \{9, \dots, 16\}. \end{cases} \quad (27)$$

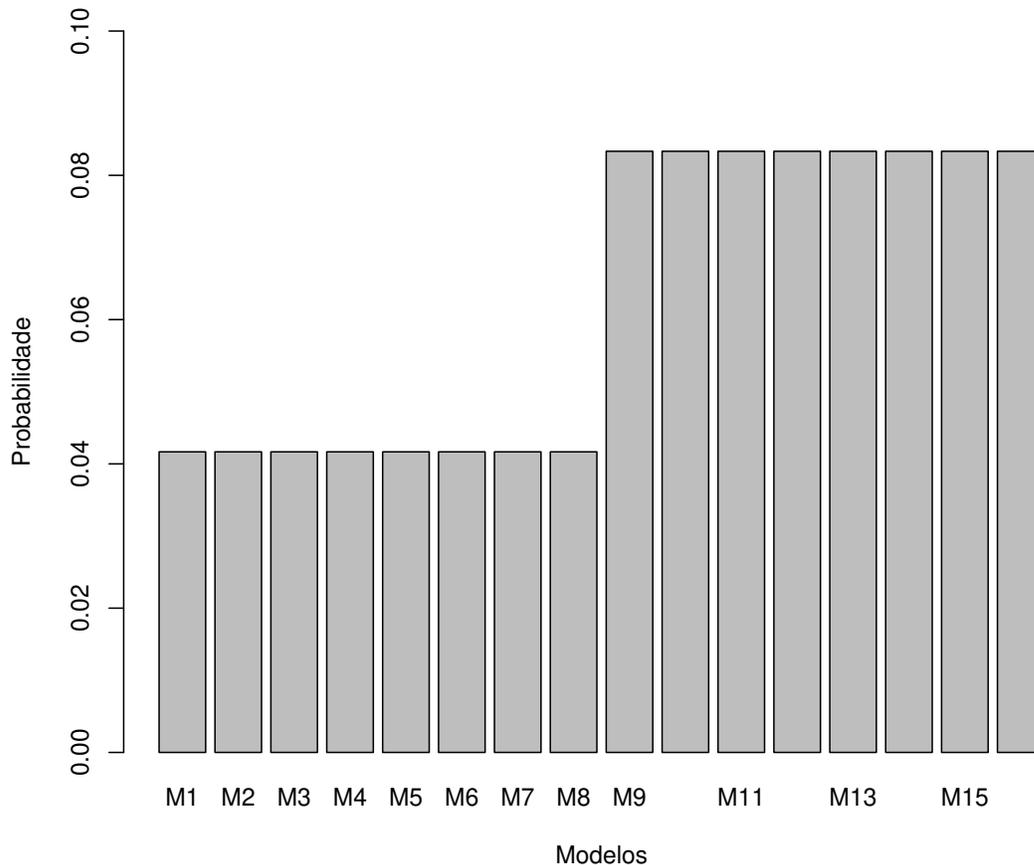
As Figuras 6, 7 e 8 apresentam graficamente as distribuições  $p(\Psi = m)$ ,  $p(\Psi = m | \mathbf{x})$  pelo método ABC-MC, e  $p(\Psi = m | \mathbf{x})$  pelo método ABC-PPMC, respectivamente. Notemos que, pelo primeiro método, o modelo escolhido é  $M_{11}$ , que, apesar de não ser o modelo original, está bem “próximo” a ele. Contudo, para o segundo método, o modelo escolhido é  $M_{12}$ , que é exatamente o modelo original.

Em sequência, realizamos um exercício de simulação para testar a acurácia dos dois métodos, ou seja, queremos conhecer a proporção de vezes em que cada abordagem seleciona o modelo correto. Para isso, selecionamos ao acaso um dos elementos apresentados na Tabela B.9 e estabelecemos uma distribuição a priori uniforme para os modelos. Calculamos  $p(\Psi = m | \mathbf{x})$  para os métodos ABC-MC e ABC-PPMC e selecionamos o modelo cujo índice maximize  $p(\Psi = m | \mathbf{x})$ . Repetimos o procedimento cem vezes. Para o método ABC-MC, o modelo original foi selecionado em 42% das vezes. No entanto, se considerarmos modelos vizinhos, a proporção de “acertos” sobe para 91%. Para o método ABC-PPMC, o modelo original foi selecionado em 100% das vezes, com um notável fator discriminatório entre modelos em  $p(\Psi = m | \mathbf{x})$ .

No Exemplo 36, averiguamos a consistência dos resultados obtidos pela abordagem ABCME. Comparamos estimativas pontuais para o índice de variação regular  $\alpha$  em três casos: temos acesso à distribuição a posteriori exata; estimamos  $\pi(\alpha | \mathbf{x})$  calculando a função de verossimilhança (MCMC); e estimamos  $\pi(\alpha | \mathbf{x})$  sem calcularmos  $f(\mathbf{x} | \alpha)$  (ABC). Notamos que a aproximação MCMC apresenta-se mais próxima da distribuição posterior exata do que a aproximação ABC. Além disso, o estimador  $\hat{\alpha}_{MCMC}$  apresentou EMQ muito menor do que  $\hat{\alpha}_{ABCME}$ . Tais resultados já eram esperados, pois o método MCMC conta com o cálculo  $f(\mathbf{x} | \alpha)$ , que carrega grande informação sobre  $\pi(\alpha | \mathbf{x})$ .

No Exemplo 37, vimos a relação das medidas-resumo utilizadas no método ABCME com o peso caudal da distribuição. Tal relação enaltece a importância de considerarmos um conjunto  $\Omega_S$  bem diversificado e coerente com a natureza do parâmetro.

No Exemplo 38, comparamos duas possíveis abordagens no caso em que queremos fazer inferência sobre o peso caudal de uma mistura de distribuições de variação regular. Vimos que

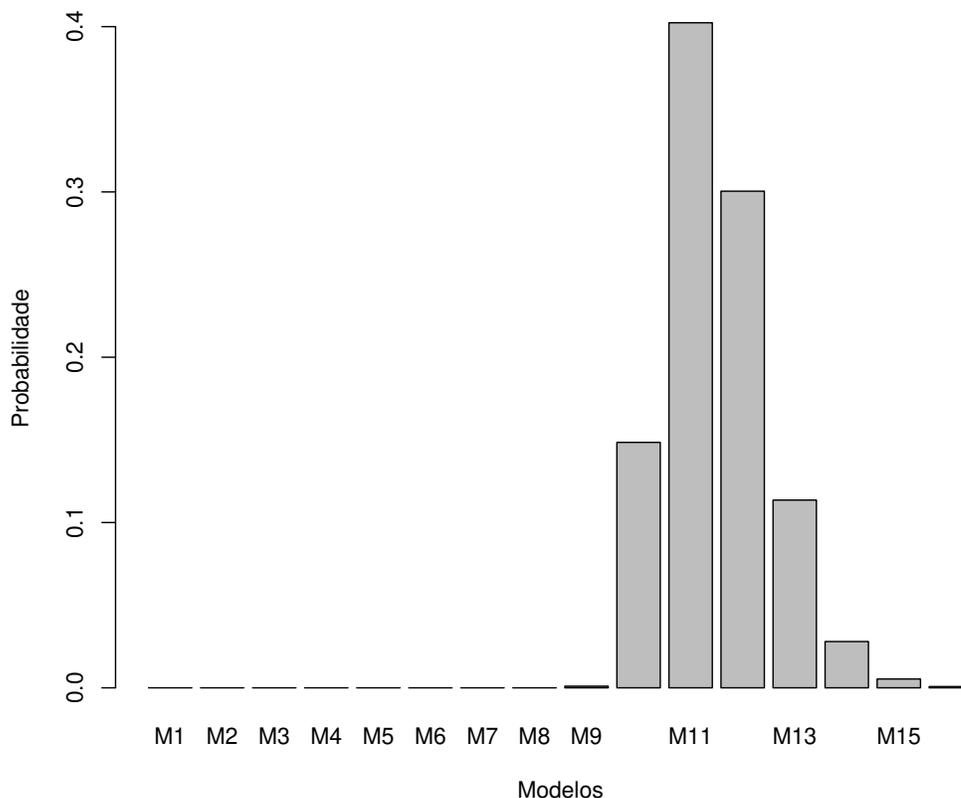


**Figura 6:** Gráfico da distribuição a priori para os modelos apresentados na Tabela B.9.

a metodologia ABC nos permite trabalhar de maneira fácil e interessante com o parâmetro de variação regular da mistura, nos possibilita até declarar diretamente uma distribuição a priori para  $\alpha = \min\{\alpha_1, \alpha_2\}$ . Tal maleabilidade é responsável por nos entregar uma distribuição a posteriori para  $\alpha$  mais concentrada do que a obtida pelo método MCMC proposto.

No Exemplo 39, verificamos a importância de se incluir a distribuição preditiva no critério de seleção de modelos. Vimos que o critério ABC-PPMC apresenta grande eficiência, e que seu uso pode ser mais indicado do que o do critério ABC-MC.

Atentamos que, com exceção do Exemplo 38, as técnicas utilizadas podem ser diretamente estendidas para distribuições subexponenciais.

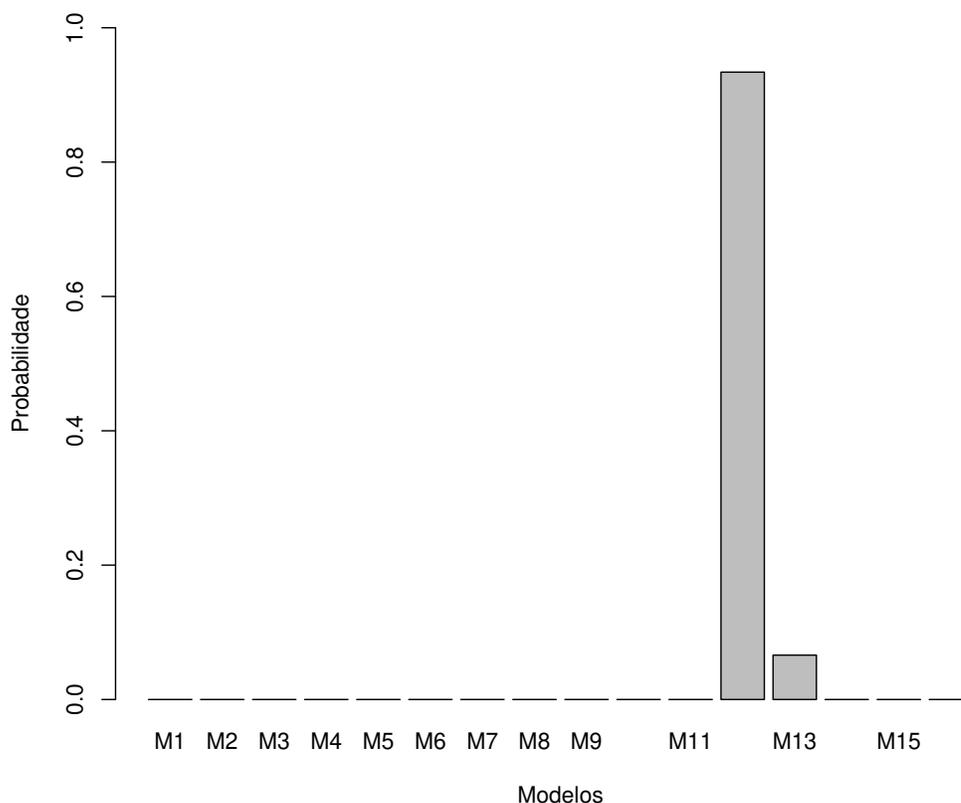


**Figura 7:** Gráfico da distribuição a posteriori pela abordagem ABC-MC para os modelos apresentados na Tabela B.9.

### 3.3 Aplicação em dados reais

A abordagem de distribuições de cauda pesada aparece em diversas áreas do conhecimento, com destaque para hidrologia e finanças. Tessier et al. (1996) modela as precipitações e vazões de rios via modelos de variação regular com índice  $2 < \alpha < 4$ . Mandelbrot (1967) apresenta evidências gráficas de que a variação diária do preço de algodão possui distribuição de variação regular com índice  $\alpha \approx 1.7$ .

Em 1998, Anderson e Meerschaert (1998) estudaram a vazão média mensal do Rio do Sal (*Salt River*), próximo a Roosevelt, Arizona. Tal rio nasce da confluência dos rios *White River* e *Black River*, e é um dos principais afluentes do rio Gila. Possui extensão aproximada de  $322 \text{ km}$  e sua bacia cobre uma área de  $35483 \text{ km}^2$ . O Rio do Sal apresenta grande variação em sua vazão mensal, oscilando entre  $2 \text{ m}^3/\text{s}$  e  $1698 \text{ m}^3/\text{s}$  ao longo dos anos. Em particular, tipicamente, o mês de janeiro apresenta valores razoavelmente altos de fluxo d'água (em relação aos demais meses)

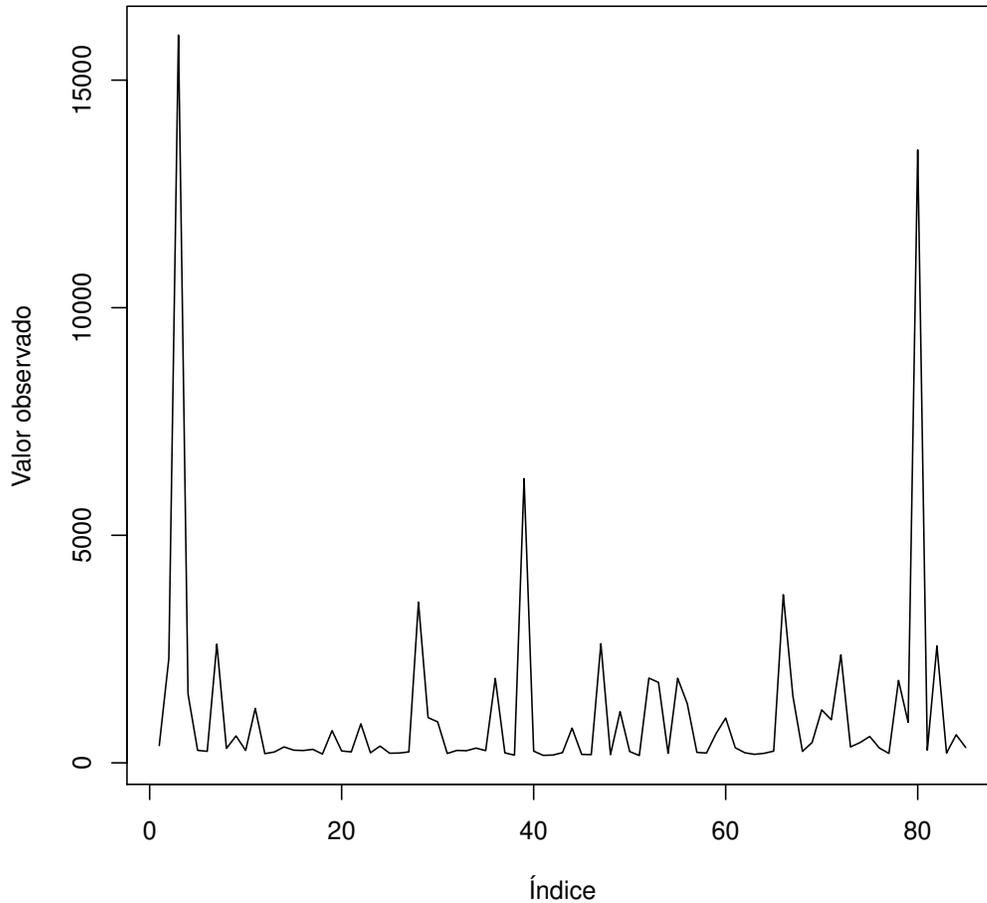


**Figura 8:** Gráfico da distribuição a posteriori pela abordagem ABC-PPMC para os modelos apresentados na Tabela B.9.

devido a cheia de *White River* que ocorre após as tempestades de inverno. No entanto, há anos em que a vazão do Rio do Sal atinge níveis muito mais altos que o esperado. Dessa forma, a modelagem estatística utilizando distribuições de cauda pesada pode ser bem razoável.

Na literatura é muito comum o uso de distribuições de variação regular para modelos hidrológicos, como exemplo, Hosking e Wallis (1987), Benson et al. (2000) e Anderson e Meerschaert (1998). No entanto, muitas ferramentas utilizadas no passado, como o estimador de Hill, Hill (1975), e suas generalizações (ver Beirlant et al. (2005)), para estimar o índice de variação da distribuição da amostra, tem seu uso comprometido quando esta não é bem ajustada a uma distribuição de Pareto. Daí a necessidade de introduzirmos novas abordagens ao problema.

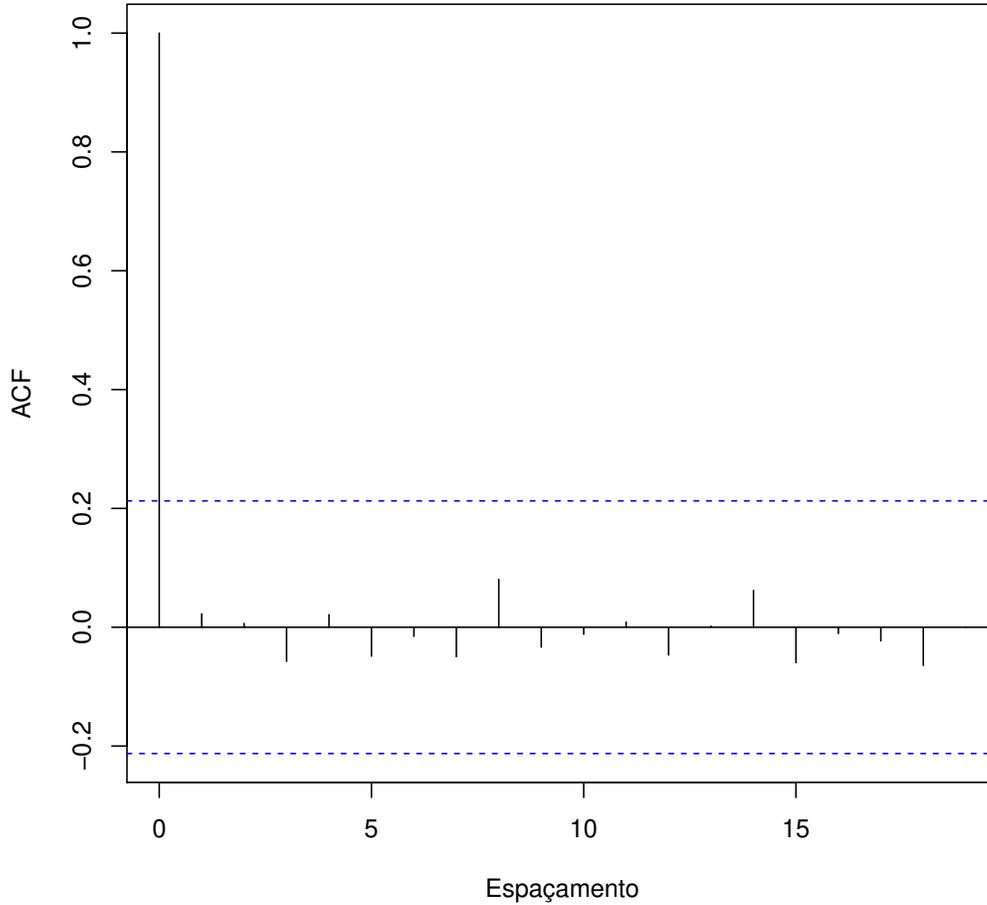
Consideremos as medições (em pés/s) das vazões médias do mês de janeiro do Rio do Sal, próximo a Roosevelt, nos anos de 1914 a 1998, como nossa amostra observada (os dados podem ser acessados gratuitamente em <http://waterdata.usgs.gov>).



**Figura 9:** Gráfico de traços para a amostra do exemplo de dados reais.

Desejamos, inicialmente, verificar se a suposição de independência das observações é razoável para este caso. As Figuras 9 e 10 apresentam os gráficos de traços e de ACF da amostra observada. Podemos notar que a correlação obtida é baixa e que não há fortes indícios de dependência entre as medições. Podemos notar também na Figura 9, alguns picos, ou seja, algumas observações com valores muito maiores que os demais registrados na amostra. A Tabela 6 apresenta algumas medidas de evidência de cauda pesada para os dados observados. A Figura 11 apresenta o gráfico da função de excesso médio para a amostra em questão. Pelo obtido, notamos que os altos valores das medidas de evidência da Tabela 6 e a tendência linear na Figura 11 justificam uma modelagem utilizando distribuições de variação regular.

Desejamos selecionar, em um conjunto de modelos  $\mathbf{P}$ , o elemento mais adequado para



**Figura 10:** Gráfico de autocorrelações para a amostra do exemplo de dados reais.

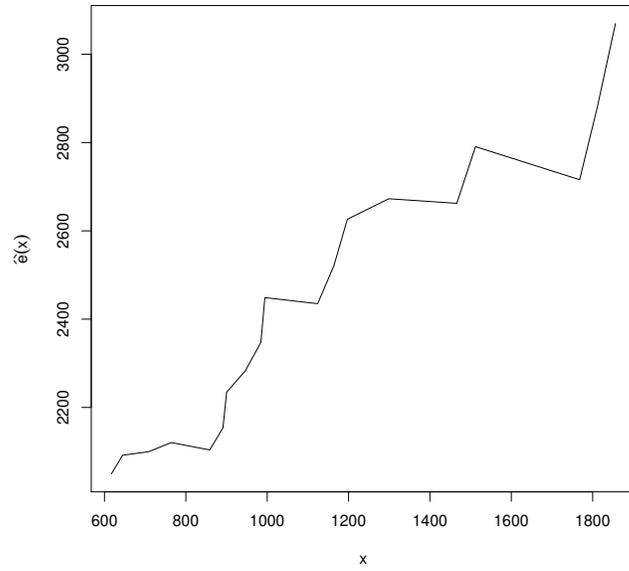
explicar o comportamento da amostra observada. A Tabela B.10 apresenta a descrição de  $\mathbf{P}$ . Consideremos uma distribuição a priori uniforme para o conjunto de índices  $\Psi$ , ou seja,  $p(\Psi = m) = 1/30$ ,  $m \in \{1, \dots, 30\}$ . Para cada modelo, consideramos o mesmo  $\Omega_S$  apresentado na Tabela B.6 e aplicamos o método ABCME. Em seguida, aplicamos o método ABC-PPMC como apresentado na Seção 2.5, e obtivemos a distribuição a posteriori dos modelos representada na Figura 12. Notemos que nosso método seleciona o modelo M21. Se definirmos como estimador de  $\alpha$ ,  $\hat{\alpha}$ , a esperança da distribuição a posteriori obtida, observamos estimativa de  $\hat{\alpha} = 5.4941$ .

Para verificarmos se o modelo está bem ajustado, utilizaremos o p-valor preditivo a posteriori, apresentado na Seção 2.4. Utilizaremos três medidas de teste: quantil amostral de 90%;  $\hat{O}b(X, 8)$ ; e  $\hat{R}a_{8,2}(X)$ . A Figura 13 apresenta o comportamento amostral a posteriori de

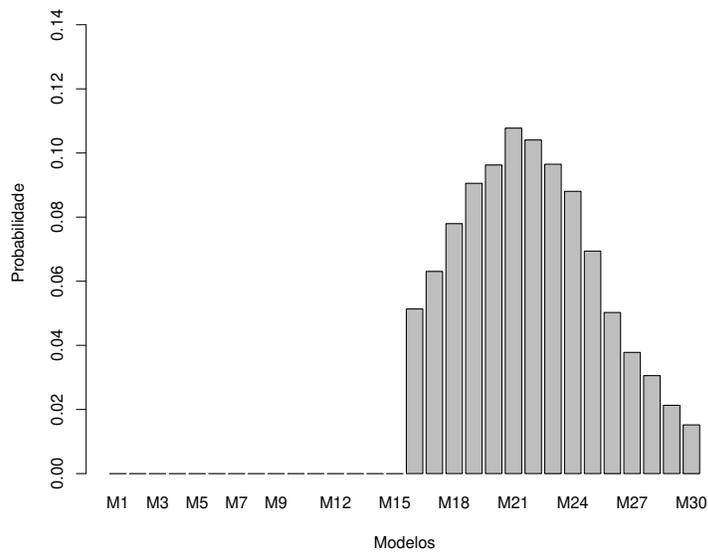
**Tabela 6:** Medidas de evidência de cauda pesada para o exemplo de dados reais.

Medidas-resumo	Valor observado na amostra	Valor exato para uma distribuição $Exp(\lambda)$
$\hat{O}b(X, 4)$	0.8683	0.7500
$\hat{O}b(X, 8)$	0.8020	0.7500
$\hat{R}a_{4,2}(X)$	0.5563	0.4000
$\hat{R}a_{8,2}(X)$	0.4580	0.2222

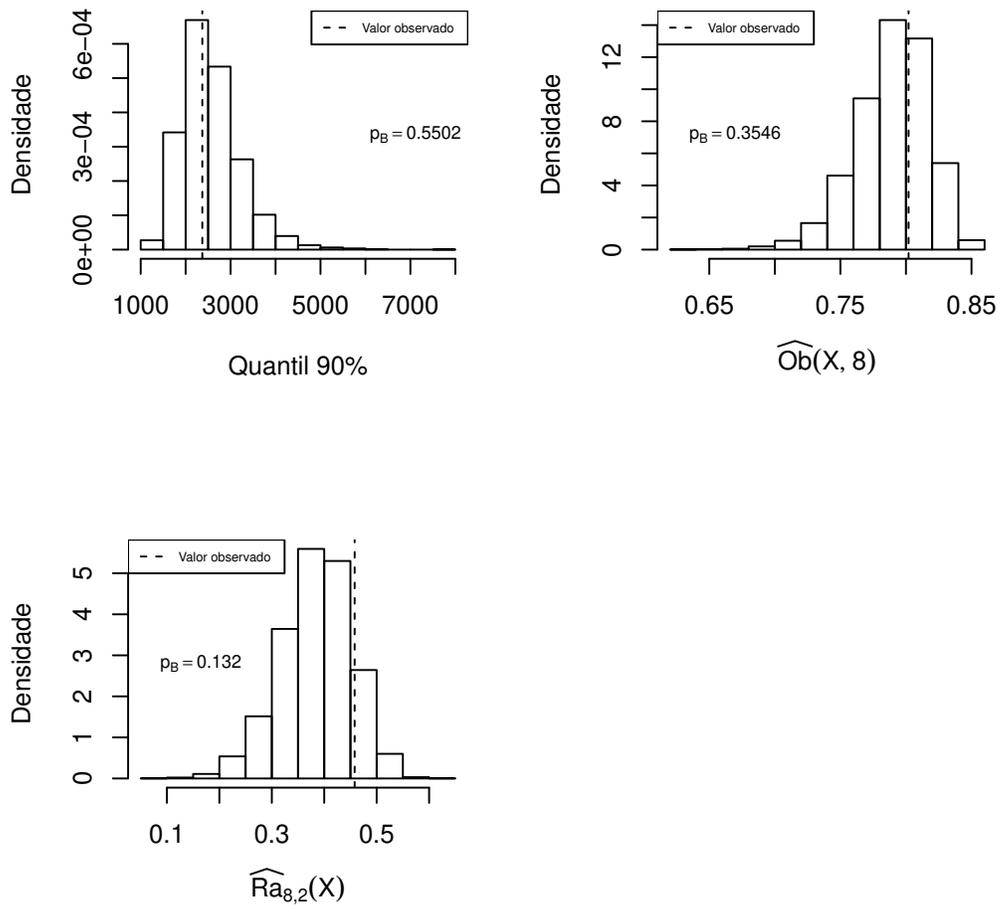
tais medidas, condicionadas ao modelo M21 e aos dados observados. Observamos que todos os p-valores preditivos obtidos foram maiores que 10%, o que nos indica que o modelo, ao que se diz respeito à cauda da distribuição, está bem ajustado. Assim como Gelman et al. (2013), reforçamos que um p-valor maior que 10% não deve ser interpretado como uma afirmação de que o modelo é verdadeiro, mas sim de que o modelo se ajusta bem a um aspecto particular dos dados.



**Figura 11:** Gráfico da função de excesso médio estimada para a amostra do exemplo de dados reais.



**Figura 12:** Gráfico da distribuição a posteriori para os modelos apresentados na Tabela B.10.



**Figura 13:** Histogramas de 10000 simulações das estatísticas de interesse via distribuição preditiva a posteriori sob o modelo M21.

# Capítulo 4

## Considerações finais

Ao longo deste texto, analisamos e desenvolvemos algumas técnicas de inferência bayesiana para distribuições de cauda pesada, que não envolvessem o cálculo explícito da função de verossimilhança. Apresentamos desde técnicas de análise exploratória até técnicas de seleção e verificação de ajuste de modelos. Estudamos algumas medidas-resumo (índice de obesidade generalizado e razões de estatísticas de ordem) que nos auxiliam na análise do peso caudal em questão sem depender de suposições sobre a distribuição dos dados. No entanto, vimos que tais estatísticas são dependentes de *bootstrap* para sua estimação. Futuramente, desejamos desenvolver medidas mais simples e mais precisas que tenham as mesmas vantagens que as duas técnicas supracitadas.

No capítulo 3, ilustramos, a partir de diversos exemplos, o uso e algumas características das ferramentas apresentadas neste trabalho. O Exemplo 36 apresenta o uso do algoritmo ABCME e compara seus resultados com os obtidos pelo algoritmo MCMC-MH. Vimos que, como esperado, a aproximação obtida pelo método MCMC-MH está mais próxima da densidade a posteriori exata do que aquela obtida pelo método ABCME (ver página 39). O Exemplo 37 ilustra a mudança da estatística responsável pela distribuição a posteriori com menor entropia do método ABCME, quando os dados provém de uma distribuição de cauda muito pesada. O Exemplo 38 propõe duas diferentes abordagens de inferência bayesiana quando o modelo considerado é uma mistura de distribuições de variação regular. O Exemplo 39 compara os métodos de seleção de modelos ABC-MC e ABC-PPMC, e apresenta um forte indício de maior acertividade do segundo método considerado.

Na Seção 1.3.3, apresentamos uma abordagem gráfica, via FEM, para identificarmos dois possíveis padrões: logarítmico (pode caracterizar amostragem Weibull); linear (pode caracterizar amostragem de variação regular). Vimos também que o uso do coeficiente angular da reta de mínimos quadrados da FEM,  $\hat{\phi}$ , para estimar o índice de variação regular,  $\alpha$ , tende a entregar

estimativas que subestimam o peso caudal da distribuição. No Exemplo 38, ilustramos o uso de  $\hat{\phi}$  como uma possível medida-resumo para a abordagem ABCME para o índice de variação da mistura modelada. Pelo apresentado na Tabela B.7, vimos que tal estatística não estava entre as que produziram as estimativas da distribuição a posteriori com menores entropias. No entanto, em estudos preliminares, notou-se que o uso de técnicas de redução de viés (*jackknife*) na estimação de  $\hat{\phi}$ , produz uma redução notável na entropia obtida no método ABCME, tornando tal estatística a escolhida pelo algoritmo. Apesar da melhoria obtida quanto a entropia, o método passa a ser muito menos eficiente, pois o cálculo da estimativa *jackknife* de  $\hat{\phi}$  é muito caro computacionalmente. Em estudos futuros, desejamos implementar no método ABCME, um balanceamento para a escolha das estatísticas que valorize a menor entropia e penalize a baixa eficiência do algoritmo.

Foss et al. (2013) apresenta que a classe  $\mathcal{L}$  é fechada com relação a misturas e Foss et al. (2009) discute sobre condições suficientes e necessárias para que misturas de distribuições subexponenciais continuem sendo subexponenciais. Assim, desejamos explorar o uso do método ABCME no estudo de misturas de distribuições de cauda pesada que não pertençam à classe  $\mathcal{VR}_\alpha$ .

Por fim, desejamos estudar o uso da técnica de seleção de modelos aqui desenvolvida, ABC-PPMC, para modelagens de distribuições pertencentes à classe  $\mathcal{L}$ .

# Referências

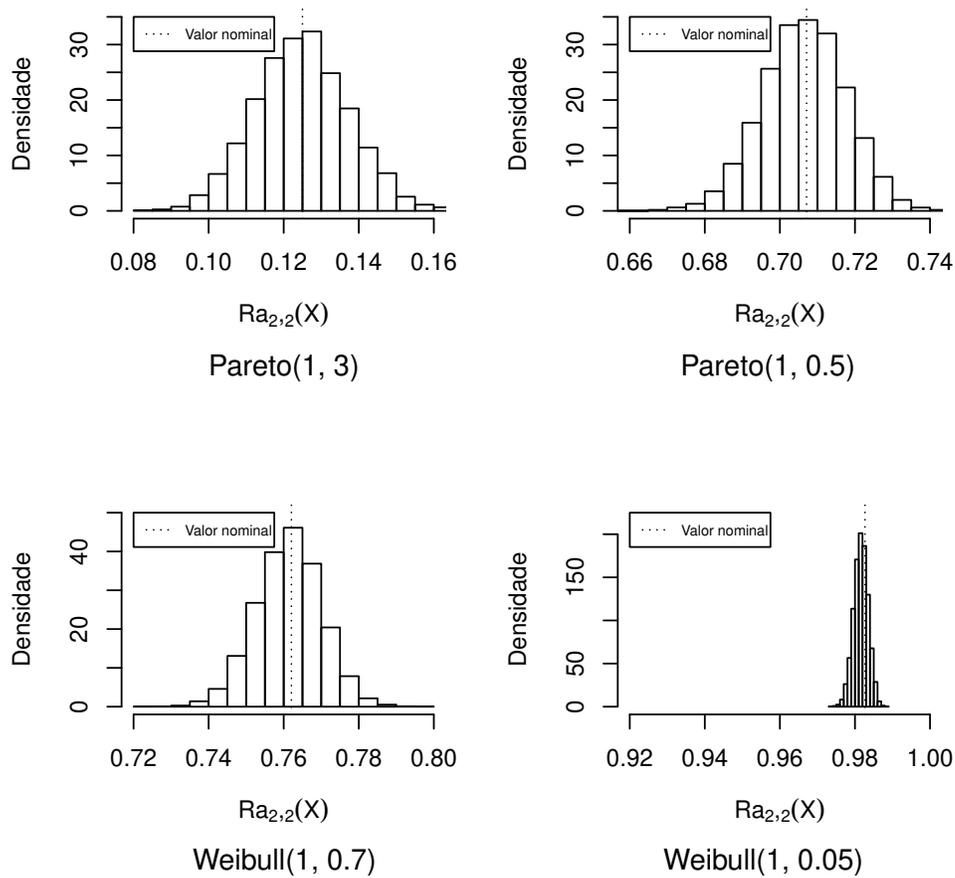
- Anderson, P. L. Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resources Research*, 34, 2271–2280.
- Beirlant, J. Teugels, J. L. (1992). Modeling large claims in non-life insurance. *Insurance: Mathematics and Economics*, 11, 17–29.
- Benson, D. A., Wheatcraft, S. W., Meerschaert, M. M. (2000). Application of a fractional advection-dispersion equation. *Water Resources Research*, 36, 1403–1412.
- Bingham, N. H., Goldie, C. M., Teugels, J. L. (1989). *Regular variation*. Cambridge University Press, Cambridge.
- Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 1). Springer New York.
- Boltz, S., Debreuve, E., Barlaud, M. (2007). Knn-based high-dimensional kullback-leibler distance for tracking. *Image analysis for multimedia interactive services, 2007. wiamis'07. eighth international workshop on* (p. 16). IEEE.
- Chib, S. Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49, 327–335.
- Chistyakov, V. P. (1964). A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability & Its Applications*, 9, 640–648.
- Cooke, R. M. Nieboer, D. (2011). Heavy-tailed distributions: data, diagnostics, and new developments. *Resources for the Future Discussion Paper*.
- DeGroot, M. H. Schervish, M. J. (2011). *Probability and statistics* (4<sup>a</sup> ed.). Addison-Wesley Reading, Massachusetts.
- Diggle, P. J. Gratton, R. J. (1984). Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 193–227.
- Embrechts, P. Goldie, C. M. (1982). On convolution tails. *Stochastic Processes and their Applications*, 13, 263–278.

- Embrechts, P., Goldie, C. M., Veraverbeke, N. (1979). Subexponentiality and infinite divisibility. *Probability Theory and Related Fields*, 49, 335–347.
- Feller, W. (1971). *An introduction to probability theory and its applications* (Vol. 2). John Wiley & Sons, New York.
- Foss, S., Korshunov, D., Zachary, S. (2009). Convolutions of long-tailed and subexponential distributions. *Journal of Applied Probability*, 756–767.
- Foss, S., Korshunov, D., Zachary, S. (2013). *An introduction to heavy-tailed and subexponential distributions*. Springer.
- Geisser, S. Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. *Markov chain monte carlo in practice* (pp. 145–161). Springer.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Halliwell, L. J. (2013). *Classifying the tails of loss distributions*. Springer.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57, 97–109.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3, 1163–1174.
- Hosking, J. R. M. Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29, 339–349.
- Joyce, P. Marjoram, P. (2008). Approximately sufficient statistics and bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7.
- Klüppelberg, C. (1989). Subexponential distributions and characterizations of related classes. *Probability Theory and Related Fields*, 82, 259–269.
- Laud, P. W. Ibrahim, J. G. (1995). Predictive model selection. *Journal of the Royal Statistical Society. Series B (Methodological)*, 247–262.
- Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 393–413.
- McCullagh, P. (2002). What is a statistical model? *Annals of statistics*, 1225–1267.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21, 1087–1092.
- Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory*. Eindhoven University of Technology.

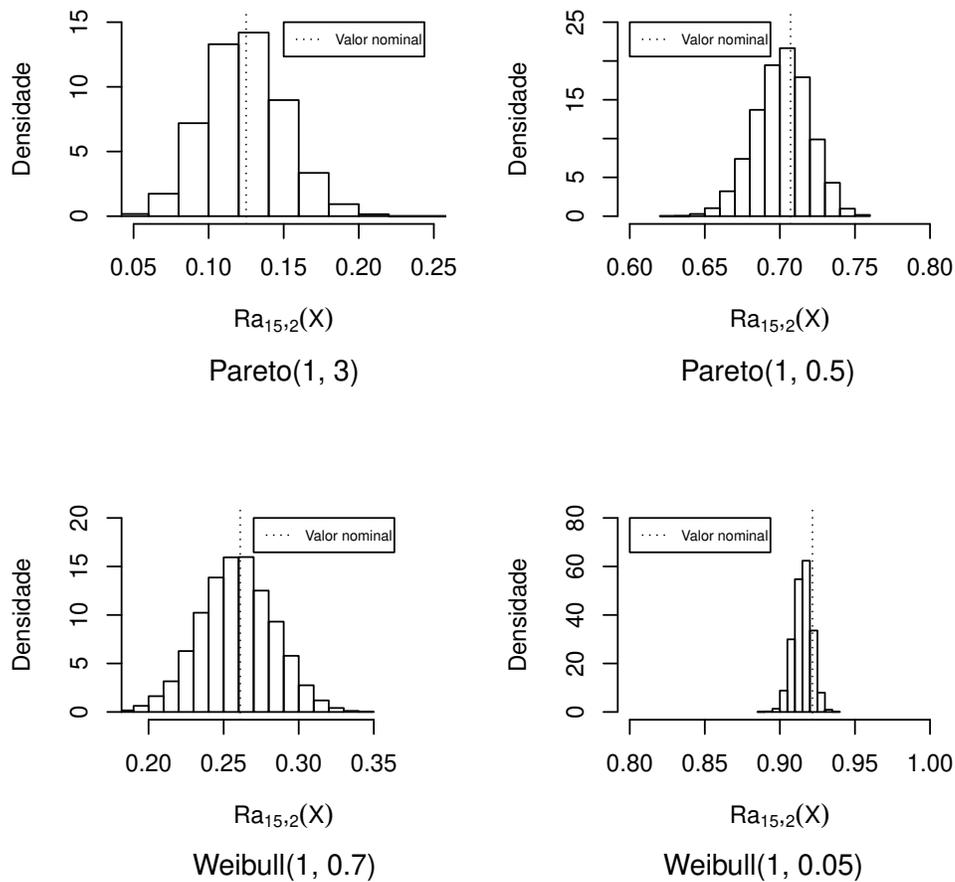
- Nunes, M. A. Balding, D. J. (2010). On optimal selection of summary statistics for approximate bayesian computation. *Statistical applications in genetics and molecular biology*, 9.
- Ross, S. (2009). *A first course in probability* (8<sup>a</sup> ed.). Pearson.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, 12, 1151–1172.
- Santos, J. D. (2007). *Discussões sobre a relação entre distribuições de cauda pesada e conflitos de informação em inferência bayesiana* (Dissertação de mestrado, IMECC, UNICAMP).
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Singh, H., Misra, N., Hnizdo, V., Fedorowicz, A., Demchuk, E. (2003). Nearest neighbor estimates of entropy. *American journal of mathematical and management sciences*, 23, 301–321.
- Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., Pecknold, S. (1996). Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 101, 26427–26440.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M. P. H. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6, 187–202.
- Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54–59.

# Apêndice A

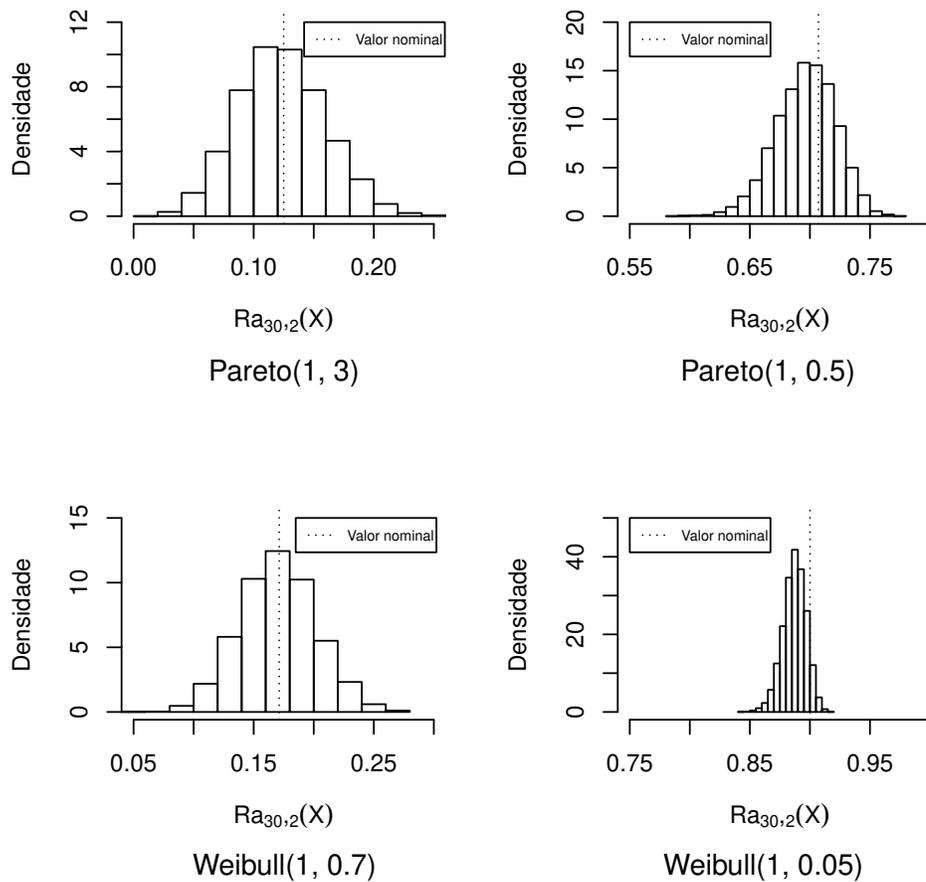
## Figuras



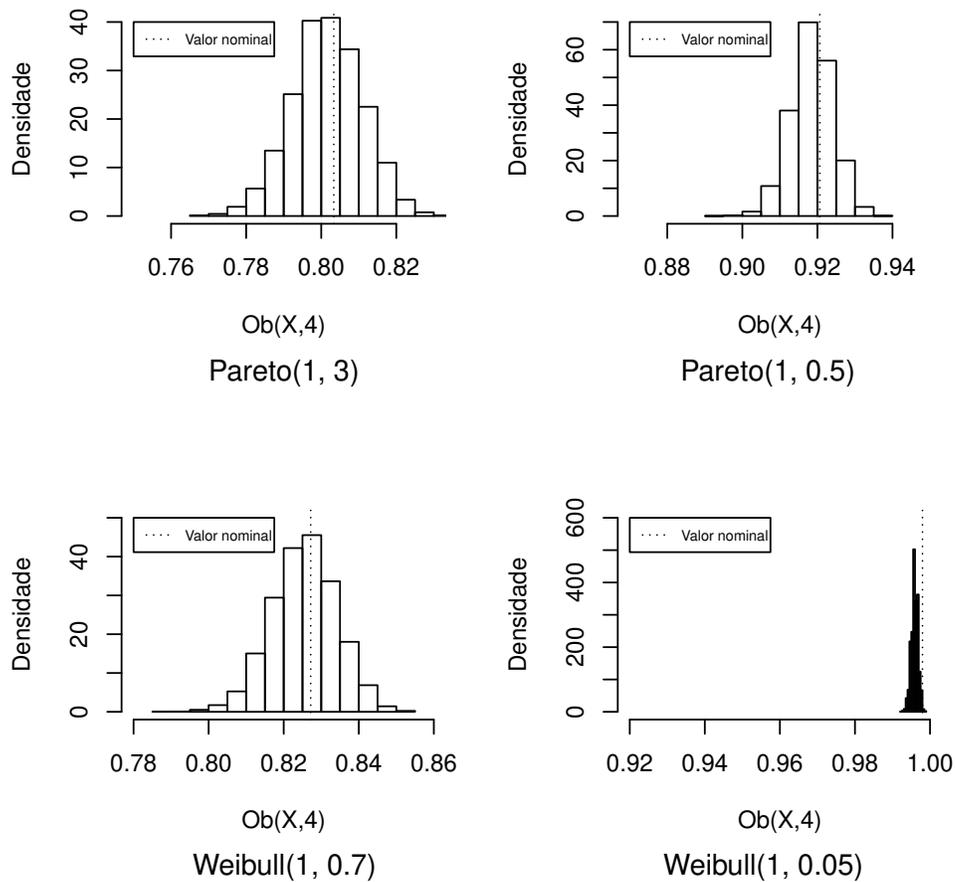
**Figura A.1:** Histogramas de 10000 simulações de  $\hat{Ra}_{2,2}(X)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



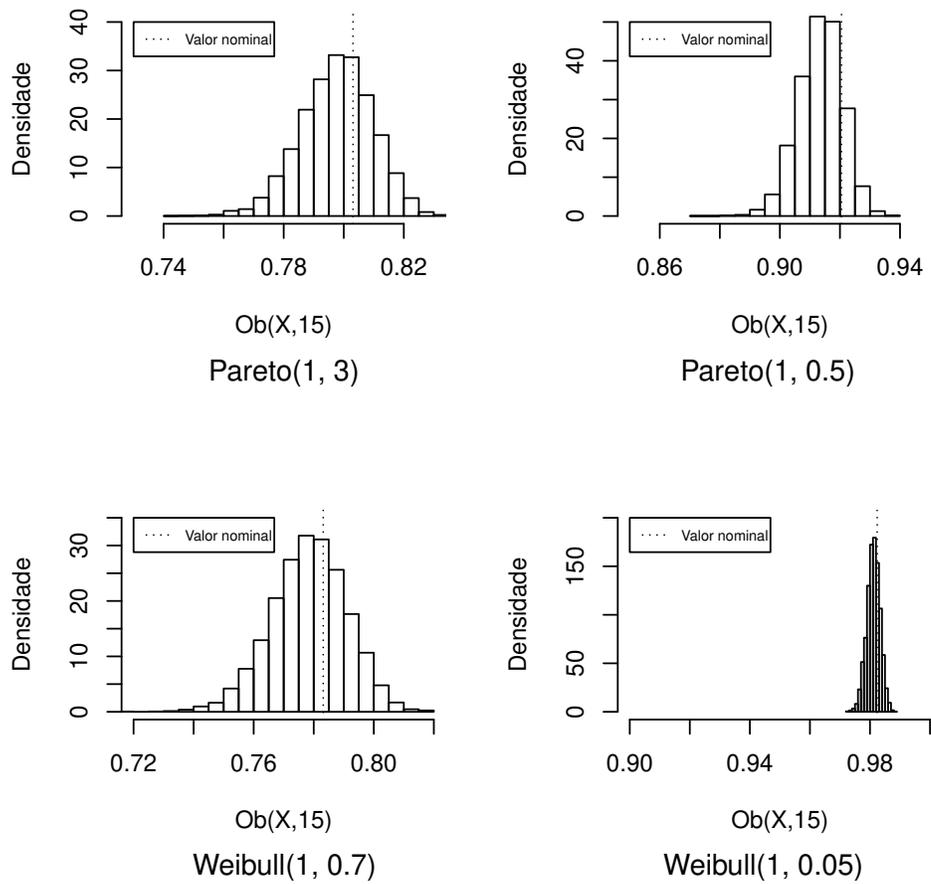
**Figura A.2:** Histogramas de 10000 simulações de  $\hat{Ra}_{15,2}(X)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



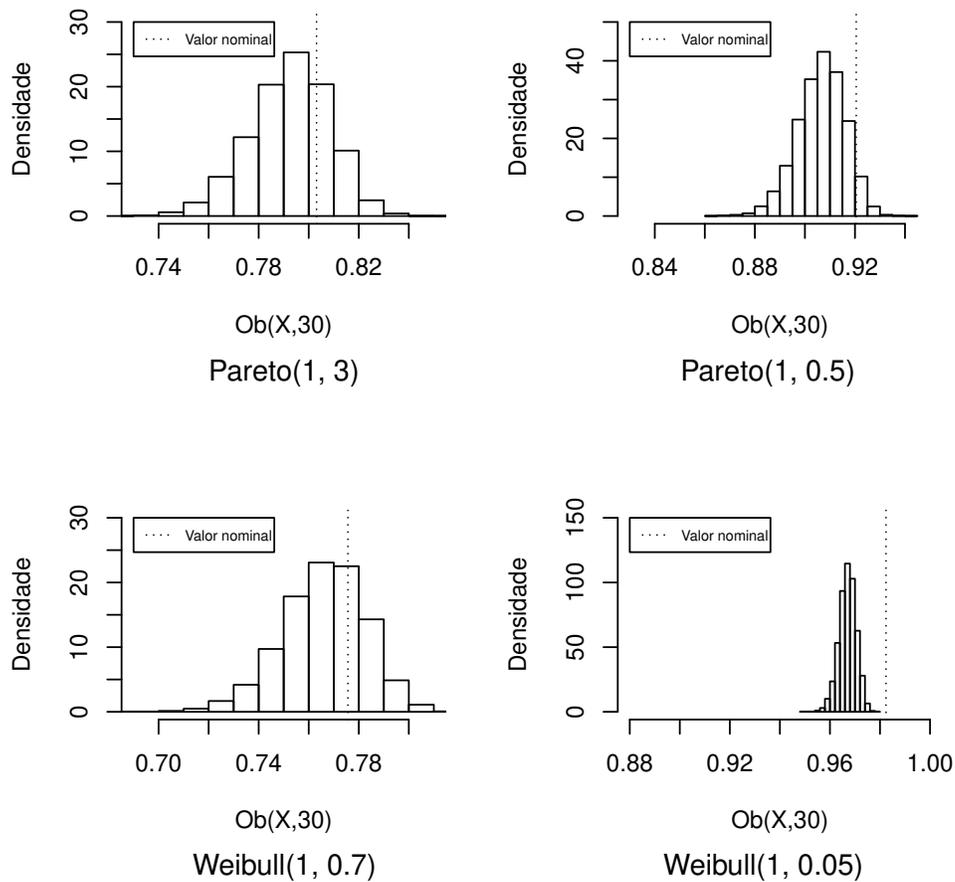
**Figura A.3:** Histogramas de 10000 simulações de  $\hat{Ra}_{30,2}(X)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



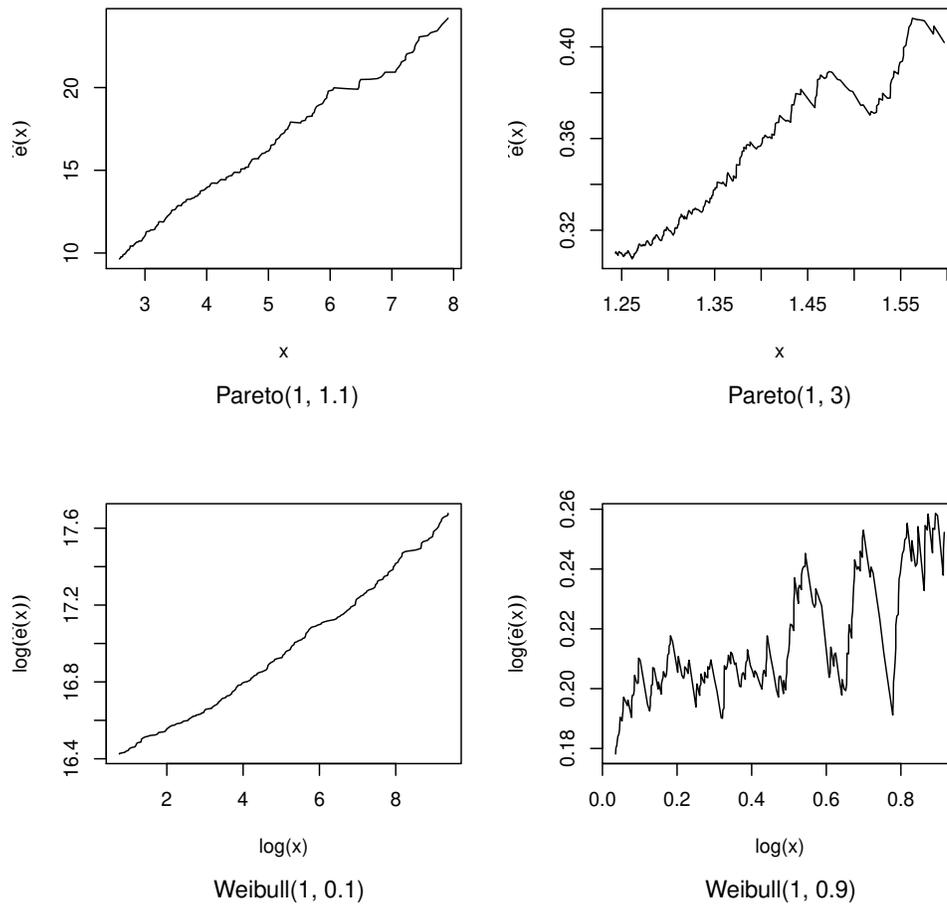
**Figura A.4:** Histogramas de 10000 simulações de  $\hat{O}b(X, 4)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



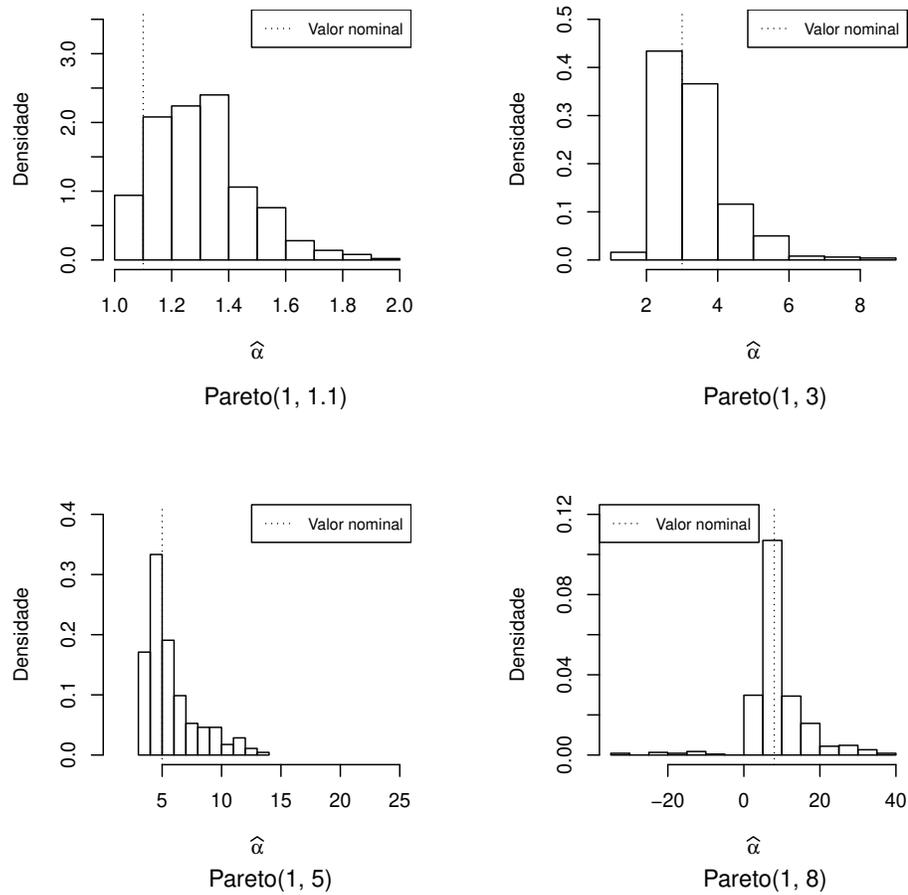
**Figura A.5:** Histogramas de 10000 simulações de  $\hat{Ob}(X, 15)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



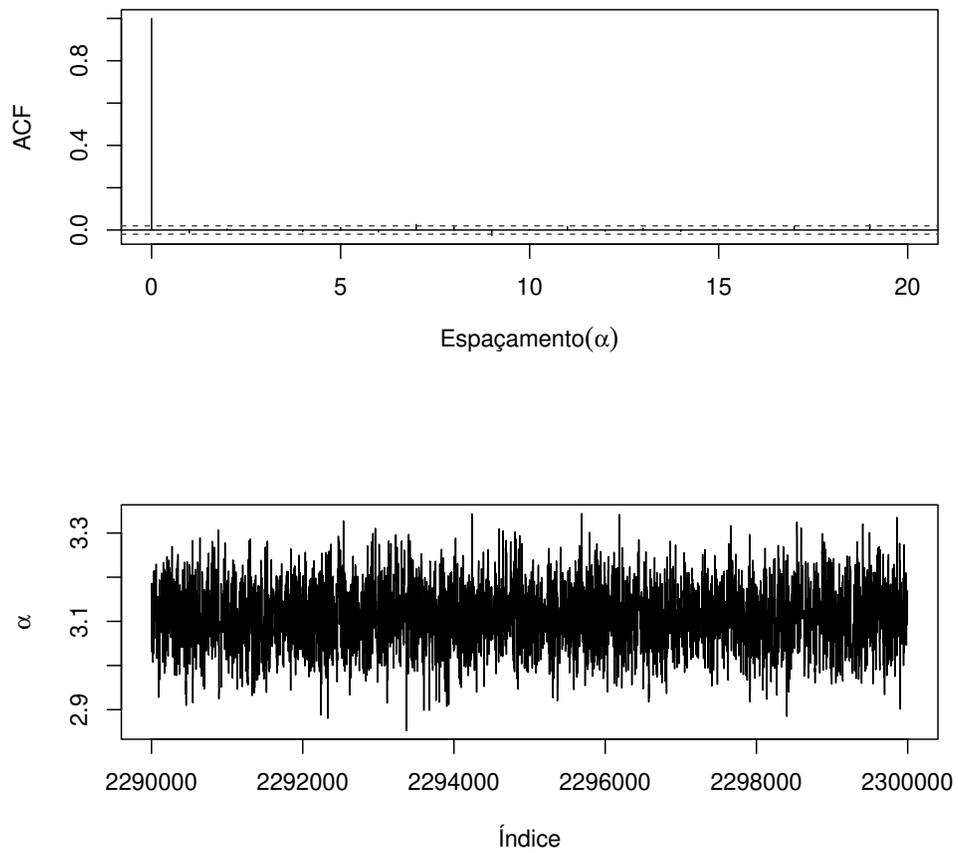
**Figura A.6:** Histogramas de 10000 simulações de  $\hat{Ob}(X, 30)$  para diversas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.



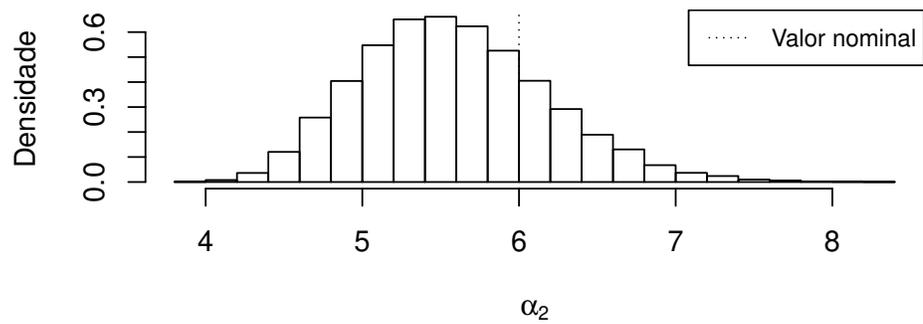
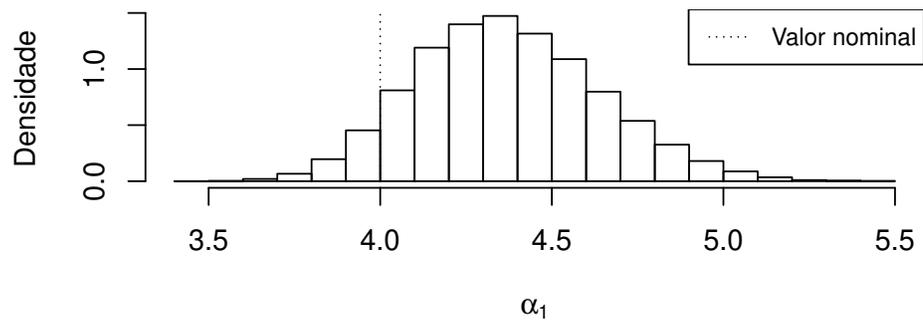
**Figura A.7:** Gráficos de  $\hat{e}_X(x)$  para diversas distribuições considerando tamanho de amostra 1000.



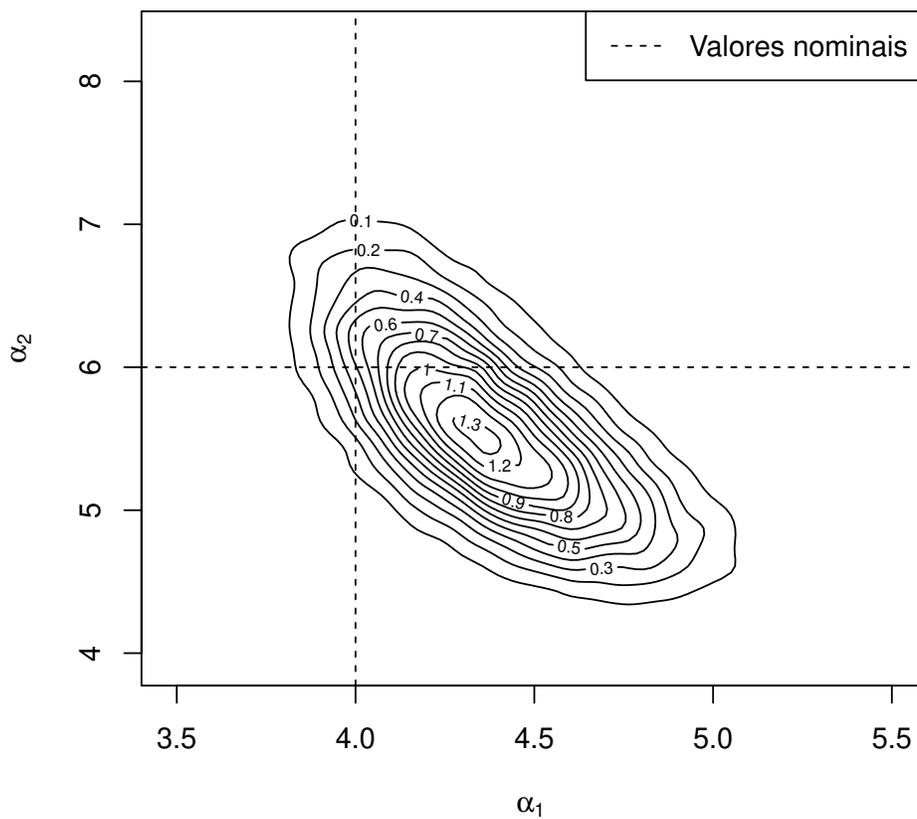
**Figura A.8:** Histogramas de 10000 simulações da estimativa do índice de variação regular,  $\alpha$ , via FEM,  $\hat{\alpha}_{FEM}$ , considerando tamanho de amostra 1000.



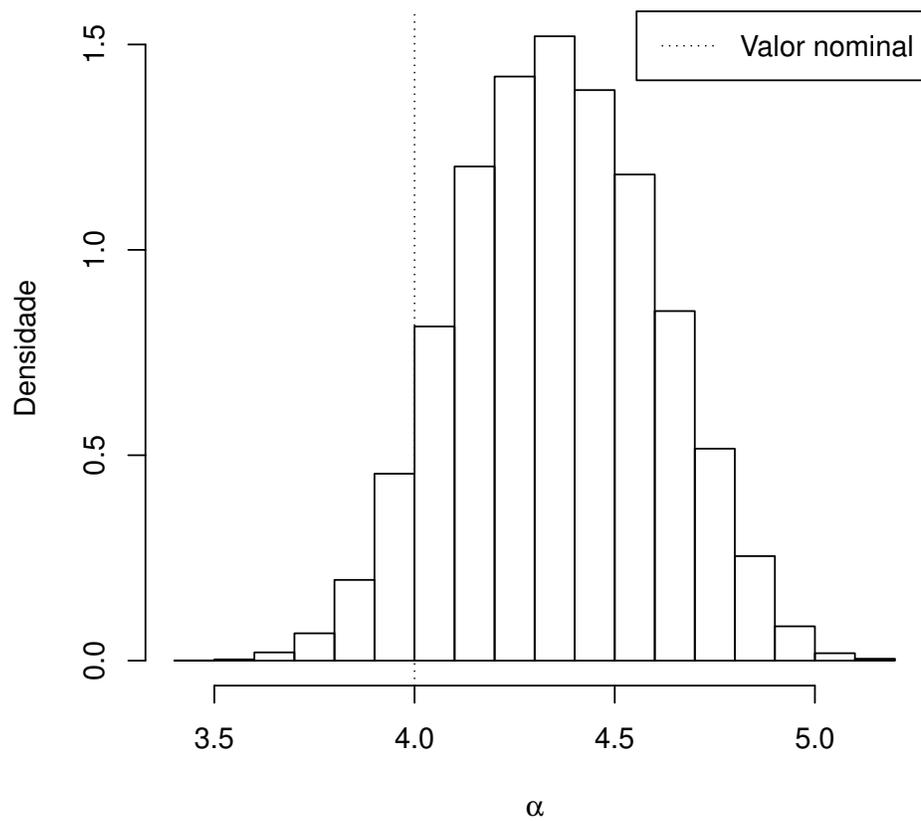
**Figura A.9:** Gráficos de autocorrelação e de traços para estimação MCMC para o Exemplo 36.



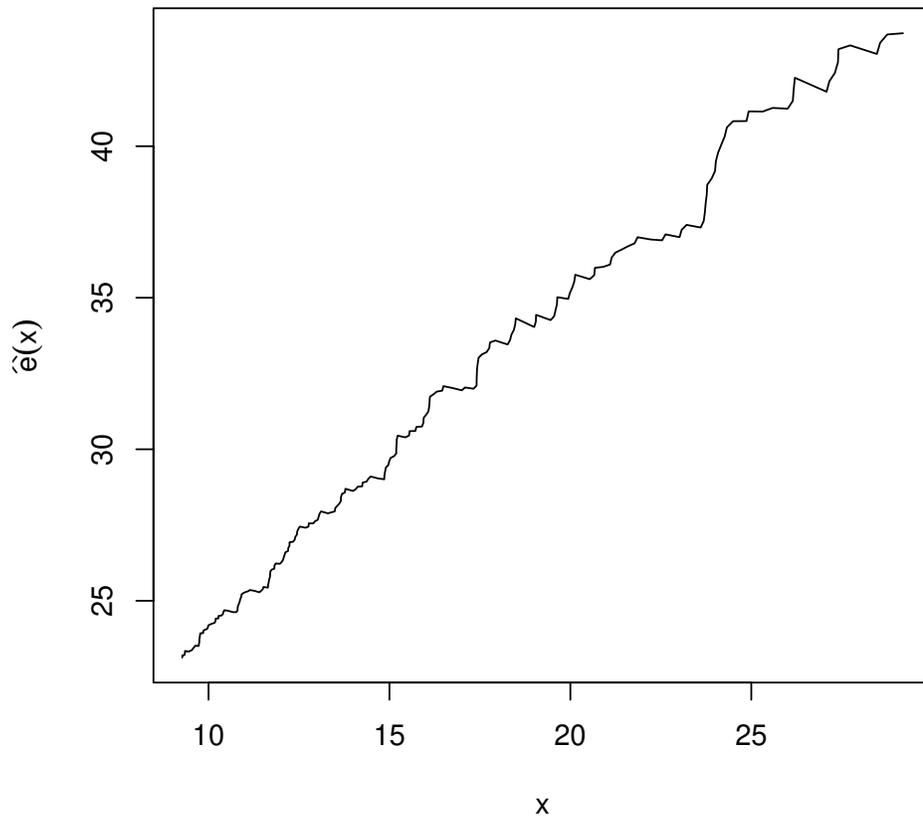
**Figura A.10:** Histogramas das amostras MCMC da distribuição marginal a posteriori dos índice de variação regular  $\alpha_1$  e  $\alpha_2$  que compõe a distribuição dos dados do Exemplo 38.



**Figura A.11:** Curvas de nível da densidade conjunta da amostra MCMC da distribuição a posteriori dos índice de variação regular  $\alpha_1$  e  $\alpha_2$ , que compõe a distribuição dos dados do Exemplo 38.



**Figura A.12:** Histograma da amostra da distribuição a posteriori do índice de variação regular,  $\alpha$ , via método MCMC para o Exemplo 38.



**Figura A.13:** Gráfico da função de excesso médio estimada para o Exemplo 39.

# Apêndice B

## Tabelas

**Tabela B.1:** Viés médios observados em 10000 simulações de  $\hat{R}a_{2,2}(X)$ ,  $\hat{R}a_{15,2}(X)$  e  $\hat{R}a_{30,2}(X)$  para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.

Distribuição	$\hat{R}a_{2,2}(X)$	$\hat{R}a_{15,2}(X)$	$\hat{R}a_{30,2}(X)$
<i>Pareto</i> (1, 3)	$-2 \times 10^{-6}$	-0.0005	-0.0010
<i>Pareto</i> (1, 0.5)	-0.0006	-0.0052	-0.0106
<i>Weibull</i> (1, 0.7)	-0.0006	-0.0011	-0.0012
<i>Weibull</i> (1, 0.05)	-0.0010	-0.0068	-0.0132

**Tabela B.2:** EQM observados em 10000 simulações de  $\hat{R}a_{2,2}(X)$ ,  $\hat{R}a_{15,2}(X)$  e  $\hat{R}a_{30,2}(X)$  para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.

Distribuição	$\hat{R}a_{2,2}(X)$	$\hat{R}a_{15,2}(X)$	$\hat{R}a_{30,2}(X)$
<i>Pareto</i> (1, 3)	0.0015	0.0067	0.0128
<i>Pareto</i> (1, 0.5)	0.0012	0.0037	0.0075
<i>Weibull</i> (1, 0.7)	0.0008	0.0059	0.0100
<i>Weibull</i> (1, 0.05)	$5 \times 10^{-5}$	0.0008	0.0027

**Tabela B.3:** Viés médios observados em 10000 simulações de  $\hat{O}b(X, 4)$ ,  $\hat{O}b(X, 15)$  e  $\hat{O}b(X, 30)$  para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.

Distribuição	$\hat{O}b(X, 4)$	$\hat{O}b(X, 15)$	$\hat{O}b(X, 30)$
<i>Pareto</i> (1, 3)	-0.0016	-0.0051	-0.0105
<i>Pareto</i> (1, 0.5)	-0.0020	-0.0069	-0.0138
<i>Weibull</i> (1, 0.7)	-0.0017	-0.0049	-0.0098
<i>Weibull</i> (1, 0.05)	-0.0020	-0.0077	-0.0152

**Tabela B.4:** EQM observados em 10000 simulações de  $\hat{O}b(X, 4)$ ,  $\hat{O}b(X, 15)$  e  $\hat{O}b(X, 30)$  para algumas distribuições. Em cada repetição, utilizamos uma amostra de tamanho 500, e a estimativa calculada foi obtida através de 20000 reamostragens.

Distribuição	$\hat{O}b(X, 4)$	$\hat{O}b(X, 15)$	$\hat{O}b(X, 30)$
<i>Pareto</i> (1, 3)	0.0009	0.0017	0.0037
<i>Pareto</i> (1, 0.5)	0.0003	0.0010	0.0028
<i>Weibull</i> (1, 0.7)	0.0008	0.0018	0.0037
<i>Weibull</i> (1, 0.05)	$5 \times 10^{-5}$	0.0006	0.0024

**Tabela B.5:** Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 36.

$\Omega_S$	Valor observado na amostra	Entropia estimada
$\hat{O}b(X, 15)$	0.8040	0.1680
$\hat{O}b(X, 30)$	0.7910	0.2411
$\hat{R}a_{15,2}(X)$	0.1970	-0.5250
$\hat{R}a_{30,2}(X)$	0.1530	-0.4823
$\frac{1}{1000} \sum_{i=1}^{1000} \mathbf{X}_i$	<b>2.1789</b>	<b>-1.1005</b>
$\frac{1}{1000} \sum_{i=1}^{1000} X_i^2$	7.2879	-0.8230
$\frac{1}{1000} \sum_{i=1}^{1000} \log X_i$	0.6435	-1.0983
<i>Quantil 10%</i>	1.1631	-0.8083
<i>Quantil 90%</i>	3.4596	-0.9124
<i>Mediana</i>	1.7137	-0.9551
<i>Moda</i>	1.1062	-0.621

**Tabela B.6:** Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 37.

$\Omega_S$	Valor observado na amostra	Entropia estimada
$\hat{O}b(X, 15)$	0.8905	-0.7028
$\hat{O}b(X, 30)$	0.8855	-0.5480
$\hat{R}a_{15,2}(X)$	0.6273	-1.1119
$\hat{R}a_{30,2}(X)$	0.6290	-0.9636
$\frac{1}{1000} \sum_{i=1}^{1000} X_i$	997.5902	-0.9666
$\frac{1}{1000} \sum_{i=1}^{1000} X_i^2$	$\approx 3 \times 10^8$	-0.5341
$\frac{1}{1000} \sum_{i=1}^{1000} \log \mathbf{X}_i$	<b>2.4815</b>	<b>-1.6799</b>
<i>Quantil 10%</i>	2.0254	-1.3459
<i>Quantil 90%</i>	125.3399	-1.5446
<i>Mediana</i>	7.7588	-1.5536
<i>Moda</i>	2.1383	-0.8212

**Tabela B.7:** Conjunto de medidas-resumo para o método ABCME e entropias das densidades a posteriori obtidas para o Exemplo 38.

$\Omega_S$	Valor observado na amostra	Entropia estimada
$\hat{O}b(X, 15)$	0.7470	0.2865
$\hat{O}b(X, 30)$	0.7655	0.2899
$\hat{R}a_{15,2}(X)$	0.0675	-0.1516
$\hat{R}a_{30,2}(X)$	0.1080	-0.1233
$\frac{1}{1000} \sum_{i=1}^{1000} \mathbf{X}_i$	<b>1.7215</b>	<b>-0.3182</b>
$\frac{1}{1000} \sum_{i=1}^{1000} X_i^2$	3.7555	-0.1725
$\frac{1}{1000} \sum_{i=1}^{1000} \log X_i$	0.4808	-0.1648
<i>Quantil 10%</i>	1.1496	0.3151
<i>Quantil 90%</i>	2.4042	-0.0023
<i>Mediana</i>	1.5246	0.2863
<i>Moda</i>	1.2646	0.3212
$\hat{\phi}$	0.6901	-0.0912

**Tabela B.8:** Medidas de evidência de cauda pesada para o Exemplo 39.

Medidas-resumo	Valor observado na amostra	Valor exato para uma distribuição $Exp(\lambda)$
$\hat{O}b(X, 15)$	0.8225	0.75
$\hat{O}b(X, 30)$	0.8015	0.75
$\hat{R}a_{15,2}(X)$	0.3688	0.125
$\hat{R}a_{30,2}(X)$	0.3048	0.0645

**Tabela B.9:** Conjunto de modelos candidatos para o Exemplo 39.

Modelos	Distribuição das variáveis observáveis	Distribuição a priori de $\theta$
$M_1$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (1, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_2$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (2, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_3$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (3, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_4$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (4, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_5$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (5, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_6$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (6, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_7$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (7, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_8$	$X_1, \dots, X_{900}$ a.a. <i>Weibull</i> (8, $\theta$ )	$\theta \sim Unif(0, 1)$
$M_9$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 1)	$\theta \sim Gama(4, 3)$
$M_{10}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 2)	$\theta \sim Gama(4, 3)$
$M_{11}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 3)	$\theta \sim Gama(4, 3)$
$M_{12}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 4)	$\theta \sim Gama(4, 3)$
$M_{13}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 5)	$\theta \sim Gama(4, 3)$
$M_{14}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 6)	$\theta \sim Gama(4, 3)$
$M_{15}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 7)	$\theta \sim Gama(4, 3)$
$M_{16}$	$X_1, \dots, X_{900}$ a.a. <i>Log - gama</i> ( $\theta$ , 8)	$\theta \sim Gama(4, 3)$

**Tabela B.10:** Conjunto de modelos candidatos para o exemplo de dados reais.

Modelos	Distribuição das variáveis observáveis	Distribuição a priori de $\theta$
$M_1$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (10, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_2$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (20, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_3$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (30, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_4$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (40, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_5$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (50, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_6$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (60, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_7$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (70, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_8$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (80, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_9$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (90, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{10}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (100, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{11}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (110, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{12}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (120, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{13}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (130, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{14}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (140, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{15}$	$X_1, \dots, X_{85}$ a.a. <i>Pareto</i> (150, $\theta$ )	$\theta \sim \text{Gama}(8, 3)$
$M_{16}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 30)	$\theta \sim \text{Gama}(8, 3)$
$M_{17}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 31)	$\theta \sim \text{Gama}(8, 3)$
$M_{18}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 32)	$\theta \sim \text{Gama}(8, 3)$
$M_{19}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 33)	$\theta \sim \text{Gama}(8, 3)$
$M_{20}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 34)	$\theta \sim \text{Gama}(8, 3)$
$M_{21}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 35)	$\theta \sim \text{Gama}(8, 3)$
$M_{22}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 36)	$\theta \sim \text{Gama}(8, 3)$
$M_{23}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 37)	$\theta \sim \text{Gama}(8, 3)$
$M_{24}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 38)	$\theta \sim \text{Gama}(8, 3)$
$M_{25}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 39)	$\theta \sim \text{Gama}(8, 3)$
$M_{26}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 40)	$\theta \sim \text{Gama}(8, 3)$
$M_{27}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 41)	$\theta \sim \text{Gama}(8, 3)$
$M_{28}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 42)	$\theta \sim \text{Gama}(8, 3)$
$M_{29}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 43)	$\theta \sim \text{Gama}(8, 3)$
$M_{30}$	$X_1, \dots, X_{85}$ a.a. <i>Log - gama</i> ( $\theta$ , 44)	$\theta \sim \text{Gama}(8, 3)$