



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE ESTATÍSTICA

CAROLINE TENÓRIO MENDES DE AQUINO

RESOLUÇÃO DE CONFLITOS DE INFORMAÇÃO UTILIZANDO A
DISTRIBUIÇÃO FLOOR

FORTALEZA, CEARÁ

2014

CAROLINE TENÓRIO MENDES DE AQUINO

**RESOLUÇÃO DE CONFLITOS DE INFORMAÇÃO UTILIZANDO A
DISTRIBUIÇÃO FLOOR**

Monografia submetida à Coordenação do Curso de Graduação em Estatística da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. José Ailton Alencar
Andrade

FORTALEZA, CEARÁ

2014

CAROLINE TENÓRIO MENDES DE AQUINO

**RESOLUÇÃO DE CONFLITOS DE INFORMAÇÃO UTILIZANDO A
DISTRIBUIÇÃO FLOOR**

Aprovada em: __/__/____

BANCA EXAMINADORA

Prof. Dr. José Ailton Alencar Andrade
Universidade Federal do Ceará - UFC
Orientador

Prof. Dr. João Maurício Araújo Mota
Universidade Federal do Ceará - UFC

Prof. Dr. Rafael Bráz Azevedo Farias
Universidade Federal do Ceará - UFC

Dedico este trabalho aos meus pais,
Francisco José e Ana Vladia. Por
acreditarem em mim e me darem forca
todos os dias para seguir em frente.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à minha família: meus pais Francisco José e Ana Vlândia, e meus irmãos Catherine e Bruno. O apoio dessas pessoas tão importantes na minha vida foi fundamental ao longo deste trabalho.

Ao professor José Ailton, que muito acrescentou para minha formação acadêmica. Pelas inúmeras conversas que contribuíram para meu crescimento, ao longo dos últimos anos, e por ter depositado sua confiança em mim para a realização deste trabalho.

A todos os professores do Departamento de Estatística e Matemática Aplicada, em especial aqueles que mais despertaram meu interesse em estudar e crescer academicamente: Professores João Maurício e Juvêncio Santos Nobre. Serei sempre grata por todo o empenho e dedicação que estes tiveram em minha formação.

Um agradecimento especial ao professor Julio Barros, por ter sido para mim um orientador desde o dia de matrícula no curso de estatística. Por ser um exemplo de dedicação aos alunos e se preocupar com sua formação acadêmica e profissional.

Aos meus amigos e colegas de turma: Eduardo Cândido, Hellano, José Eduardo, Jucelino, Rodney e Widemberg, por serem meus companheiros nos últimos quatro anos. Obrigada pelas conversas, horas de estudo e momentos de distração.

A todos aqueles que contribuíram direta ou indiretamente para a produção deste trabalho, agradeço imensamente.

“A vida é uma tempestade, meu jovem amigo. Deleitar-se-á ao sol, num momento, para ser arremessado contra as rochas, no momento seguinte. O que fará de si um homem é como reagirá, chegada a tempestade. Deverá olhar para ela e gritar, como fez em Roma: ‘Façam o vosso pior... pois eu farei o mesmo’. Então, o destino conhecê-lo-á, tal como nós o conhecemos”

(O Conde de Monte Cristo - Alexandre Dumas)

RESUMO

Do ponto de vista bayesiano, duas fontes de informação são utilizadas para estudar um fenômeno de interesse: a distribuição a priori e a verossimilhança. Com base nessas duas fontes de informação, é possível utilizar o teorema de Bayes para obter a distribuição a posteriori, e, a partir dela, realizar inferências sobre um fenômeno de interesse. Os modelos representam apenas uma imagem da realidade, e por isso é inevitável o aparecimento de erros. A má especificação de um modelo pode acontecer de diversas formas; entre elas está a presença de *outliers* na amostra. Quando os dados que estão sendo estudados são perturbados por *outliers*, a distribuição a priori poderá entrar em conflito com a distribuição dos dados, gerando assim um conflito de informações. Neste trabalho, estaremos interessados no problema de conflito de informações quando há presença de *outliers*. Na literatura, foi visto que conflitos de informações geralmente são resolvidos fazendo-se uso de densidades com caudas pesadas. Para estudar distribuições de caudas pesadas, as classes de distribuição de variação regular e \mathcal{O} -regular foram utilizadas. Neste trabalho, vimos que essas classes de distribuições com caudas pesadas podem contribuir para obter uma modelagem robusta. O objetivo principal desta monografia é discutir o uso de distribuições de caudas pesadas em problemas de conflito de informações usando a teoria de variação regular e \mathcal{O} -regular, especificadamente com o uso da distribuição floor. Um estudo de simulação ao final deste trabalho mostrou que a distribuição floor é robusta a *outliers*, comparando-a com a distribuição exponencial.

Palavras-chave: Caudas pesadas. Modelagem robusta bayesiana. Variação Regular.

ABSTRACT

From the Bayesian point of view, two sources of information are used to study a phenomenon of interest: the prior distribution and the likelihood. Based on these two sources of information, it is possible to use Bayes' theorem to obtain the posterior distribution, and, from it, make inferences about a phenomenon of interest. The models represent only a picture of reality, and so it is inevitable the occurrence of errors. The misspecification of a model can happen in several ways; among them is the presence of outliers in the sample. When the data being studied are disturbed by outliers, the prior distribution may conflict with the data distribution, thus creating a conflict of information. In this work, we are interested in the problem of conflicting information when there is presence of outliers. In the literature, it was seen that conflict of information are usually resolved by making use of densities with heavy tails. To study heavy-tailed distributions, the concepts of regular and \mathcal{O} -regular variation were used. In this study, we saw that these classes of heavy-tailed distributions can contribute to obtain a robust modeling. The main objective of this paper is to discuss the use of heavy-tailed distributions in problems of conflicting information using the theory of regular and \mathcal{O} -regular variation, specifically with the use of the floor distribution. A simulation study at the end of this study showed that the floor distribution is robust to outliers, when comparing it to the exponential distribution.

Keywords: Bayesian robustness modelling . Heavy tails. Regular variation.

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Motivação	10
1.2	Objetivos	10
1.3	Revisão de literatura	10
2	FUNDAMENTOS DA ESTATÍSTICA BAYESIANA	12
2.1	Priori, verossimilhança e posteriori	12
2.2	Teorema de Bayes	13
2.3	Função de verossimilhança padronizada	13
2.4	Uso sequencial do teorema de Bayes	13
2.5	Distribuição preditiva	14
3	CAUDAS PESADAS E VARIAÇÃO REGULAR	15
3.1	Variação regular	15
3.2	Variação regular e caudas pesadas	17
3.3	Variação θ-regular	17
4	RESOLUÇÃO DE CONFLITOS DE INFORMAÇÃO	20
4.1	Condições de Dawid	20
4.2	Estrutura com parâmetro de escala	21
4.2.1	Rejeição parcial da informação dos dados	21
4.2.2	Muitas observações	22
4.3	Rejeição no caso de variação θ-regular	23
5	A DISTRIBUIÇÃO FLOOR	25
5.1	Densidade	25
5.2	Constante normalizadora	26
5.3	Função de distribuição acumulada	26
5.4	Momento de ordem r	27
5.5	Variância	27
5.6	Coeficiente de assimetria	27
5.7	Excesso de Curtose	28

5.8	Geração de valores aleatórios	28
5.9	A distribuição floor e variação \mathcal{O}-regular	29
6	APLICAÇÃO	31
7	CONCLUSÕES	33
	REFERÊNCIAS BIBLIOGRÁFICAS	34
	APÊNDICE A – FUNÇÕES NO R PARA A DISTRIBUIÇÃO FLOOR	37
	APÊNDICE B – MODELOS NO OPENBUGS	40
	ANEXO A – DADOS	41

1 INTRODUÇÃO

1.1 Motivação

Em estatística bayesiana, duas fontes de informação são utilizadas para estudar um fenômeno de interesse: a distribuição a priori, representando o conhecimento prévio sobre o problema; e a verossimilhança, representando a informação dos dados. Com base na distribuição a priori e na verossimilhança, é possível utilizar o teorema de Bayes para obter a distribuição a posteriori, e, a partir dela, realizar inferências sobre um fenômeno de interesse.

Como os modelos representam apenas uma imagem da realidade, é inevitável o aparecimento de erros. A má especificação de um modelo pode acontecer de diversas formas; entre elas está a presença de *outliers* na amostra. Quando os dados que estão sendo estudados são perturbados por *outliers*, a distribuição a priori poderá entrar em conflito com a distribuição dos dados. E, como ambas distribuições são componentes da distribuição a posteriori, de onde todas as inferências sobre o parâmetro de interesse são realizadas, pode haver, então, um conflito de informações.

De acordo com Santos (2007), para evitar o problema de conflito de informações, alguns critérios podem ser estabelecidos. Por exemplo, se um fenômeno de interesse puder assumir apenas valores positivos, o pesquisador deverá levar o suporte da distribuição a priori em consideração, respeitando o que já é conhecido sobre o fenômeno. Porém, nem sempre as fontes de informação irão concordar entre si e então haverá um conflito de informações. Isso pode ser verificado visualmente pela distância entre as duas fontes de informação, isto é, a distribuição a priori e a verossimilhança.

Neste trabalho, estaremos interessados no problema de conflito de informações quando há presença de *outliers*. Segundo Desgagné (2013), o conflito de informações geralmente é resolvido fazendo-se uso de densidades com caudas suficientemente pesadas. Assim, neste trabalho também será discutida a relação entre distribuições de caudas pesadas e conflito de informações.

1.2 Objetivos

O objetivo principal desta monografia é discutir o uso de distribuições de caudas pesadas em problemas de conflito de informações usando a teoria de variação regular e \mathcal{O} -regular, especificadamente com o uso da distribuição floor.

1.3 Revisão de literatura

Para uma revisão de literatura extensa, consultar O'Hagan e Pericchi (2012). Muitos estudos abordam o tema de conflito de informações. Dentre eles, o seguinte estudo é destacado: de Finetti (1961) enfatiza que no contexto bayesiano observações não devem ser descartadas,

mesmo se tratando de *outliers*. Em casos em que o procedimento de “rejeitar algumas observações” é adotado e justificado, o método bayesiano leva automaticamente a resultados similares em que a influência dessas observações é fraca ou insignificante.

Um procedimento chamado de “método bayesiano de rejeição de *outliers*” foi descrito por de Finetti (1961), que afirma que, se observações possuem variâncias desconhecidas, diferentes e independentes, então um *outlier* teria um erro grande na estimativa, e então possuiria menos peso. Quanto maior o conflito, i.e., quanto mais distante for o *outlier* do restante das observações, menor peso teria. No limite, se o *outlier* ficasse infinitamente distante das outras observações, possuiria peso zero. No entanto, de Finetti não considerou casos específicos.

Lindley (1968), respondendo à comentários de E.M.L. Beale, sugere que o uso da distribuição t como priori seria mais robusto a *outliers* em comparação com a distribuição normal. A partir da sugestão de Lindley, Dawid (1973) estabeleceu condições para estudar o comportamento da distribuição a posteriori quando $x \rightarrow \infty$, quando a densidade de $x|\theta$ possui parâmetro de locação. O teorema de Dawid dá condições na densidade $f(x|\theta)$ e na distribuição a priori que asseguram que a esperança da função a posteriori tende para sua esperança a priori quando $x \rightarrow \infty$.

Segundo Neyman e Scott (1971 apud O’HAGAN, 1979, p. 358),

Há situações em que os *outliers* não devem ser tratados como aberrantes ou sem sentido, mas como um processo natural do processo de geração de dados. Amostras de distribuições com caudas pesadas são passíveis de terem observações extremas, porém esses aparentes *outliers* deveriam carregar menor peso do que observações moderadas. Rejeitá-los completamente seria inapropriado. Neyman e Scott (1971) introduziram o termo “propenso a *outliers*” para distribuições que tendem a revelar valores extremos, enquanto outras distribuições são chamadas de “resistentes a *outliers*”. (O’HAGAN, 1979, p. 358, tradução livre.)

O’Hagan discute neste artigo as definições de distribuições propensas e resistentes a *outliers*, em termos do comportamento da distribuição a posteriori dada uma priori arbitrária. O autor mostra que a distribuição t é propensa a *outliers* e a distribuição normal é resistente a *outliers*. O’Hagan também sugeriu condições alternativas que são mais fáceis de verificar do que as apresentadas por Dawid (1973). O’Hagan (1990) introduz o conceito e propriedades da medida *credence*, para o estudo de parâmetros de locação no caso de variâncias conhecidas.

Andrade e O’Hagan (2006) usam uma classe de distribuições de caudas pesadas diferente das consideradas por Dawid (1973) e O’Hagan (1979): a classe de *distribuições de variação regular*, para resolver conflitos de informação quando o parâmetro pertence ao caso univariado de locação-escala. O estudo de robustez quando o parâmetro pertence à família de escala foi considerado pela primeira vez nesse estudo.

2 FUNDAMENTOS DA ESTATÍSTICA BAYESIANA

Neste capítulo, noções de inferência bayesiana necessárias ao desenvolvimento deste trabalho serão apresentadas.

Os dois paradigmas mais importantes encontrados na estatística são o paradigma clássico ou frequentista e o paradigma bayesiano. A estatística bayesiana é uma importante alternativa à estatística clássica e será considerada com maior atenção neste trabalho.

De acordo com Bernardo (2003), há muitas vantagens ao se utilizar métodos bayesianos, como: métodos bayesianos podem ser derivados de um sistema axiomático, possibilitando assim uma metodologia coerente; muitos procedimentos frequentistas comuns são casos particulares de métodos bayesianos; através da distribuição a priori, a estatística bayesiana torna possível a incorporação de hipóteses na análise, que pode ser implementada em problemas complexos; a inferência estatística é descrita como a modificação da incerteza sobre o valor de um parâmetro à luz de evidências, e o teorema de Bayes mostra como essa modificação deve ser feita.

Referências sobre a história da estatística bayesiana podem ser encontradas em Stigler (1983), com um estudo sobre a origem do teorema de Bayes, e Leonard (2014), com um estudo vasto sobre as contribuições de diversos pesquisadores da área, seguindo uma ordem cronológica. Os seguintes conceitos e resultados já estão bem estabelecidos na literatura, e serão usados como fonte os livros de Box e Tiao (1973) e Lee (2012).

2.1 Priori, verossimilhança e posteriori

Ao considerar-se um vetor de quantidades de interesse, denotado por

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k),$$

atribui-se uma densidade de probabilidade para representar os possíveis valores que tal vetor de variáveis pode assumir. No início, tem-se a distribuição a priori, representada por

$$p(\boldsymbol{\theta}),$$

obtida geralmente através de opinião de profissional da área ou por meio de estudos anteriores. Pode-se obter uma amostra $\mathbf{x} = (x_1, x_2, \dots, x_n)$, que possui uma distribuição que depende dos k parâmetros, de forma que a densidade de probabilidade de \mathbf{x} depende do vetor $\boldsymbol{\theta}$. Assim, a dependência de \mathbf{x} pode ser escrita como

$$p(\mathbf{x}|\boldsymbol{\theta}),$$

que pode ser interpretada como a forma em que os dados estão relacionados com $\boldsymbol{\theta}$. Finalmente, através do teorema de Bayes, pode-se chegar à seguinte conclusão, em que \propto é um sinal de

proporção:

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta}).$$

Sabe-se que, se $p(\mathbf{x}|\boldsymbol{\theta})$ é considerada como uma função de \mathbf{x} para parâmetros $\boldsymbol{\theta}$ fixos, essa função é uma densidade. Porém, muitas vezes podemos achar interessante considerar essa função como de $\boldsymbol{\theta}$ para \mathbf{x} fixo. Com essa interpretação, segundo Fisher (1922), essa função é chamada de *função de verossimilhança* e é denotada por $l(\boldsymbol{\theta}|\mathbf{x})$. É importante notar que as propriedades não serão as mesmas. Por exemplo, não há mais razão para que a soma (ou integral) seja igual a um. No contexto bayesiano, a função de verossimilhança tem um papel muito importante, porque é através dessa função que a informação sobre $\boldsymbol{\theta}$ através dos dados é considerada. Agora pode-se escrever de forma simplificada o resultado acima.

$$\text{posteriori} \propto \text{priori} \times \text{verossimilhança}$$

2.2 Teorema de Bayes

A relação entre distribuição a posteriori, distribuição a priori e a função de verossimilhança é mostrada na Equação (2.1):

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{x})}, \quad (2.1)$$

em que $p(\boldsymbol{\theta}|\mathbf{x})$ é a distribuição a posteriori, $p(\boldsymbol{\theta})$ é a distribuição a priori, $p(\mathbf{x}|\boldsymbol{\theta})$ é a densidade dos dados, que está relacionada com a função de verossimilhança, e $p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ é a constante normalizadora. No caso discreto, pode-se trocar a integral em $p(\mathbf{x})$ por uma soma. A expressão acima é chamada de *teorema de Bayes*. A constante normalizadora assegura que a distribuição a posteriori integrará (ou somará) um.

2.3 Função de verossimilhança padronizada

É importante notar que a função de verossimilhança pode ser multiplicada por uma constante ou qualquer função que dependa apenas de \mathbf{x} sem alterar a distribuição a posteriori, por causa da forma que se escreve o teorema de Bayes com o sinal de proporcionalidade. Quando a integral $\int l(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$ é finita, a quantidade seguinte é chamada de *verossimilhança padronizada*, cuja integral corresponde à unidade.

$$\frac{l(\boldsymbol{\theta}|\mathbf{x})}{\int l(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}}$$

2.4 Uso sequencial do teorema de Bayes

Uma das grandes vantagens do uso do teorema de Bayes está em não apenas combinar conhecimentos prévios e recentes, mas também em atualizar sempre que possível o modelo,

fazendo uso sequencial do teorema de Bayes. Começando com uma amostra \mathbf{x} , tem-se através da Equação (2.1) que

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x})$$

Supondo que outra amostra, denotada por \mathbf{y} e independente da anterior, esteja disponível. Assim

$$p(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) \propto p(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})$$

Usando a suposição de independência entre as amostras, tem-se que

$$p(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$$

Sabendo-se que a função de verossimilhança pode ser vista como qualquer múltiplo constante de $p(\mathbf{x}|\boldsymbol{\theta})$,

$$l(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) \propto l(\boldsymbol{\theta}|\mathbf{x})l(\boldsymbol{\theta}|\mathbf{y})$$

E assim

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) &\propto p(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x})l(\boldsymbol{\theta}|\mathbf{y}) \\ &\propto p(\boldsymbol{\theta}|\mathbf{x})l(\boldsymbol{\theta}|\mathbf{y}). \end{aligned}$$

De forma que a distribuição a posteriori da observação \mathbf{x} é depois tratada como priori da observação \mathbf{y} .

2.5 Distribuição preditiva

Em alguns casos, a seguinte distribuição marginal é considerada:

$$p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Essa distribuição é chamada de distribuição preditiva e é utilizada em teoria da decisão bayesiana e em métodos empíricos bayesianos, e representa predições sobre o valor de \mathbf{x} levando em consideração tanto a incerteza sobre o valor de $\boldsymbol{\theta}$ e a incerteza residual sobre \mathbf{x} quando $\boldsymbol{\theta}$ é conhecido (LEE, 2012).

3 CAUDAS PESADAS E VARIAÇÃO REGULAR

Não existe uma definição clara sobre o que é uma distribuição de cauda pesada, sendo importante considerar o contexto do problema. Várias classes de caudas pesadas foram propostas na literatura, destacando-se a classe de cauda longa, a classe de variação dominada, a classe subexponencial e a classe de variação regular. Neste trabalho a última classe será analisada e estudada.

3.1 Variação regular

O estudo de variação regular foi introduzido por Karamata (1930). Feller (1971) dedicou-se às aplicações dessa teoria em probabilidade e Andrade e O'Hagan (2006 e 2011) utilizaram distribuições de variação regular em modelagem robusta bayesiana. Outros estudos sobre variação regular encontram-se em Landau (1911), Valiron (1913), Pólya (1917) - estudos preliminares realizados antes de Karamata (1930) - de Haan (1970) e Seneta (1976). A principal referência sobre variação regular utilizada neste trabalho é o livro de Bingham, Goldie e Teugels (1987), que chamaremos de BGT para fins de simplificação. A definição de função de variação regular é dada por:

Definição 3.1.1. (Variação Regular) Uma função mensurável f é de variação regular no infinito com índice ou ordem ρ se $\rho \in \mathbb{R}$ e se $\forall \lambda > 0$ tem-se que

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho \quad (3.1)$$

e escreve-se $f \in R_\rho$.

Um caso particular ocorre quando $\rho = 0$. Nesse caso, f é dita ser de variação lenta e escreve-se $f \in R_0$. Dessa forma, pode-se escrever o conjunto de todas as funções de variação regular como $R = \{R_\rho : \rho \in (-\infty, \infty)\}$.

Estendendo a definição acima para estudos do comportamento da cauda esquerda, pode-se dizer que f tem variação regular em $-\infty$ com índice ρ se $\lim_{x \rightarrow -\infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho$; e escreve-se $f \in R_\rho(-\infty)$. Se $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, pode-se dizer que f possui variação regular na origem com índice ρ se $\lim_{x \rightarrow 0} \frac{f(\lambda x)}{f(x)} = \lambda^\rho$; e escreve-se $f \in R_\rho(0+)$ (ANDRADE; O'HAGAN, 2006).

Teorema 3.1.1. TEOREMA DE CARACTERIZAÇÃO. Se $f > 0$ for mensurável tal que $\frac{f(\lambda x)}{f(x)} \rightarrow g(\lambda) \in \mathbb{R}^+$ quando $x \rightarrow \infty$ é verdade para todo λ em um conjunto de medida positiva, então

1. $\frac{f(\lambda x)}{f(x)} \rightarrow g(\lambda) \in \mathbb{R}^+$ quando $x \rightarrow \infty$ é verdade para todo $\lambda > 0$;
2. Há um número real ρ tal que $g(\lambda) \equiv \lambda^\rho \forall \lambda > 0$;
3. $f(x) = x^\rho \ell(x)$ com ℓ tendo variação lenta.

Prova: ver BGT (1987, p.17)

Em suma, o Teorema 3.1.1 diz que se f é uma função que pertence à classe de variação regular, então pode ser escrita como x^ρ vezes uma função ℓ que possui variação lenta.

A Tabela 3.1 mostra exemplos de distribuições que pertencem à classe de variação regular:

Tabela 3.1: Exemplos de distribuições de variação regular

Nome	Função densidade de probabilidade	Suporte e parâmetros
Burr	$f(x) = \frac{\kappa \eta x^{\eta-1}}{(1+x^\eta)^{\kappa+1}}$	$x > 0, \eta > 0, \kappa > 0$
Cauchy	$f(x) = \frac{1}{\pi(1+x^2)}$	$-\infty < x < \infty$
Log-gama	$f(x) = \frac{\log(x)^{\alpha-1} e^{-\log(x)/\beta}}{x\beta^\alpha \Gamma(\alpha)}$	$x \geq 0, \alpha > 0, \beta > 0$
Pareto	$f(x) = \frac{\delta \gamma^\delta}{x^{\delta+1}}$	$\gamma \leq x < \infty, \gamma > 0, \delta > 0$

A distribuição de Pareto, por exemplo, possui uma propriedade interessante, chamada de “lei de Pareto”. Após estudar intensivamente a renda de diversas populações, o economista social Vilfredo Pareto (1848-1923) observou que não importava a população ou o lugar: a distribuição de renda em uma população em uma economia estável seguia a fórmula empírica $y \propto x^{-\nu}$, em que y representa o número de pessoas com renda igual ou superior a x e $\nu \cong 1,5$ (PARETO, 1897; DAVIS, 1941). Para Pareto, esse comportamento não poderia ser devido ao acaso, e não deveria depender fortemente de variáveis econômicas, pois lugares tão diferentes como Inglaterra, Peru, Alemanha, Irlanda e algumas cidades italianas obedecia a mesma lei. Além das aplicações em dados de distribuição de renda, a distribuição de Pareto também tem aplicações em diversas outras fontes de dados (NEWMAN, 2006), como frequência de palavras, diâmetro de crateras lunares e frequência de nomes de família.

Outras propriedades de funções de variação regular são apresentadas (BGT, 1987, p. 26).

- (i) Se $f \in R_\rho$, então $f^\alpha(x) \in R_{\rho\alpha}$, em que $\rho, \alpha \in \mathbb{R}$;
- (ii) Sejam $f_1 \in R_{\rho_1}$ e $f_2 \in R_{\rho_2}$. Pode-se mostrar que $f_1 + f_2 \in R_{\max\{\rho_1, \rho_2\}}$ e $f_1 f_2 \in R_{\rho_1 + \rho_2}$;
- (iii) Se $f_i \in R_{\rho_i}$ ($i = 1, 2$), $f_2(x) \rightarrow \infty$ quando $x \rightarrow \infty$, então $f_1(f_2(x)) \in R_{\rho_1 \rho_2}$
- (iv) Se $f_i \in R_{\rho_i}$ ($i = 1, \dots, k$), e $r(x_1, \dots, x_k)$ é uma função racional com coeficientes positivos, então $r(f_1(x), \dots, f_k(x)) \in R$.

Mais propriedades de variação regular podem também ser encontradas em BGT (1987) e Andrade (2005).

Funções que não satisfazem a Definição 3.1.1 são chamadas de funções de variação rápida (DE HAAN, 1970). São funções cujo limite na Equação (3.1) é infinito, isto é, decrescem ou crescem mais rapidamente do que qualquer função potência de ordem finita. Assim, tem-se a definição (BGT, 1987, p. 83):

Definição 3.1.2. (Variação Rápida) Uma função mensurável $f : (0, \infty) \rightarrow (0, \infty)$ é de variação rápida com índice ou ordem ∞ se

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \begin{cases} 0, & \text{se } 0 < \lambda < 1 \\ 1, & \text{se } \lambda = 1 \\ \infty, & \text{se } \lambda > 1 \end{cases} \quad (3.2)$$

e escreve-se $f \in R_{\infty}$. Também tem-se que f é de variação rápida com índice $-\infty$ se

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \begin{cases} \infty, & \text{se } 0 < \lambda < 1 \\ 1, & \text{se } \lambda = 1 \\ 0, & \text{se } \lambda > 1 \end{cases} \quad (3.3)$$

e escreve-se $f \in R_{-\infty}$.

Note que as Equações (3.2) e (3.3) representam o limite de λ^{ρ} quando ρ tende a ∞ ou $-\infty$, respectivamente.

3.2 Variação regular e caudas pesadas

A definição 3.1.1, de função de variação regular, pode ser interpretada como o comportamento da cauda de uma densidade de probabilidade, isto é, há uma relação entre variação regular e o peso das caudas. Por exemplo, se $f \in R_{-\rho}$ quando x tende a infinito, isso quer dizer que a cauda direita decresce com taxa ρ , ou em outras palavras, como $\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^{-\rho}$, conforme x cresce, a cauda direita da função decresce como uma potência de ordem ρ .

Neste trabalho, uma nova distribuição que pertence à família de distribuições com variação \mathcal{O} -regular será estudada. A variação \mathcal{O} -regular é uma generalização da variação regular e será apresentada a seguir.

3.3 Variação \mathcal{O} -regular

A classe de variação \mathcal{O} -regular (que chamaremos de ORV) estende o conceito de variação regular e foi introduzida por Avakumović (1936). Essa classe envolve distribuições cujas caudas decrescem em qualquer comportamento entre duas funções de variação regular. Por exemplo, as caudas de uma distribuição ORV podem decrescer oscilando entre duas funções potência (ANDRADE; OMEY, 2014).

As funções de variação \mathcal{O} -regular satisfazem a relação (DJURČIĆ, 1998):

$$0 < \liminf_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} \leq \limsup_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} < \infty,$$

para todo $\lambda \geq 1$.

Definição 3.3.1. (Variação \mathcal{O} -regular) Uma função mensurável f possui variação \mathcal{O} -regular no infinito, se f satisfaz

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} < \infty, \forall \lambda > 0,$$

e escreve-se $f \in \text{ORV}$.

Juntamente com a definição de função de variação \mathcal{O} -regular, os *índices de Matuszewska* são definidos. Se $f \in \text{ORV}$, o índice superior de f é dado por

$$\alpha(f) = \lim_{\lambda \rightarrow \infty} \frac{\log \limsup_{x \rightarrow \infty} f(\lambda x)/f(x)}{\log(\lambda)},$$

e o índice inferior de f é dado por

$$\beta(f) = \alpha(1/f) = \lim_{\lambda \rightarrow \infty} \frac{\log \liminf_{x \rightarrow \infty} f(\lambda x)/f(x)}{\log(\lambda)}.$$

Se $f \in \text{ORV}$, então para qualquer $\beta < \beta(f)$ e $\alpha > \alpha(f)$, há constantes C, D e x° de forma que (BGT, p. 65):

$$C \lambda^\beta \leq \frac{f(\lambda x)}{f(x)} \leq D \lambda^\alpha, \forall \lambda \geq 1, \forall x \geq x^\circ$$

O Teorema 3.3.1, apresentado a seguir, é provado em Aljančić e Arandelović (1977):

Teorema 3.3.1. TEOREMA DE CARACTERIZAÇÃO PARA VARIAÇÃO \mathcal{O} -REGULAR. *Uma função $f \in \text{ORV}$ se e somente se pode ser escrita como*

$$f(x) = \exp \left\{ \mu(x) + \int_B^x \frac{\varepsilon(t)}{t} dt \right\}, x \geq B,$$

em que μ e ε são funções reais mensuráveis limitadas no intervalo $[B, \infty)$, $B > 0$.

Uma propriedade interessante de funções ORV é que uma função $f \in \text{ORV}$ se e somente se $f(\lambda x) \asymp f(x)$, para todo $\lambda > 0$ e $x \rightarrow \infty$, isto é, existem dois números $0 < m < M < \infty$ de forma que $m \leq f(\lambda x)/f(x) \leq M$, para x suficientemente grande (ALJANČIĆ; ARANDELOVIĆ, 1977).

Como as caudas de distribuições ORV decrescem entre dois polinômios, essa classe é bastante abrangente, e assim, há diversas formas de construir distribuições ORV. Andrade e Omeij (2014) sugerem uma maneira de criar distribuições de variação \mathcal{O} -regular, apresentada no Teorema 3.3.2

Teorema 3.3.2. CONDIÇÕES SUFICIENTES PARA ORV (ANDRADE; OMEY, 2014) *Seja f uma função densidade de probabilidade da forma*

$$f(x) = C b(x) A(x), \quad (3.4)$$

tal que C é a constante normalizadora, b é delimitado de zero a infinito quando x tende a infinito.

(i) *Se $x A'(x)/A(x)$ for limitada, então $f \in ORV$.*

(ii) *Se $A(x) \in RV_{-\alpha}$, então $f \in ORV$.*

Demonstração. Ver Andrade e OmeY (2014), Teorema 5. ■

No capítulo seguinte, condições serão dadas para que a resolução de conflito de informações possa ser feita através do uso de distribuições com variação regular e \mathcal{O} -regular.

4 RESOLUÇÃO DE CONFLITOS DE INFORMAÇÃO

Na literatura, a distribuição t -Student foi largamente utilizada na resolução de informações conflitantes. Contudo, os resultados de Dawid (1973) e estudos posteriores não assumiram distribuições específicas, contanto que tivessem certas propriedades de caudas pesadas. Neste capítulo introduziremos essas propriedades e daremos maior atenção à classe de distribuições com variação regular, em particular, a classe de variação \mathcal{O} -regular.

4.1 Condições de Dawid

Em seu estudo de 1973, Dawid estabeleceu condições para que a distribuição a posteriori se aproximasse da distribuição a priori, assintoticamente, quando $x \rightarrow \infty$. Ele considerou o caso de uma única observação $X = x$ de uma distribuição com parâmetro de locação desconhecido Θ e aplicou as condições quando a distribuição a priori é normal e o erro segue uma distribuição t de Student. Para entender melhor o problema de como informações conflitantes podem trazer consequências à distribuição a posteriori, Dawid considerou o caso em que tem-se o seguinte modelo ¹:

$$\begin{cases} X|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2) \\ \mu \sim \mathcal{N}(m_0, \sigma_0^2) \end{cases}$$

Sabe-se que a esperança a posteriori para este modelo é dada por:

$$\mathbb{E}(\mu|x) = x - \frac{\sigma^2}{\sigma^2 + \sigma_0^2} (x - m_0),$$

assim, a diferença entre $\mathbb{E}(\mu|x)$ e x é dada por um múltiplo de $(x - m_0)$, a diferença entre o valor observado e a média a priori. Quanto maior a diferença entre x e m_0 , maior será a diferença entre $\mathbb{E}(\mu|x)$ e x , i.e., a estimativa para μ dada pelo valor médio da distribuição a posteriori e a verossimilhança, representada pela observação nova x . As condições de Dawid são mostradas no Teorema 4.1.1.

Teorema 4.1.1. TEOREMA DE DAWID (1973). *Sejam $Y|\mu$ e μ variáveis aleatórias independentes absolutamente contínuas com densidades $f(y - \mu)$ e $p(\mu)$, respectivamente. A seguir, condições suficientes serão dadas para que $\mathbb{E}[p(\mu|y)] \rightarrow p(\mu)$ quando y tende a infinito.*

1. *Dados $\varepsilon > 0$, $\delta > 0$, existe A tal que, se $y > A$, então $|f(y') - f(y)| < \varepsilon f(y)$ sempre que $|y' - y| < \delta$;*
2. *Para algum B e M , $0 < f(y') < Mf(y)$ sempre que $y' > y > B$;*

¹Note que θ é o valor de um parâmetro Θ desconhecido, e poderíamos ter escrito $\Theta \sim \mathcal{N}(m_0, \sigma_0^2)$. Mesmo com a convenção de que usa-se letras maiúsculas para representar variáveis aleatórias e letras minúsculas seus valores observados, no contexto bayesiano isso se torna bastante restrito, pois trabalha-se com muitas variáveis aleatórias.

3. Seja $k(\mu) = \sup_y [f(y - \mu)/f(y)]$, então $\int_{\Theta} k(\mu) p(\mu) d\mu < \infty$.

De forma simplificada, podemos dizer que quando uma observação y tende a infinito, a distribuição a posteriori irá tender à distribuição a priori, se a densidade dos dados f tiver caudas pesadas e a distribuição a priori $p(\mu)$ tiver caudas mais leves do que f (ANDRADE, 2005).

A ideia central da resolução de conflito de informações, no contexto de modelagem robusta, é designar distribuições de caudas mais pesadas ou mais leves às fontes de informação, dependendo do grau de conflito que cada fonte causa. Por exemplo, no Teorema 4.1.1, se a distribuição a priori é a causa do conflito, então pode-se atribuir à essa fonte de informação uma distribuição com cauda pesada, e uma distribuição com cauda mais leve é atribuída à distribuição dos dados. Neste trabalho o interesse será dado a estruturas com parâmetro de escala.

4.2 Estrutura com parâmetro de escala

Uma forma de conflito de informação simples e fácil de ser verificada é a presença de *outliers*. Deve-se escolher uma distribuição adequada caso haja suspeita de que os dados de interesse possam ter *outliers*, seja pela natureza do fenômeno estudado ou pela forma que os dados são coletados.

4.2.1 Rejeição parcial da informação dos dados

Uma função densidade de probabilidade pertence à família de escala se pode ser escrita como

$$f(x) = \frac{1}{\theta} h\left(\frac{x}{\theta}\right),$$

em que θ é um parâmetro de escala. O Teorema a seguir, apresentado em ANDRADE (2005) dá condições suficientes para rejeitar a informação dos dados, isto é, favorecer a distribuição a priori.

Teorema 4.2.1. REJEIÇÃO PARCIAL DA INFORMAÇÃO DOS DADOS (ANDRADE, 2005) *Sejam $Y|\theta$ e θ com densidades $f(y|\theta) = 1/\theta h(y/\theta)$ e $p(\theta)$, respectivamente. Se para algum $\delta > 0$ e $\rho > 0$*

(i) $h \in R_{-(\rho+1)}$ conforme $y \rightarrow \infty$ e

(ii) $\int_{\Theta} \theta^{\rho+\delta} p(\theta) d\theta < \infty$,

então

$$\lim_{y \rightarrow \infty} p(\theta|y) = \frac{\theta^{\rho} p(\theta)}{\int_{\Theta} \theta^{\rho} p(\theta) d\theta} \quad (4.1)$$

Demonstração. Ver Andrade e O'Hagan (2006), Teorema 3.1, p. 175. ■

A primeira condição garante que a distribuição de y é de cauda pesada. A segunda condição diz que a distribuição a priori, $p(\theta)$, deve ter cauda mais leve do que a distribuição de y , além de ser uma condição suficiente para que a constante de proporcionalidade da distribuição a posteriori exista.

Em suma, o Teorema 4.2.1 diz que, se uma distribuição de variação regular for designada para y e uma distribuição com cauda mais leve for designada para a informação a priori, então conforme y se distancia da informação a priori, a distribuição a posteriori dará menos importância à informação de y , se tornando independente dela, resolvendo assim o conflito de informação.

No caso descrito por Dawid (1973), para o parâmetro de posição, a informação conflitante é rejeitada. Isso não acontece no caso do parâmetro de escala, pois mesmo que distribuição a posteriori limite seja independente da informação conflitante, não há rejeição completa do *outlier*. O numerador da distribuição a posteriori na Equação (4.1) é dado por $\theta^\rho p(\theta)$, e θ^ρ é a única influência que o *outlier* exerce sobre a distribuição a posteriori. De acordo com Andrade e O'Hagan (2006, p. 176), pode-se dizer que θ^ρ é uma distribuição imprópria que carrega pouca informação sobre os dados. Em outras palavras, também pode-se dizer que a distribuição dos dados se torna menos informativa.

4.2.2 Muitas observações

O Teorema 4.2.1 pode ser generalizado para se estudar o caso de muitas observações. Neste caso, um ou mais *outliers* podem ser verificados nos dados, de forma que as demais observações não entram em conflito com a distribuição a priori e nem com os outros dados, isto é, as informações restantes concordam entre si e também com a distribuição a priori.

Diferentemente da seção anterior, agora um grupo de *outliers* será (parcialmente) rejeitado, ao invés de apenas um. A generalização do Teorema 4.2.1, sob as mesmas condições, pode ser escrita como (ANDRADE; O'HAGAN, 2006):

$$p(\theta|\mathbf{x}) \approx \frac{\theta^\rho \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)}{\int_{\Theta} \theta^\rho \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta) d\theta}, \text{ para } x_n \text{ suficientemente grande,} \quad (4.2)$$

em que $\mathbf{x} = (x_1, \dots, x_n)$, é uma amostra aleatória independentemente distribuída com $f_i(x_i|\theta) = 1/\theta \times h_i(x_i|\theta)$, $i = 1, \dots, n$. Assim, a distribuição conjunta dos dados é dada por $f(\mathbf{x}|\theta) = \prod_{i=1}^n f_i(x_i|\theta)$. A distribuição a priori é dada por $p(\theta)$ e supõe-se que x_n é um *outlier*. Separando a densidade proveniente da informação conflitante, tem-se que

$$f(\mathbf{x}, \theta) = f_n(x_n|\theta) \times \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)$$

Note que essa é a configuração do Teorema 4.2.1, em que $f_n(x_n|\theta)$ é a verossimilhança que envolve o outlier x_n e $\prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)$ faz o papel de distribuição a priori, com

a informação concordante dos dados que não envolvem o outlier e a própria informação a priori sobre o parâmetro θ .

Assim, as condições do Teorema 4.2.1 serão satisfeitas se

- (i) $f_n(x_n|\theta) \in R_{-(\rho+1)}$ e
- (ii) $\int_{\Theta} \theta^{\rho+\delta} \prod_{i=1}^{n-1} f_i(x_i|\theta) p(\theta) d\theta < \infty$

A segunda condição dependerá do comportamento das caudas da distribuição a priori, $p(\theta)$, e das funções de verossimilhança, $f_i(x_i|\theta)$, para $i = 1, \dots, n-1$. Se as funções de verossimilhança ($i = 1, \dots, n-1$) ou $p(\theta)$ forem de caudas leves quando $\theta \rightarrow \infty$, pode-se verificar que a condição (ii) é automaticamente satisfeita (ANDRADE; O'HAGAN, 2006). Se as funções de verossimilhança forem de variação regular quando $\theta \rightarrow \infty$ com índices $-v_i \in \mathbb{R}^-$ e se $p(\theta) \in R_{-(c+1)}$, então a condição (ii) será satisfeita se $\rho + 1 < \sum v_i + (c + 1) + \delta$.

A distribuição a posteriori na Equação (4.2) também pode ser adaptada para o caso de um grupo de *outliers*. Com essa interpretação, um grupo de *outliers* será (parcialmente) rejeitado. Considere $X_i \sim f_i(x_i|\theta) \in R_{-(v_i+1)}$, $v_i > 0, i = 1, \dots, n$. Isto é, distribuições de variação regular com índice $-(v_i + 1)$ são atribuídas para cada observação. Também considere que $p(\theta) \notin R$ e (x_1, \dots, x_k) são *outliers*, com $k < n$. Assim,

$$f(\mathbf{x}|\theta) = \prod_{i=1}^k f_i(x_i|\theta) \times \prod_{i=k+1}^n f_i(x_i|\theta)$$

Usando os resultados do Teorema 4.2.1, fazendo os *outliers* tenderem a infinito e o papel de distribuição a priori sendo dado para $p(\theta|x_{k+1}, \dots, x_n) = \prod_{i=k+1}^n f_i(x_i|\theta) p(\theta) \notin R$. Assim, a combinação de $p(\theta|x_1, \dots, x_{n-k})$ e $\theta^{-(v_1+v_2+\dots+v_k+k)}$ resultará na distribuição a posteriori, dada por

$$p(\theta|\mathbf{x}) \approx \frac{\theta^{-(v_1+v_2+\dots+v_k+k)} \prod_{i=k+1}^n f_i(x_i|\theta) \times p(\theta)}{\int_{\Theta} \theta^{-(v_1+v_2+\dots+v_k+k)} \prod_{i=k+1}^n f_i(x_i|\theta) \times p(\theta) d\theta}$$

Nesta seção, mostramos como as observações conflitantes podem ser parcialmente rejeitadas usando-se o conceito de caudas pesadas através de variação regular, dentro do contexto de parâmetro de escala. Na próxima seção, o mesmo problema será abordado, porém com a verossimilhança tendo uma distribuição de variação \mathcal{O} -regular.

4.3 Rejeição no caso de variação \mathcal{O} -regular

Considere $\mathbf{x} = (x_1, \dots, x_n) | \theta \stackrel{i.i.d}{\sim} f(x|\theta) = 1/\theta h(x/\theta)$ e $\theta \sim p(\theta)$. Seguindo a notação presente em Andrade e Omey (2014), os dados serão particionados em dois conjuntos, chamados de \mathbf{x}^L e \mathbf{x}^U , definidos por $L = f(\mathbf{x}^L|\theta) = \prod_{i=1}^k h(x_i|\theta)$ e $U = f(\mathbf{x}^U|\theta) = \prod_{i=k+1}^n h(x_i|\theta)$,

em que \mathbf{x}^L são os *outliers*. Em outros termos, $f(\mathbf{x}|\theta) = \theta^{-n} \times L \times U$, e a distribuição a posteriori é dada por

$$p(\theta|\mathbf{x}) = \frac{\theta^{-n} \times L \times U \times p(\theta)}{\int_0^\infty \theta^{-n} \times L \times U \times p(\theta) d\theta}$$

Teorema 4.3.1. REJEIÇÃO DOS DADOS QUANDO HÁ DISTRIBUIÇÃO \mathcal{O} -REGULAR (ANDRADE; OMEY, 2014). *Considere $\mathbf{x} = (x_1, \dots, x_n)$ uma amostra aleatória proveniente de uma população com distribuição $f(x|\theta) = 1/\theta h(x/\theta)$, i.i.d., $\theta \sim p(\theta)$, e h e p densidades de probabilidade delimitadas e contínuas. Se as condições a seguir forem satisfeitas,*

- (i) $h \in ORV$ com índices $\alpha(h) < 0$;
- (ii) $\int_x^\infty \theta^{-n} p(\theta) d\theta = O(1) \prod_{i=1}^k x_i^{\beta-\varepsilon}$, onde $x = \min_{i \leq k}(x_i)$ e
- (iii) $\int_1^\infty \theta^{-n-k\beta} U p(\theta) d\theta < \infty$

então

$$0 < \liminf_{x \rightarrow \infty} \frac{p(\theta|\mathbf{x})}{U p(\theta)} \leq \limsup_{x \rightarrow \infty} \frac{p(\theta|\mathbf{x})}{U p(\theta)} < \infty \quad (4.3)$$

Demonstração. Ver Andrade e Omei (2014), Teorema 4. ■

O Teorema 4.3.1 estabelece que, conforme x tende a infinito, a distribuição a posteriori será limitada por duas quantidades independentes de x . Assim, da mesma forma como foi visto nas duas seções anteriores, a distribuição a posteriori terá como base a distribuição a priori e o restante das observações, isto é, aquelas que não são *outliers*.

5 A DISTRIBUIÇÃO FLOOR

A distribuição floor, proposta por Andrade e Omev (2013) é uma alternativa à distribuição exponencial. Sua vantagem em relação à distribuição exponencial está na cauda, que é pesada, o que é uma característica de robustez para informações conflitantes na modelagem bayesiana, como já foi visto ao longo desta monografia. A distribuição floor tem variação \mathcal{O} -regular, com índices de Matuszewska $\alpha(f) = \beta(f) = -a + 1$. Neste capítulo, as principais propriedades da distribuição floor são exibidas, como função de distribuição acumulada, momentos, coeficiente de assimetria e curtose; além da utilização do método de aceitação-rejeição para gerar valores aleatórios da distribuição.

5.1 Densidade

Dizemos que uma variável aleatória X tem distribuição floor com parâmetro a , com notação $X \sim \text{floor}(a)$, se sua função densidade de probabilidade (f.d.p.) é dada por

$$f(x) = C(a)x^{-a}e^{\lfloor \log x \rfloor}, \quad 1 \leq x < \infty, \quad (5.1)$$

em que $C(a)$ é a constante normalizadora (ver seção 5.3), $a > 2$ e $\lfloor \cdot \rfloor$ é a função piso. Se desejamos incluir parâmetros de posição e escala, a densidade de $Z = \sigma X + \mu$ torna-se

$$h(z) = C(a) \frac{1}{\sigma} \left(\frac{z - \mu}{\sigma} \right)^{-a} e^{\lfloor \log(\frac{z - \mu}{\sigma}) \rfloor}, \quad \mu + \sigma \leq z < \infty,$$

em que μ é o parâmetro de posição e σ é o parâmetro de escala. Quando os parâmetros de posição e escala não forem mencionados, refere-se à distribuição floor padrão, com $\mu = 0$ e $\sigma = 1$. A Figura 5.1 mostra alguns gráficos para diferentes valores de a , μ , e σ .

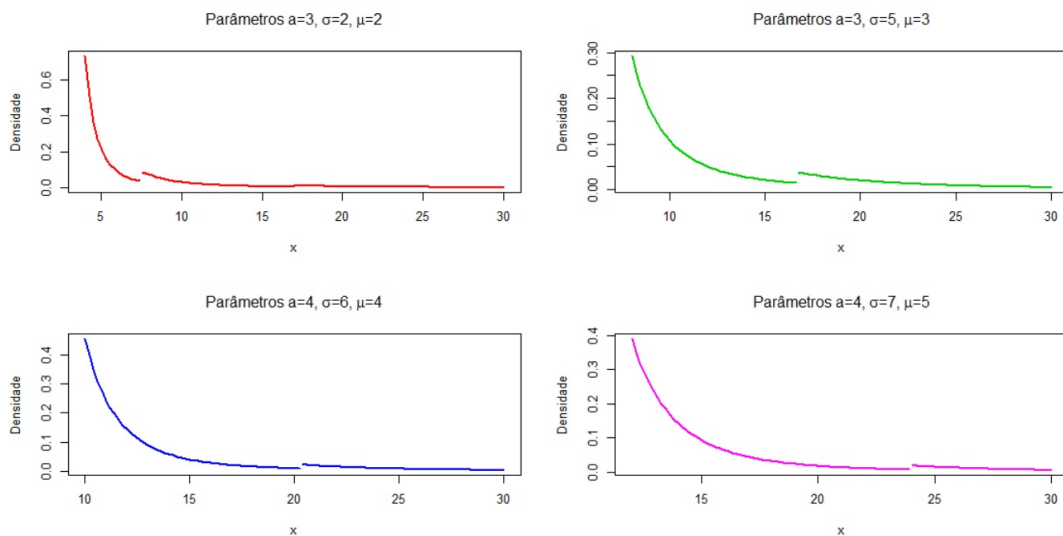


Figura 5.1: Densidades para diferentes valores de a , μ , e σ

5.2 Constante normalizadora

A constante normalizadora, $C(a)$, é dada por

$$C(a) = \frac{(a-1)(e^2 - e^a)}{e - e^a}, \quad a > 2.$$

Demonstração. Para que a Equação (5.1) seja uma legítima função densidade de probabilidade, é necessário que

$$\int_1^{\infty} C(a) x^{-a} e^{\lfloor \log x \rfloor} dx = 1 \quad (5.2)$$

Com a transformação $t = \log(x)$,

$$\begin{aligned} \frac{1}{C(a)} &= \int_0^{\infty} e^{-at} e^{\lfloor t \rfloor} e^t dt \\ &= \int_0^{\infty} e^{-t(a-1)} e^{\lfloor t \rfloor} dt \\ &= \int_0^1 e^{-t(a-1)} e^0 dt + \int_1^2 e^{-t(a-1)} e^1 dt + \dots \\ &= \sum_{n=0}^{\infty} e^n \int_n^{n+1} e^{-t(a-1)} dt \\ &= \frac{1}{a-1} \left[\sum_{n=0}^{\infty} e^{-n(a-2)} - \sum_{n=0}^{\infty} e^{-(n+1)(a-1)+n} \right] \end{aligned} \quad (5.3)$$

As duas somas da Equação (5.3) são séries geométricas com razão $e^{-(a-2)}$, de forma que

$$C(a) = \frac{(a-1)(e^2 - e^a)}{e - e^a}, \quad a > 2 \quad \blacksquare$$

5.3 Função de distribuição acumulada

A função de distribuição acumulada de $X \sim \text{floor}(a)$ é dada por:

$$F(x) = C(a) \int_1^x t^{-a} e^{\lfloor \log t \rfloor} dt$$

Fazendo uso da transformação $v = \log t$, tem-se que

$$F(x) = \frac{C(a)}{a-1} \left[\sum_{n=0}^{\lfloor \log x \rfloor} e^{-n(a-2)} - e \sum_{n=0}^{\lfloor \log x \rfloor} e^{-a(n+1)} \right] \quad (5.4)$$

As somas na Equação (5.4) são séries geométricas de razões $e^{-(a-2)}$ e e^{-a} , respectivamente. A soma parcial de uma série geométrica com razão R é dada por: $\sum_{n=0}^k R^n = \frac{1-R^{k+1}}{1-R}$ e $\sum_{n=0}^k R^{n+1} = \frac{R-R^{k+2}}{1-R}$. Assim, a função de distribuição acumulada (f.d.a) pode ser escrita como

$$F(x) = -\frac{C(a)}{a-1} \left\{ e^{-\log(x)(a-1)+[\log(x)]} + (e^{-1} - 1) \left[\frac{1 - e^{-[\log(x)]}(a-2)}{e^{a-2} - 1} \right] - 1 \right\},$$

para $x > 1$ e $a > 2$.

5.4 Momento de ordem r

O momento de ordem r é calculado substituindo a por $(a-r)$ na integral da Equação (5.2). Assim,

$$E(X^r) = C(a) \int_1^{\infty} x^{-(a-r)} e^{[\log x]} dx = \frac{C(a)}{C(a-r)}, \quad (5.5)$$

que existe apenas quando $a-r > 2$.

5.5 Variância

Através da Equação (5.5), a variância é obtida.

$$Var(X) = \frac{C(a)[C^2(a-1) - C(a)C(a-2)]}{C(a-2)C^2(a-1)}, \quad a > 4. \quad (5.6)$$

5.6 Coeficiente de assimetria

O coeficiente de assimetria de Pearson, usado para quantificar o grau de simetria de uma distribuição, é dado por

$$\gamma_1(X) = \frac{m_3}{\sigma^3},$$

em que $m_i = \mathbb{E}[X - \mathbb{E}(X)]^i$, isto é, o i -ésimo momento central em torno da média, e σ é o desvio padrão da distribuição. Se $\gamma_1 > 0 (< 0)$, diz-se que a distribuição é assimétrica à direita (esquerda). Se $\gamma_1 = 0$, diz-se que a distribuição é simétrica.

Para encontrar o terceiro momento central (em torno da média), a Equação (5.5) foi utilizada e o resultado obtido foi

$$m_3 = \frac{C(a)[C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)]}{C^3(a-1)C(a-2)C(a-3)}$$

Assim, usando a Equação (5.6), $[Var(X)]^{3/2} = \sigma^3 = \frac{C^{3/2}(a)[C^2(a-1) - C(a)C(a-2)]^{3/2}}{C^{3/2}(a-2)C^3(a-1)}$.

Agora o coeficiente pode ser calculado: $\gamma_1(X) = \frac{m_3}{\sigma^3}$:

$$\gamma_1(X) = \frac{C^{1/2}(a-2)[C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)]}{C^{1/2}(a)C(a-3)[C^2(a-1) - C(a)C(a-2)]^{3/2}}, \quad a > 5$$

O valor de $\gamma_1(X)$ sempre será maior do que *zero*, o que indica que a distribuição floor é assimétrica à direita, para qualquer valor de $a > 5$.

5.7 Excesso de Curtose

O excesso de curtose, $K(X)$, é usado para comparar o formato de uma distribuição em relação à distribuição normal. Quando $K(X) = 0$, então a distribuição tem o mesmo achatamento do que o encontrado na distribuição normal, e tal distribuição recebe o nome de mesocúrtica. Quando $K(X) > 0$, a distribuição é menos achatada do que a distribuição normal, recebendo o nome de leptocúrtica. Finalmente, quando $K(X) < 0$, diz-se que tal distribuição é mais achatada do que a distribuição normal, sendo chamada de platicúrtica. Sua forma de calcular é dada por

$$K(X) = \frac{m_4}{\sigma^4} - 3,$$

A curtose da distribuição floor é obtida de forma semelhante à usada para calcular o coeficiente de assimetria, obtendo-se primeiro o quarto momento central e a variância ao quadrado, σ^4 . Assim, a curtose é dada por:

$$\begin{aligned} K(X) &= \frac{m_4}{\sigma^4} - 3 \\ &= C(a-1)C(a-2) \left\{ C^3(a-1) \prod_{i=2}^3 C(a-i) - 4C(a-4) \prod_{i=0}^2 C(a-i) + 6C^2(a)C(a-1) \right\} \times \\ &\quad \times \prod_{i=3}^4 C(a-i) - 4C(a) \prod_{i=0}^4 C(a-i) + C^2(a) \prod_{i=2}^4 C(a-i) \times \\ &\quad \times \{C(a)C(a-3)C(a-4)\}^{-1} \{C^2(a-1) - C(a)C(a-2)\}^{-2} - 3, \quad a > 6 \end{aligned}$$

5.8 Geração de valores aleatórios

A obtenção de valores aleatórios da distribuição floor é feita através do método de aceitação-rejeição (KRONMAL; PETERSON, 1981) implementado no software livre R. No Apêndice A este algoritmo e outras funções para a distribuição floor foram implementados no R. O algoritmo do método de aceitação-rejeição é como segue:

1. gere um valor de uma variável aleatória Y com distribuição g ;
2. gere um valor de uma variável U com distribuição uniforme padrão (independente de Y);

3. se $U \leq \frac{f(y)}{cg(y)}$, então faça $X = Y$ (“aceite”); caso contrário volte ao passo 1 (“rejeite”).

No algoritmo, $f(y)$ é a distribuição alvo, da qual queremos gerar valores (neste caso, a floor). Além da distribuição alvo, é necessária outra distribuição com comportamento semelhante, a distribuição candidata. Para isso, a distribuição exponencial com dois parâmetros (DETZEL; MINE, 2011), com parâmetro de posição igual a 1 foi utilizada. No algoritmo acima, ela ocupa o lugar de $g(y)$. Sua f.d.p. é definida por

$$g(y) = \frac{1}{\phi} e^{-\frac{(y-1)}{\phi}}, 1 \leq y < \infty,$$

em que $\phi > 0$.

A Figura 5.2 mostra um histograma com a distribuição de dez mil valores aleatórios, gerados através do método de aceitação-rejeição para $a = 9$, com a curva da densidade sobre o histograma.

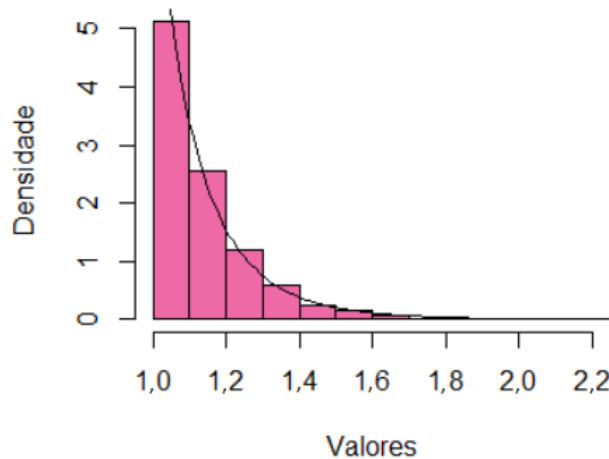


Figura 5.2: Histograma dos valores gerados com $a = 9$, $\mu = 0$, $\sigma = 1$

Além do histograma, foram calculadas medidas como média e variância dos valores gerados, e estes foram comparados com os valores da esperança e variância da distribuição. No caso para $a = 9$, a média dos valores foi igual a 1,147 (com três casas decimais), enquanto que a esperança da distribuição floor é aproximadamente igual a 1,144. A variância amostral foi igual a 0,020817 (com seis casas decimais), enquanto a variância da distribuição floor quando $a = 9$ é aproximadamente igual a 0,029560.

5.9 A distribuição floor e variação \mathcal{O} -regular

Andrade e Omev (2014) mostram que a distribuição floor possui variação \mathcal{O} -regular. Note que a densidade da distribuição floor pode ser escrita como

$$f(x) = C(a)x^{-a+1}e^{[\log x] - \log x},$$

para $x > 1$. Essa representação é como a mostrada na Equação (3.4), em que $A(x) = x^{a-1}$ e $b(x) = e^{\lfloor \log x \rfloor - \log x}$. É fácil notar que $A(x)$ pertence à classe de variação regular com índice $-a + 1$, pois

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lim_{x \rightarrow \infty} \frac{(\lambda x)^{a-1}}{x^{a-1}} = \lambda^{-a+1}$$

Como $\lfloor \log x \rfloor \leq \log x \leq \lfloor \log x \rfloor + 1$, então tem-se também que $-1 \leq \log x \leq 0$, o que mostra que $e^{-1} \leq b(x) \leq 1$.

Dessa forma, através do Teorema 3.3.2, tem-se que a distribuição floor é de variação \mathcal{O} -regular. Os índices de Matuszewska são dados por $\alpha(f) = \beta(f) = -a + 1$ (ANDRADE; OMEY, 2014).

Nesta seção, várias propriedades da distribuição floor foram apresentadas. No próximo capítulo, o fato de a distribuição floor ser de variação \mathcal{O} -regular será utilizado e poderá ser observado que esta distribuição é adequada para solucionar conflitos de informação, em que nos dados há uma informação conflitante.

6 APLICAÇÃO

Neste capítulo, a vantagem da distribuição floor sobre a exponencial será evidente na resolução de conflitos de informação. Será mostrado que a distribuição a posteriori automaticamente rejeitará a observação conflitante, se a distribuição floor for utilizada como distribuição dos dados. O software livre `OpenBugs` será utilizado.

O banco de dados utilizado neste trabalho refere-se ao tempo médio de espera na fila do restaurante universitário (RU) da Universidade Federal de Minas Gerais (UFMG) (FUMP, 2014). Os dados disponíveis em FUMP (2014) mostram os tempos mínimo e máximo de espera nas filas do RU no horário de pico, de 11h 20 às 12h 30. Neste trabalho serão utilizados os dados referentes ao RU Setorial I, que compreende 16 observações realizadas ao longo dos meses de agosto a novembro de 2014 (nem todos os dias são observados).

Os dados podem ser encontrados no ANEXO A. Seguindo a mesma metodologia encontrada no exemplo dado por Andrade e Omei (2014), os dados são modelados usando duas distribuições diferentes com parâmetro de escala: exponencial, com densidade dada por $f(x|\theta) = 1/\theta e^{-x/\theta}, x > 0$ e floor, com densidade dada por

$$f(x|\theta) = \frac{C(a)}{\theta} \left(\frac{x}{\theta}\right)^{-a} e^{\log\lfloor x/\theta \rfloor},$$

para $x > \theta$. A informação a priori considerada é $\theta \sim \text{Gama}(2, 10)$, para os dois casos. No Apêndice B os modelos implementados no `OpenBugs` podem ser encontrados.

O gráfico mostrado na Figura 6.1 é similar ao apresentado por Andrade e Omei (2014). O gráfico é gerado após simularmos a distribuição a posteriori para cada modelo, fazendo com que uma das observações tenda a infinito.

No caso da distribuição exponencial, conforme o *outlier* se distancia dos outros dados, o valor médio a posteriori, que é a estimativa para o parâmetro de escala θ , é afetado pela informação conflitante. Isso não ocorre com a distribuição floor, que permanece robusta ao *ou-*

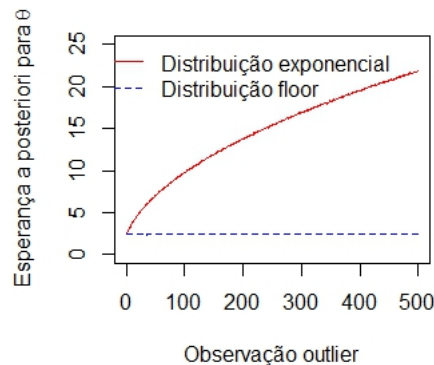


Figura 6.1: Estimativas a posteriori para θ segundo os dois modelos

tlier. Em termos práticos, pode-se supor que, se alguém por engano registrasse incorretamente o tempo de espera, de forma que essa observação destoasse das demais, não haveria problemas ao utilizar a distribuição floor, pois as estimativas a posteriori não seriam afetadas, mesmo com a presença da informação conflitante.

7 CONCLUSÕES

Neste trabalho, foi visto que as classes de variação regular, e, em seu caso mais abrangente, a classe de variação \mathcal{O} -regular, podem contribuir bastante para o estudo de caudas pesadas, com o objetivo de obter modelos mais robustos a informações conflitantes.

Os primeiros dois capítulos foram responsáveis por elucidar o processo histórico do tema, como também apresentar o contexto bayesiano, muito importante neste trabalho.

Os capítulos 3 e 4 são os que dão a base teórica desta monografia. O primeiro mostra como o conceito de caudas pesadas está intrinsecamente atrelado ao de distribuições de variação regular e \mathcal{O} -regular. O segundo, em como essas classes de distribuições podem ser utilizadas para resolver conflitos de informação, especificadamente no caso de parâmetro de escala e rejeição de *outliers*.

Para a descrição da distribuição floor, no capítulo 5, o método de aceitação-rejeição foi utilizado para gerar valores aleatórios, e se mostrou bastante razoável. Para as outras propriedades da distribuição, apenas álgebra e cálculo integral foram utilizados, mostrando assim, que a distribuição floor não é muito difícil de ser estudada.

No estudo de simulação, no capítulo 6, vimos que a distribuição floor é robusta a informação conflitante proveniente da amostra, e que em casos em que há suspeita de presença de *outliers*, é aconselhável trabalhar com a distribuição floor.

Neste trabalho, o interesse centralizou na rejeição de informações conflitantes provenientes dos dados e em famílias com parâmetro de escala. Na literatura, casos que fugiram ao tema deste trabalho foram estudados. Assim, pode-se notar que outras contribuições podem ser feitas. Como trabalhos futuros, a distribuição floor pode ser estudada no caso em que há parâmetro de posição, e também na rejeição de informação a priori.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALJANČIĆ, S.; ARANDELOVIĆ, D. O-regularly varying functions. *Publ. Inst. Math.* (Beograd). Vol. 22, No. 36., p. 5-22. 1977.
- ANDRADE, J. A. A. **Bayesian Robustness Modelling Using Regularly Varying Distributions**. Tese (Doutorado em Estatística). Department of Probability and Statistics, School of Mathematics and Statistics. University of Sheffield. 2005
- ANDRADE, J. A. A.; O'HAGAN, A. Bayesian robustness modelling using regularly varying distributions. *Bayesian Analysis*, 1: p. 169-188. 2006.
- ANDRADE, J. A. A.; O'HAGAN, A. Bayesian robustness modelling of location and scale parameters. *Scandinavian Journal of Statistics* 38, p. 691-711. 2011.
- ANDRADE, J. A. A.; OMEY, E. Modelling conflicting information using subexponential distributions and related classes. *Annals of the Institute of Statistical Mathematics*. Vol. 65 (3), 491 - 511. 2013.
- ANDRADE, J. A. A.; OMEY, E. Resolution of conflict of information using O-regularly varying functions. (Em submissão). 2014.
- AVAKUMOVIĆ, V. G. Über einen O-inversionsatz. *Bull. Int. Acad. Youg. Sci.*, 29-30, p. 107-17. 1936.
- BERNARDO, J. M. Bayesian Statistics. *Encyclopedia of Life Support Systems (EOLSS)*. Probability and Statistics, (R. Viertl, ed). Oxford, UK: UNESCO. 2003.
- BINGHAM, N. H.; GOLDIE, C. M.; TEUGELS, J. L. **Regular variation**. Vol. 27 da "Encyclopedia of Mathematics and Its Applications". Cambridge: Cambridge University Press. 1987.
- BOX, G. E. P.; TIAO, G. C. **Bayesian Inference in Statistical Analysis**. 1ª edição. Addison-Wesley. 1973.
- DAVIS, H. T. **Theory of Econometrics**. Bloomington, Indiana. 1941
- DAWID, A. P. Posterior Expectations for Large Observations. *Biometrika*, Vol. 60, No. 3, p. 664-667. Publicado por: Biometrika Trust. 1973. Disponível em: < <http://dx.doi.org/10.2307/2335019> >. Data de acesso: 15/05/2014.
- DE FINETTI, B. The Bayesian Approach to the Rejection of Outliers. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 199-210, University of California Press, Berkeley, California. 1961.
- DE HAAN, L. (1970). **On regular variation and its application to the weak convergence of sample extremes**. Mathematical Centre Tracts 32, Mathematisch Centrum, Amsterdam.
- DESGAGNÉ, Alain. Full Robustness in Bayesian Modelling of a Scale Parameter. *Bayesian Analysis*, Vol. 8, No. 1, p. 187 - 220. 2013. Disponível em: < <http://dx.doi.org/10.1214/13-BA808> >. Data de acesso: 06/05/2014.

- DETZEL, D. H. M.; MINE, M. R. M. Modelagem de Quantidades Precipitadas em Escala Diária: Uma Análise Comparativa. *RBRH — Revista Brasileira de Recursos Hídricos*. Vol. 16, No. 2, p. 101-110. 2011.
- DJURČIĆ, D. \mathcal{O} -Regularly Varying Functions and Strong Asymptotic Equivalence. *Journal of Mathematical Analysis and Applications*. Vol. 220, 2. p. 451-461. 1998.
- FELLER, W. **An Introduction to Probability Theory and Its Applications**, Vol. II. 2^a edição. John Wiley & Sons. 1971
- FISHER, R. A. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London*. Series A. Vol. 222, p. 309-368. 1922. Disponível em: < <http://www.jstor.org/stable/91208> >. Data de acesso: 15/05/2014.
- FUMP. Assistência estudantil da UFMG. Monitoramento das filas. Disponível em: < <http://www.fump.ufmg.br/conteudo.aspx?pagina=174> >. Data de acesso: 01/11/2014.
- KARAMATA. Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)*. 4, p. 38-53. 1930.
- KRONMAL, R. A.; PETERSON JR, A. V. A Variant of the Acceptance-Rejection Method for Computer Generation of Random Variables *Journal of the American Statistical Association*. Vol. 76, No. 374, p. 446-451. 1981.
- LANDAU, E. Sur les valeurs moyennes de certaines fonctions arithmetiques. *Bull. Acad. Roy. Belgique*, p. 443-72. 1911.
- LEE, P. M. **Bayesian Statistics: An Introduction**. 4^a edição. Chichester, John Wiley & Sons. 2012.
- LEONARD, T. H. A Personal History of Bayesian Statistics. *WIREs Comp Stat*, Vol. 6, p. 80-115. 2014. Disponível em: < <http://dx.doi.org/10.1002/wics.1293> >. Data de acesso: 21/05/2014.
- LINDLEY, D. V. The choice of variables in multiple regression (with discussion). *Journal of the Royal Statistical Society B* 30, 31-66. 1968
- NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46, p. 323-351. 2006.
- O'HAGAN, A. On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society B* 41, p. 358-367. 1979.
- O'HAGAN, A. Outliers and Credence for Location Parameter Inference. *Journal of the American Statistical Association*, Vol. 85, No. 40, p. 172-176. 1990.
- O'HAGAN, A.; PERICCHI, L. Bayesian heavy-tailed models and conflict resolution: a review. *Brazilian Journal of Probability and Statistics* 26, 372-401. 2012.
- PARETO, V. **Cours d'Economie Politique**. Lausanne. 1897.
- PÓLYA, G. Über eine neue Weise, bestimmte Integrale in der analytischen Zahlentheorie zu gebrauchen. *Nachrichten Göttinger*, p 149-159. 1917.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2011. Disponível em < <http://www.R-project.org> >. Data de acesso: 10/05/2014.

SANTOS JUNIOR, J. D. O. **Discussões sobre a Relação entre Distribuições de Cauda Pesada e Conflitos de Informação em Inferência Bayesiana**. 2007. Dissertação (Mestrado em Estatística). Instituto de Matemática, Estatística e Computação Científica. Universidade Estadual de Campinas. Campinas, São Paulo.

SENETA, E. **Functions of regular variation**. Lecture Notes in Mathematics Vol. 506, New York: Springer. 1976

STIGLER, S. M. Who Discovered Bayes's Theorem? *The American Statistician*. Vol. 37, No. 4, Parte 1, p. 290-296. Publicado por: American Statistical Association. 1983. Disponível em: < <http://dx.doi.org/10.2307/2682766> >. Data de acesso: 21/05/2014.

VALIRON. G. Sur les fonctions entières d'ordre fini, et d'ordre nul, et en particulière les fonctions à correspondance régulière. *Ann. Fac. Sci. Univ. Toulouse* (3), 5, 117-257. 1913.

APÊNDICE A – FUNÇÕES NO R PARA A DISTRIBUIÇÃO FLOOR

```

# Constante normalizadora
C= function(a){(((a-1)*(exp(2)-exp(a)))/(exp(1)-exp(a)))}

# Função densidade de probabilidade (dfloor)
dfloor <- function(x) {C(a)*s^(-1)*((x-m)/s)^(-a)*
exp(floor(log((x-m)/s)))}

# Gráfico
a=3;s=2;m=2
curve(dfloor(x),xlim=c(4,30),
main=expression(paste("Parâmetros ", a, "=3, ", sigma, "=2, ", mu, "=2"))
, col=2,lwd=2)

# Função de distribuição acumulada (pfloor)

pfloor=function(x,a){
-C(a)/(a-1)*
(exp(-log(x)*(a-1)+floor(log(x))))+
(exp(-1)-1)*((1-exp(-floor(log(x)))*(a-2)))/(exp(a-2)-1))-
1
)
}

curve(pfloor(x,4),xlim=c(1,5), ylim=c(0,1), ylab=
'Função de distribuição acumulada')

# Quantis (qfloor)

# usando a função optimize

qfloor=function(p){
variable=function(x,a) pfloor(x,a)-p
z=optimize(variable, c(1,100000), tol=0.0001, a=3)
return(z)
}
qfloor(0.75)

# Esperança

E=function(r,a){C(a)/C(a-r)}

```

```

# Gerando valores aleatórios (rfloor) para a=9

E(1,9)-1
  #0 valor de lambda é escolhido de forma que a esperança da dist.
  #exponencial seja igual à esperança da distribuição floor menos 1.
#Este "menos um" ocorre porque a cada valor aceito no algoritmo
#será acrescentada uma unidade, para que os valores
#sejam maiores que 1.
rfloor=function(n,a=9){
  lambda=6.94
  a=9
  c=8.2
  sample.x = rexp(5000,lambda)
  accept = c()
  for(i in 1:length(sample.x)){
    U = runif(1, 0, 1)
    if(c*U <= lambda*C(a)*((sample.x[i])^-a)*
exp(floor(log(sample.x[i]))+(sample.x[i]/lambda))) {
      accept[i] = '1'
    }
    else if(c*U > lambda*C(a)*((sample.x[i])^-a)*
exp(floor(log(sample.x[i]))+(sample.x[i]/lambda))) {
      accept[i] = '0'
    }
  }
  T = data.frame(sample.x, accept =
factor(accept, levels= c('1','0')))
del.t=T[T$accept=="1",]
rx=del.t[,1] # valores aceitos no método
  #(são os possíveis valores aleatórios
#de  $X \sim \text{Floor}(8)$ )

srx=sort(rx) # ordenando
Valores=srx+1 # aumentando uma unidade para que os
#valores sejam acima de 1
return{Valores}
}
rfloor(100)

# testes

#media amostral

```

```
mean(valores)
#media populacional(distribuição)
E(1,9)

#variancia amostral
var(valores)
#variancia da distribuição
Var=function(a){E(2,a)-(E(1,a))^2}
Var(9)

#assimetria amostral
am1=(3*(mean(valores)-median(valores)))/sd(valores);am1
#assimetria da distribuição
m3 = function (a) {E(3, a)-3*E(1, a)*E(2, a)+2*E(1, a)^3}
assim=function(a){m3(a)/(Var(a))^(3/2)}
assim(9)
```


APÊNDICE B – MODELOS NO OPENBUGS

As distribuições a posteriori para os modelos exponencial e floor foram obtidas através de métodos Monte Carlo via Cadeias de Markov (MCMC) com utilização do software OpenBugs. Foram usados $update=10000$ e $burn=100$.

```
#Exponencial
model;{
for (k in 1:500){
for (i in 2:N) { X.rep[i,k] <- X[i]}
X.rep[1,k] <- k*10
for( i in 1:N){
X.rep[i,k] ~ dexp(is[k])
}
is[k] <- 1/s[k]
s[k] ~ dgamma(2,10)
}
}
list(N=16, X=c(4.083333333,5.416666667,7.383333333,7.308333333,
2.358333333,8.883333333,2.666666667,3.308333333,
5.541666667,5.9,3.408333333,4.108333333,
5.016666667,9.05,9.616666667,6.075))

#Floor
model;{
for (k in 1:500){
for (i in 2:N) { X.rep[i,k] <- X[i]}
X.rep[1,k] <- k*10
for( i in 1:N){
zeros[i,k] <- 0
zeros[i,k] ~ dloglik(phi[i,k])
phi[i,k] <- -a*(log(X.rep[i,k]/sigma[k]))-log(sigma[k])
+round(log(X.rep[i,k]/sigma[k])-0.5)
}
sigma[k]~dgamma(2,10)
}
}
list(N=16,a=3, X=c(4.083333333,5.416666667,7.383333333,7.308333333,
2.358333333,8.883333333,2.666666667,3.308333333,
5.541666667,5.9,3.408333333,4.108333333,
5.016666667,9.05,9.616666667,6.075))
```

ANEXO A – DADOS

Os dados em FUMP (2014) mostram o tempo mínimo e máximo de espera. Além desses dados, neste trabalho foi adicionada a variável tempo médio de espera, que é a média aritmética entre os tempos mínimo e máximo. Nas colunas 2-4 os tempos estão em segundos, e na última coluna os dados foram convertidos em minutos (min) e implementados no OpenBugs.

Tabela A.1: Tempo de espera, em segundos, na fila do RU (UFMG) de 30 de agosto a 5 de novembro de 2014

Observação	Tempo mínimo	Tempo máximo	Tempo médio	Tempo médio (min)
30/08	29	700	364,5	6,075
2/09	261	893	577	9,617
3/09	273	813	543	9,050
4/09	0	602	301	5,017
5/09	0	493	246,5	4,108
9/09	0	409	204,5	3,408
10/09	196	512	354	5,900
11/09	74	591	332,5	5,542
12/09	78	319	198,5	3,308
16/09	0	320	160	2,666
18/09	286	780	533	8,883
19/09	0	283	141,5	2,358
1/10	147	730	438,5	7,308
9/10	141	745	443	7,383
29/10	135	515	325	5,417
5/11	138	352	245	4,083